

NMSA202 Pravděpodobnost a matematická statistika

POZNÁMKY K PŘEDNÁŠCE

Naposledy upraveno dne 23. června 2020.



matfyz

Katedra pravděpodobnosti a matematické statistiky
Matematicko-fyzikální fakulta University Karlovy

Tento učební text obsahuje přehled všech vět, definic, tvrzení a poznámek probíraných v přednášce „NMSA202 Pravděpodobnost a matematická statistika“ v rámci bakalářského studia oboru „Obecná matematika“ na MFF UK. Nejedná se o plnohodnotnou učebnici ani skriptu, protože zde nejsou uvedeny důkazy vět a tvrzení, chybí některé příklady a není zde obsažena látka probíraná na cvičení. Při přípravě na zkoušku je nutné tento text doplnit poznámkami z přednášek a cvičení.

Od čtvrtého týdne semestru (od přednášky č.7) je text skript kompletní, plně odpovídající přednášce, tedy obsahuje všechny důkazy, poznámky, příklady a vysvětlení proč, co, a jak děláme.

OBSAH

ZNAČENÍ	6
1 PRAVDĚPODOBNOSTNÍ PROSTOR	8
1.1 Axiomatická definice pravděpodobnosti	8
1.2 Podmíněná pravděpodobnost	9
2 NÁHODNÉ VELIČINY	11
2.1 Náhodná veličina a její rozdělení	11
2.2 Momenty náhodné veličiny	14
2.3 Některá rozdělení náhodné veličiny	16
2.4 Rozdělení funkce náhodné veličiny	16
3 NÁHODNÉ VEKTORY	19
3.1 Náhodné vektory	19
3.2 Nezávislé náhodné veličiny	24
3.3 Momenty náhodného vektoru	27
3.4 Rozdělení transformovaného náhodného vektoru	33
3.5 Příklady rozdělení náhodného vektoru	38
3.5.1 Multinomické rozdělení	38
3.5.2 Mnohorozměrné normální rozdělení	39
4 LIMITNÍ VĚTY	45
4.1 Cantelliho a Borelova věta	46
4.2 Konvergence posloupnosti náhodných veličin	49
4.3 Slabý zákon velkých čísel	52
4.4 Silný zákon velkých čísel	53
4.5 Centrální limitní věta	61
5 STATISTIKA	70
5.1 Bodový odhad	71
5.1.1 Metoda momentů	75
5.1.2 Metoda maximální věrohodnosti	78
5.2 Intervalový odhad	82
5.2.1 Intervalové odhady v normálním modelu	84
5.2.2 Intervalové odhady založené na CLV	91
5.3 Testování hypotéz	93

ZNAČENÍ

\mathbb{R}	množina reálných čísel
\mathbb{N}	množina přirozených čísel
\mathbb{Z}	množina celých čísel
\mathbf{a}^T	transpozice vektoru \mathbf{a}
$\ \mathbf{a}\ $	eukleidovská norma vektoru \mathbf{a}
\mathbb{I}_n	jednotková matice $n \times n$
Ω	prostor elementárních jevů
\mathcal{A}	σ -algebra náhodných jevů na Ω
\mathcal{B}	borelovská σ -algebra na \mathbb{R}
\mathcal{B}^n	borelovská σ -algebra na \mathbb{R}^n
$\mathbb{1}_M$	indikátor množiny M
M^c	doplněk množiny M
δ_x	Diracova míra v bodě x
P	pravděpodobnost
P_X	rozdělení náhodné veličiny X
S_X	nosič rozdělení náhodné veličiny X
$P_{\mathbf{X}}$	rozdělení náhodného vektoru \mathbf{X}
$S_{\mathbf{X}}$	nosič rozdělení náhodného vektoru \mathbf{X}
$X \sim \mathcal{L}$	X má rozdělení \mathcal{L}
λ	Lebesgueova míra na \mathbb{R}
λ^n	Lebesgueova míra na \mathbb{R}^n
L^p	množina náhodných veličin na (Ω, \mathcal{A}, P) s konečným p -tým absolutním momentem
$\ h\ _\infty$	supremová norma funkce h
$E X$	střední hodnota náhodné veličiny X
$\text{var } X$	rozptyl náhodné veličiny X
σ_X	směrodatná odchylka náhodné veličiny X
σ_X^2	rozptyl náhodné veličiny X
μ_k	k -tý moment náhodné veličiny X
$\text{cov}(X_1, X_2)$	kovariance náhodných veličin X_1 a X_2
$\text{cor}(X, Y)$	korelační koeficient náhodných veličin X a Y
$\text{Var}(\mathbf{X})$	varianční matice náhodného vektoru \mathbf{X}

f_X	hustota náhodné veličiny X
$f_{\mathbf{X}}$	hustota náhodného vektoru \mathbf{X}
F_X	distribuční funkce náhodné veličiny X
$F_{\mathbf{X}}$	distribuční funkce náhodného vektoru \mathbf{X}
F_X^{-1}	kvantilová funkce náhodné veličiny X
$u_X(\alpha)$	α -kvantil náhodné veličiny X
$N(0, 1)$	normované normální rozdělení
u_α	α -kvantil rozdělení $N(0, 1)$
φ	hustota normovaného normálního rozdělení
Φ	distribuční funkce normovaného normálního rozdělení
$\chi_n(\alpha)$	α -kvantil rozdělení χ_n^2
$t_n(\alpha)$	α -kvantil Studentova t_n -rozdělení
\widehat{F}_n	empirická distribuční funkce
\bar{X}_n	výběrový průměr náhodného výběru X_1, \dots, X_n
S_n^2	výběrový rozptyl náhodného výběru X_1, \dots, X_n
\widehat{m}_r	r -tý výběrový moment náhodného výběru X_1, \dots, X_n
\mathcal{F}	pravděpodobnostní model pro pozorovaná data
$\widehat{\theta}_n$	odhad parametru θ na základě náhodného výběru X_1, \dots, X_n
$L(\theta, \mathbf{x})$	věrohodnost θ pro pozorování \mathbf{x}
$l(\theta, \mathbf{x})$	logaritmická věrohodnost θ pro pozorování \mathbf{x}
$\xrightarrow[n \rightarrow \infty]{P}$	konvergence v pravděpodobnosti
$\xrightarrow[n \rightarrow \infty]{sj}$	konvergence skoro jistě
$\xrightarrow[n \rightarrow \infty]{d}$	konvergence v distribuci

1 PRAVDĚPODOBNOSTNÍ PROSTOR

1.1 AXIOMATICKÁ DEFINICE PRAVDĚPODOBNOSTI

Bud' Ω neprázdná množina a \mathcal{A} σ -algebra na množině Ω .

Definice 1.1 (pravděpodobnostní prostor) *Pravděpodobnost* P je množinová funkce $P : \mathcal{A} \rightarrow [0, 1]$ splňující

- (i) $P(A) \geq 0, \forall A \in \mathcal{A}$;
- (ii) $P(\Omega) = 1$;
- (iii) jsou-li $A_1, A_2, \dots \in \mathcal{A}$ po dvou disjunktní, pak $P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$.

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

Věta 1.1 Bud' (Ω, \mathcal{A}, P) pravděpodobnostní prostor, $A, B \in \mathcal{A}$, $\{A_n\}_{n=1}^{\infty} \subset \mathcal{A}$. Pro P platí:

- (i) $P(\emptyset) = 0$;
- (ii) P je konečně aditivní;
- (iii) $P(A^c) = 1 - P(A)$;
- (iv) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$;
- (v) P je monotónní, tj. $A \subset B \Rightarrow P(A) \leq P(B)$;
- (vi) je-li $A_1 \subseteq A_2 \subseteq \dots$, pak $\lim_{n \rightarrow \infty} P(A_n) = P\left(\bigcup_{i=1}^{\infty} A_i\right)$;
- (vii) je-li $A_1 \supseteq A_2 \supseteq \dots$, pak $\lim_{n \rightarrow \infty} P(A_n) = P\left(\bigcap_{i=1}^{\infty} A_i\right)$;
- (viii) $B \subset A \Rightarrow P(A \setminus B) = P(A) - P(B)$;
- (ix) princip inkluze a exkluze, tj.

$$P\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n P(A_i) - \sum_{1 \leq i < j \leq n} P(A_i \cap A_j) + \sum_{1 \leq i < j < k \leq n} P(A_i \cap A_j \cap A_k) - \dots + (-1)^{n-1} P\left(\bigcap_{i=1}^n A_i\right).$$

Poznámka. Speciální případ bodu (vii) z předchozího tvrzení je tzv. „spojitost pravděpodobnosti v 0“ tj. tvrzení

$$\left(A_1 \supseteq A_2 \supseteq \dots \text{ a } \bigcap_{i=1}^{\infty} A_i = \emptyset \right) \Rightarrow \lim_{n \rightarrow \infty} P(A_n) = 0.$$

1.2 PODMÍNĚNÁ PRAVDĚPODOBNOST

Definice 1.2 (podmíněná pravděpodobnost) Bud' (Ω, \mathcal{A}, P) pravděpodobnostní prostor a $A, B \in \mathcal{A}$ splňující $P(B) > 0$. Podmíněnou pravděpodobnost jevu A za podmínky (jevu) B definujeme vztahem

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

Věta 1.2 Bud' (Ω, \mathcal{A}, P) pravděpodobnostní prostor a $B \in \mathcal{A}$ takové, že $P(B) > 0$. Pak zobrazení $P(\cdot|B) : \mathcal{A} \rightarrow [0, 1]$ splňuje definici pravděpodobnosti.

*Zde končí
předn. 1
(20.2.)*

Věta 1.3 Bud' (Ω, \mathcal{A}, P) pravděpodobnostní prostor a $B \in \mathcal{A}$ takové, že $P(B) > 0$. Pro podmíněnou pravděpodobnost platí:

- (i) $P(A \cup C|B) = P(A|B) + P(C|B) - P(A \cap C|B)$;
- (ii) $B \subset A \Rightarrow P(A|B) = 1$;
- (iii) $A \cap B = \emptyset \Rightarrow P(A|B) = 0$;
- (iv) $P(A|\Omega) = P(A)$;
- (v) pokud $P(\{\omega\}) > 0$ pro nějaké $\omega \in \Omega$, pak $\forall A \in \mathcal{A}$ platí $P(A|\{\omega\}) \in \{0, 1\}$.

Poznámka. Podmíněná pravděpodobnost není míra v druhé proměnné, tj. obecně neplatí rovnost mezi $P(A|B \cup C)$ a $P(A|B) + P(A|C)$, a to ani v případě, kdy jsou B a C disjunktní!

Věta 1.4 (o násobení pravděpodobností) Bud' (Ω, \mathcal{A}, P) pravděpodobnostní prostor, jevy $A_1, \dots, A_n \in \mathcal{A}$ takové, že $P(A_1 \cap \dots \cap A_{n-1}) > 0$. Pak platí

$$P(A_1 \cap \dots \cap A_n) = P(A_n|A_1 \cap \dots \cap A_{n-1})P(A_{n-1}|A_1 \cap \dots \cap A_{n-2}) \dots P(A_2|A_1)P(A_1).$$

Věta 1.5 (o celkové pravděpodobnosti) Bud' (Ω, \mathcal{A}, P) pravděpodobnostní prostor, $A \in \mathcal{A}$ a $\{B_n\} \subset \mathcal{A}$ konečná nebo spočetná posloupnost vzájemně neslučitelných jevů splňujících $P(B_n) > 0 \forall n$ a $P(\bigcup_n B_n) = 1$. Pak platí

$$P(A) = \sum_n P(A|B_n)P(B_n).$$

Věta 1.6 (Bayesův vzorec) Za předpokladů věty 1.5 a předpokladu $P(A) > 0$ platí

$$P(B_i|A) = \frac{P(A|B_i)P(B_i)}{\sum_n P(A|B_n)P(B_n)} \text{ pro všechna } i.$$

Definice 1.3 (nezávislost 2 jevů) Náhodné jevy A a B jsou *nezávislé*, pokud platí $P(A \cap B) = P(A)P(B)$.

Poznámka. Nezávislost dvou jevů interpretujeme tak, že výskyt jednoho jevu neovlivní šanci na výskyt druhého jevu. Nejde tedy o kauzalitu, ale jen o proporcionální překryv pravděpodobností těchto dvou jevů.

Tvrzení 1.7 Jsou-li jevy A a B nezávislé, pak také jevy A a B^c jsou nezávislé. Pokud navíc $P(B) > 0$, pak $P(A|B) = P(A)$.

Zde končí
předn. 2
(21.2.)

Definice 1.4 (vzájemná nezávislost jevů) Buď $\{A_l, l \in \Lambda\}$ systém náhodných jevů. Jevy nazveme *vzájemně nezávislé*, pokud pro každé $n \in \mathbb{N}$ a každou n -prvkovou množinu $I \subset \Lambda$ platí

$$P\left(\bigcap_{i \in I} A_i\right) = \prod_{i \in I} P(A_i).$$

Věta 1.8 Buď $C = \{B_1, B_2, \dots, B_k\}$, $k \in \mathbb{N}$ systém nezávislých jevů. Nahradíme-li libovolnou podmnožinu těchto jevů jejich doplňky, dostaneme opět systém nezávislých jevů.

Tvrzení 1.9 Jsou-li jevy $A_1, \dots, A_n, B_1, \dots, B_m$ vzájemně nezávislé a $P(B_1 \cap \dots \cap B_m) > 0$, pak

$$P(A_1 \cap \dots \cap A_n | B_1 \cap \dots \cap B_m) = P(A_1 \cap \dots \cap A_n) = P(A_1) \dots P(A_n).$$

2 NÁHODNÉ VELIČINY

2.1 NÁHODNÁ VELIČINA A JEJÍ ROZDĚLENÍ

Definice 2.1 Budťe (Ω, \mathcal{A}) a (Ω', \mathcal{A}') dva stavové prostory. Pak každé měřitelné zobrazení $X : \Omega \rightarrow \Omega'$ nazveme *náhodný element* z Ω' .

Definice 2.2 Budť (Ω, \mathcal{A}, P) pravděpodobnostní prostor. Měřitelné zobrazení $X : (\Omega, \mathcal{A}) \rightarrow (\mathbb{R}, \mathcal{B})$ nazveme *(reálnou) náhodnou veličinou*.

Poznámka. Značení: Budť $A \in \mathcal{B}$. Místo $\{\omega \in \Omega : X(\omega) \in A\}$ budeme psát $\{X \in A\}$. A protože $\{X \in A\}$ je náhodný jev z \mathcal{A} , můžeme mu přiřadit $P(X \in A)$.

Definice 2.3 Budť X náhodná veličina. Množinový systém $\{X^{-1}(B), B \in \mathcal{B}\}$ nazveme σ -algebrou náhodných jevů generovaných náhodnou veličinou X nebo σ -algebrou indukovanou náhodnou veličinou X , a značíme ji $\sigma(X)$.

Poznámka.

- Měřitelnost náhodné veličiny X stačí ověřit na generátoru \mathcal{B} např. na $\{(-\infty, a], a \in \mathbb{R}\}$.
- $\sigma(X)$ je skutečně σ -algebra.
- Pro každý množinový systém $\mathcal{S} \subseteq \mathcal{P}(\mathbb{R})$ platí $\sigma(X^{-1}\mathcal{S}) = X^{-1}(\sigma\mathcal{S})$, takže

$$\sigma(X) = \sigma\{X^{-1}(-\infty, a], a \in \mathbb{R}\} = \sigma\{\{\omega \in \Omega : X(\omega) \leq a\}, a \in \mathbb{R}\}.$$

Definice 2.4 Rozdělením náhodné veličiny $X : (\Omega, \mathcal{A}) \rightarrow (\mathbb{R}, \mathcal{B})$ nazveme indukovanou pravděpodobnostní míru P_X na $(\mathbb{R}, \mathcal{B})$ definovanou jako $P_X(B) = P(\{\omega \in \Omega : X(\omega) \in B\})$, $B \in \mathcal{B}$.

Poznámka. Takže $P_X = X(P)$ je obraz míry P v zobrazení X a $(\Omega, \mathcal{A}, P) \rightarrow (\mathbb{R}, \mathcal{B}, P_X)$.

Věta 2.1 (o přenosu integrace pro P_X) Budť X náhodná veličina a budť $h : (\mathbb{R}, \mathcal{B}) \rightarrow (\mathbb{R}, \mathcal{B})$ měřitelná funkce. Pak platí

$$\int_{\Omega} h(X(\omega)) dP(\omega) = \int_{\mathbb{R}} h(x) dP_X(x),$$

pokud má alespoň jedna strana smysl.

Poznámka. Takže pro práci s náhodnou veličinou nepotřebujeme znát celé (Ω, \mathcal{A}, P) a X , ale stačí nám znát míru P_X na $(\mathbb{R}, \mathcal{B})$!

Poznámka. Navíc míru P_X můžeme vyjádřit v (pro nás) výhodném tvaru, neboť máme Radon-Nikodýmovu větu.

Definice 2.5 Buď X náhodná veličina, P_X její rozdělení a μ σ -konečná míra na $(\mathbb{R}, \mathcal{B})$ taková, že $P_X \ll \mu$. Potom $f(x) = \frac{dP_X}{d\mu}(x)$ se nazývá *hustota náhodné veličiny X vzhledem k míře μ* .

Poznámka.

- $f(x)$ je určena jednoznačně μ -skoro všude.
- Pokud je měřitelná $g : (\mathbb{R}, \mathcal{B}) \rightarrow (\mathbb{R}, \mathcal{B})$ nezáporná, nebo pro ni platí $\int_{\mathbb{R}} |g(x)| dP_X(x) < \infty$, pak $\int_{\mathbb{R}} g(x) dP_X(x) = \int_{\mathbb{R}} g(x)f(x) d\mu(x)$.

*Zde končí
předn. 3
(27.2.)*

Věta 2.2 Buď X náhodná veličina, P_X její rozdělení a $B \in \mathcal{B}$. Pak platí následující rovnosti

$$\begin{aligned} P(X \in B) &= P(\{\omega \in \Omega : X(\omega) \in B\}) \\ &= \int_{\Omega} \mathbb{1}_B(X(\omega)) dP(\omega) = \int_{\mathbb{R}} \mathbb{1}_B(x) dP_X(x) = \int_B 1 dP_X(x) = P_X(B). \end{aligned}$$

Pokud je navíc f hustota X vzhledem k σ -konečné míře μ , pak platí i

$$P(X \in B) = \int_{\Omega} \mathbb{1}_B(X(\omega)) dP(\omega) = \int_{\mathbb{R}} \mathbb{1}_B(x)f(x) d\mu(x) = \int_B f(x) d\mu(x).$$

Poznámka. Takže při znalosti P_X nebo $f(x)$ jsme schopni určit pravděpodobnosti všech jevů ze $\sigma(X)$.

Definice 2.6 Buď X náhodná veličina. Funkci $F_X : \mathbb{R} \rightarrow [0, 1]$ definovanou jako $F_X(x) = P(X \leq x)$, $x \in \mathbb{R}$, nazveme *distribuční funkcí* náhodné veličiny X .

Poznámka. $F_X(x) = P_X((-\infty, x])$ a tedy distribuční funkce jednoznačně určuje rozdělení P_X náhodné veličiny X . Ale různé náhodné veličiny mohou mít stejnou distribuční funkci (a stejné rozdělení).

Věta 2.3 (Vlastnosti distribuční funkce) Buď F_X distribuční funkce náhodné veličiny X . Pak

- (i) F_X je neklesající;
- (ii) F_X je zprava spojitá;
- (iii) $\lim_{x \rightarrow -\infty} F_X(x) = 0$, $\lim_{x \rightarrow \infty} F_X(x) = 1$.

Poznámka. Z teorie míry víme, že ke každé funkci splňující předpoklady (i)–(iii) existuje právě jedna konečná borelovská míra μ na \mathbb{R} splňující

$$\mu((a, b]) = F(b) - F(a), \quad \forall a < b \in \mathbb{R}.$$

Tato míra se nazývá Lebesgueovou-Stieltjesovou mírou příslušnou distribuční funkci F . Z bodu (iii) plyne, že tato μ je pravděpodobnostní míra. Navíc je rovna rozdělení nějaké náhodné veličiny X , jak tvrdí následující věta.

Věta 2.4 Buď $F : \mathbb{R} \rightarrow \mathbb{R}$ splňující body (i) – (iii) z věty 2.3. Pak existuje pravděpodobnostní prostor (Ω, \mathcal{A}, P) a náhodná veličina $X : (\Omega, \mathcal{A}) \rightarrow (\mathbb{R}, \mathcal{B})$ taková, že $F = F_X$.

Poznámka. Pro náhodnou veličinu X platí

$$P(X \in B) = \int_{\mathbb{R}} \mathbb{1}_B(x) dP_X(x) = \int_{\mathbb{R}} \mathbb{1}_B(x) dF_X(x),$$

kde poslední výraz je Lebesgueův-Stieltjesův integrál podle F_X . Takže distriční funkce F_X určuje rozdělení náhodné veličiny X .

Poznámka. A jak vlastně může vypadat P_X ?

Z teorie míry víme, že každou Lebesgueovu-Stieltjesovu míru $P_X = \mu$ můžeme rozložit na tři části:

$$P_X = \mu = \mu_a + \mu_d + \mu_{sc},$$

kde $\mu_a \ll \lambda$, μ_d je diskrétní míra a μ_{sc} je neatomická míra (tj. $\mu(\{x\}) = 0, \forall x \in \mathbb{R}$) splňující $\mu_{sc} \perp \lambda$.

Definice 2.7 Náhodnou veličinu X nazveme *diskrétní náhodnou veličinou*, pokud existuje (konečná nebo spočetná) posloupnost bodů $\{x_i\}_{i \in I} \subset \mathbb{R}$ a posloupnost čísel $\{p_i\}_{i \in I} \subset (0, 1)$, splňujících $\sum_{i \in I} p_i = 1$, takových, že

$$P_X = \sum_{i \in I} p_i \delta_{x_i}.$$

Odpovídající distribuční funkce tvaru

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

se pak nazývá *diskrétní distribuční funkce*.

Poznámka. Je-li X diskrétní náhodná veličina jako v definici výše a ν čítací míra na $\{x_i\}_{i \in I}$, pak $P_X \ll \nu$ a pro hustotu f_X náhodné veličiny X vzhledem k míře ν platí

$$f_X(x_i) = \begin{cases} p_i, & i \in I, \\ 0, & x \notin \{x_i\}_{i \in I}. \end{cases}$$

*Zde končí
předn. 4
(28.2.)*

Definice 2.8 Náhodnou veličinu X nazveme *absolutně spojitou náhodnou veličinou*, pokud $P_X \ll \lambda$. Odpovídající distribuční funkci F_X se pak nazveme *absolutně spojitou distribuční funkcí*.

Poznámka. Distribuční funkce F_X absolutně spojitě náhodné veličiny má derivaci F'_X λ -skoro všude a pro hustotu f_X náhodné veličiny X vzhledem k λ platí $f_X = F'_X$ λ -s.v. Tedy platí také

$$F_X(x) = \int_{-\infty}^x f_X(z) dz.$$

Definice 2.9 Buď F_X distribuční funkce náhodné veličiny X . Funkce

$$F_X^{-1}(u) = \inf\{x : F_X(x) \geq u\}, \quad u \in (0, 1),$$

se nazývá *kvantilová funkce* náhodné veličiny X .

Poznámka. Kvantilová funkce je neklesající a zleva spojitá. Lze z ní jednoznačně určit distribuční funkci F_X – takže také charakterizuje rozdělení P_X . Pro F_X striktně rostoucí a spojitou je F_X^{-1} inverzní funkcí k F_X .

2.2 MOMENTY NÁHODNÉ VELIČINY

Definice 2.10 *Střední hodnota* náhodné veličiny X je číslo $E X$ dané výrazem

$$E X = \int_{\Omega} X(\omega) dP(\omega),$$

pokud má tento integrál smysl.

Poznámka. Jako $L^1(\Omega, \mathcal{A}, P) = L^1$ značíme prostor všech reálných náhodných veličin na (Ω, \mathcal{A}, P) s konečnou střední hodnotou.

Poznámka. Jak počítat $E X$? Ne nutně z definice – nepotřebuji znát celou míru P , stačí mi P_X , resp. F_X neboť

$$E X = \int_{\Omega} X(\omega) dP(\omega) = \int_{\mathbb{R}} x dP_X(x) = \int_{\mathbb{R}} x dF_X(x),$$

z věty o přenosu integrace pro P_X a z definice Lebesgueove-Stieltjesova integrálu. Speciálně pro diskrétní náhodnou veličinu

$$E X = \sum_{i \in I} x_i p_i = \sum_{i \in I} x_i P(X = x_i),$$

a pro absolutně spojitou náhodnou veličinu

$$E X = \int_{\mathbb{R}} x f_X(x) dx.$$

Tedy $E X$ je vlastnost rozdělení P_X náhodné veličiny X . Pro náhodné veličiny X a Y se stejným rozdělením (resp. stejnou distribuční funkcí) je stejná i střední hodnota.

Definice 2.11 *Medián* rozdělení náhodné veličiny X je číslo $q_{\frac{1}{2}}$ splňující $P(X \leq q_{\frac{1}{2}}) \geq \frac{1}{2}$ a $P(X \geq q_{\frac{1}{2}}) \geq \frac{1}{2}$.

Poznámka. Medián je tedy bod, ve kterém F_X dosáhne (nebo přeskočí) hladinu $\frac{1}{2}$. Nemusí být definován jednoznačně, ale $F_X^{-1}(\frac{1}{2})$ také splňuje definici mediánu.

Věta 2.5 Buď X náhodná veličina a $g : \mathbb{R} \rightarrow \mathbb{R}$ měřitelná funkce. Pak

$$E g(X) = \int_{\mathbb{R}} g(x) dP_X(x) = \int_{\mathbb{R}} g(x) dF_X(x),$$

pokud jeden z nich existuje. Speciálně pro diskrétní n.v.

$$E g(X) = \sum_{i \in I} g(x_i) p_i,$$

a pro absolutně spojitou n.v.

$$E g(X) = \int_{\mathbb{R}} g(x) f_X(x) dx.$$

Věta 2.6 (Vlastnosti střední hodnoty) Buďte X, Y náhodné veličiny a $a, b \in \mathbb{R}$. Pak platí:

- (i) $E(a + bX) = a + bE X, \quad X \in L^1,$
- (ii) $E(X + Y) = E X + E Y, \quad X, Y \in L^1,$
- (iii) $P(X \geq 0) = 1 \Rightarrow E X \geq 0,$
- (iv) $X \in \mathcal{L}^1 \Rightarrow |X| \in L^1,$
- (v) $X \leq Y$ P-s.j. $\Rightarrow E X \leq E Y$ (pokud existují).

Definice 2.12 Buď $n \in \mathbb{N}$.

n-tý moment n.v. X definujeme jako $E X^n$.

n-tý absolutní moment n.v. X definujeme jako $E |X|^n$.

n-tý centrální moment n.v. X definujeme jako $E (X - E X)^n$, existuje-li $E X$.

n-tý absolutní centrální moment n.v. X definujeme jako $E |X - E X|^n$, existuje-li $E X$.

Poznámka. Je-li n -tý moment konečný, píšeme $X \in L^n(\Omega, \mathcal{A}, P)$, resp. $X \in L^n$.

Poznámka. První moment je střední hodnota a pro první centrální moment vždy platí $E (X - E X) = 0$.

Definice 2.13 *Rozptyl* n.v. X je definován jako $E (X - E X)^2$. Značíme $\text{var } X$.

Poznámka. Rozptyl je střední čtvercová odchylka od průměru $E X$ – měřeno/váženo pomocí P_X . Odpovídá tedy na otázku, jak moc jsou hodnoty X rozptýlené kolem $E X$.

Poznámka. Zřejmě je $\text{var } X = E (X - E X)^2 \geq 0$ a $\text{var } X = E (X - E X)^2 = 0$ právě když $X = E X$ P-s.j.

Tvrzení 2.7 Buď X náhodná veličina s konečným rozptylem a $a, b \in \mathbb{R}$. Platí

$$\text{var } (a + bX) = b^2 \text{var } X.$$

Poznámka. Jak počítat rozptyl? Buď přímo:

$$E(X - EX)^2 = \begin{cases} \sum_{i \in I} (x_i - EX)^2 p_i & \text{pro diskrétní n.v.} \\ \int_{\mathbb{R}} (x - EX)^2 f_X(x) dx & \text{pro absolutně spojitou n.v.} \end{cases}$$

Nebo se může hodit přepis

$$E(X - EX)^2 = EX^2 - (EX)^2 = EX(X - 1) + EX - (EX)^2.$$

Věta 2.8 (Čebyševova nerovnost) Buď X n.v. z L^1 a $a > 0$. Pak platí

$$P(|X - EX| \geq a) \leq \frac{\text{var } X}{a^2}.$$

Věta 2.9 (Markovova nerovnost) Buď X n.v. z L^n , $n \in \mathbb{N}$ a $a > 0$. Pak platí

$$P(|X| \geq a) \leq \frac{E|X|^n}{a^n}.$$

Poznámka. Oba výše uvedené odhady jsou velmi hrubé.

*Zde končí
předn. 5
(5.3.)*

Věta 2.10 (Nerovnost mezi L^p normami na pravděpodobnostních prostorech) Buď X náhodná veličina, $0 < \alpha < \beta$ a $E|X|^\beta < \infty$. Pak

$$(E|X|^\alpha)^{\frac{1}{\alpha}} \leq (E|X|^\beta)^{\frac{1}{\beta}},$$

a speciálně $E|X| \leq \sqrt{EX^2}$.

Poznámka. Takže jsme dokázali, že pro pravděpodobnostní míru P_X platí $L^\beta(P_X) \subseteq L^\alpha(P_X)$, $\forall 0 < \alpha < \beta$. Tedy je-li absolutní moment řádu n konečný, tak je konečný i pro každé m splňující $0 < m < n$.

2.3 NĚKTERÁ ROZDĚLENÍ NÁHODNÉ VELIČINY

2.4 ROZDĚLENÍ FUNKCE NÁHODNÉ VELIČINY

*Zde končí
předn. 6
(6.3.)*

Pro řešení teoretických i praktických problémů je dobré umět z rozdělení náhodné veličiny X odvodit i rozdělení ztransformované náhodné veličiny $Y = g(X)$. Pomocí distribuční funkce to jde vcelku snadno.

Věta 2.11 (o rozdělení funkce náhodné veličiny) Buď X náhodná veličina s distribuční funkcí F_X a $g : \mathbb{R} \rightarrow \mathbb{R}$ měřitelná funkce. Pak $Y = g(X)$ je náhodná veličina s distribuční funkcí

$$F_Y(y) = \int_{\{x: g(x) \leq y\}} dF_X(x).$$

Důkaz: Platí

$$F_Y(y) = P(Y \leq t) = P(g(X) \leq t) = P_X(\{x : g(x) \leq t\}) = \int_{\{x: g(x) \leq t\}} dP_X = \int_{\{x: g(x) \leq t\}} dF_X,$$

kde ve třetí rovnosti jsme použili větu 2.1 o přenosu integrace pro P_X . \square

Poznámka. Speciálně pro absolutně spojitou náhodnou veličinu X dostanu

$$F_Y(y) = \int_{\{x: g(x) \leq y\}} f_X(x) dx.$$

Pro diskrétní náhodnou veličinu X dostanu

$$F_Y(y) = \sum_{x_i: g(x_i) \leq y} p_i, \quad (2.1)$$

a zřejmě $P_Y = \sum_{i \in I} p_i \delta_{g(x_i)}$, kde ale Diracovy míry v sumě nemusí být všechny různé.

Poznámka. Pro $X \sim R(0, 1)$ a Y absolutně spojitou náhodnou veličinu vždy existuje taková funkce g , že $g(X)$ má stejné rozdělení jako Y . Toho se využívá například u generátorů náhodných čísel z daného absolutně spojitého rozdělení. Pokud za g volíme kvantilovou funkci F_Y^{-1} , pak $g(X)$ má skutečně rozdělení P_Y (ověřte).

Lineární transformace X se vyskytuje nejčastěji, spočtěme si tedy rozdělení $Y = aX + b$ podrobně:

Příklad. Buď X náhodná veličina a buď Y definovaná jako $Y = aX + b$, kde $a \neq 0$, $b \in \mathbb{R}$. Pak

$$\begin{aligned} F_Y(y) &= P(Y \leq y) = P(aX + b \leq y) \\ &= P(aX \leq y - b) = \begin{cases} P\left(X \leq \frac{y-b}{a}\right) = F_X\left(\frac{y-b}{a}\right), & a > 0 \\ P\left(X \geq \frac{y-b}{a}\right) = 1 - F_X\left(\left(\frac{y-b}{a}\right)_-\right), & a < 0. \end{cases} \end{aligned}$$

Pro F_X absolutně spojitou platí $F_X(x_-) = F_X(x)$, $\forall x \in \mathbb{R}$ a existuje derivace F_X pro sk. vš. x , takže náhodná veličina Y má hustotu vzhledem k λ na \mathbb{R} a platí

$$f_Y(y) = \frac{1}{|a|} f_X\left(\frac{y-b}{a}\right), \quad x \in \mathbb{R}.$$

Uvědomme si, jaký je vztah mezi lineární transformací, momenty a parametry normálního rozdělení:

Příklad. Buď X náhodná veličina s normovaným normálním rozdělením $X \sim N(0, 1)$. A buď Y lineární transformace X , tj. $Y = aX + b$, kde $a \neq 0$, $b \in \mathbb{R}$. Pak podle předchozího příkladu má Y hustotu

$$f_Y(y) = \frac{1}{\sqrt{2\pi a^2}} e^{-\frac{(y-b)^2}{2a^2}}, \quad y \in \mathbb{R},$$

tedy Y má normální rozdělení $N(b, a^2)$.

Pokud je transformace g náhodné veličiny X prostá, dostáváme jednodušší výraz pro distribuční funkci ztransformované náhodné veličiny:

Věta 2.12 (o monotónní transformaci) Buď X náhodná veličina s distribuční funkcí F_X a buď S_X nosič rozdělení P_X (tj. $P(X \in S_X) = 1$). Buď $g : S_X \rightarrow \mathbb{R}$ měřitelná funkce a položme $Y = g(X)$. Pak platí:

- (i) pro g ryze rostoucí $F_Y(y) = F_X(g^{-1}(y))$, pro $y \in g(S_X)$,
- (ii) pro g ryze klesající $F_Y(y) = 1 - F_X(g^{-1}(y)_-)$, pro $y \in g(S_X)$.

Důkaz: Důsledek věty 2.11 o rozdělení funkce náhodné veličiny. □

Pokud je náhodná veličina X absolutně spojitá s hustotou, můžeme odvodit vzorec přímo pro hustotu ztransformované náhodné veličiny:

Důsledek. Buď X absolutně spojitá náhodná veličina s hustotou f_X a $g : S_X \rightarrow \mathbb{R}$ ryze monotónní funkce diferencovatelná λ -s.v. na S_X . Pak náhodná veličina $Y = g(X)$ má hustotu

$$f_Y(y) = f_X(g^{-1}(y)) \left| \frac{dg^{-1}(y)}{dy} \right| \mathbb{1}_{g(S_X)}(y),$$

vzhledem k λ na \mathbb{R} .

I pro diskrétní náhodnou veličinu dochází pro monotónní transformaci ke zjednodušení vzorce (2.1):

Poznámka. Pro X diskrétní náhodnou veličinu a g ryze monotónní na S_X je $P_Y = \sum_{i \in I} p_i \delta_{g(x_i)}$, ale na rozdíl od obecné g jsou nyní míry zahrnuté v sumě navzájem různé.

Vyzkoušíme důsledek věty 2.12 na lineární transformaci:

Příklad. Rozdělení $Y = aX + b$, kde $a \neq 0$, $b \in \mathbb{R}$, je pro absolutně spojitou n.v. X také absolutně spojitá a má podle důsledku výše hustotu

$$f_Y(y) = \frac{1}{|a|} f_X\left(\frac{y-b}{a}\right), \quad x \in \mathbb{R},$$

(jak už jsme zjistili v příkladu za větou 2.11).

3 NÁHODNÉ VEKTORY

3.1 NÁHODNÉ VEKTORY

Pokud chceme pravděpodobnost používat, nevystačíme si typicky s jednou náhodnou veličinou, ale potřebujeme jich mít na (Ω, \mathcal{A}, P) definováno víc najednou. K tomu se dobře hodí pojem náhodného vektoru.

Definice 3.1 Bud' (Ω, \mathcal{A}, P) pravděpodobnostní prostor. Měřitelné zobrazení $X : (\Omega, \mathcal{A}) \rightarrow (\mathbb{R}^n, \mathcal{B}^n)$ nazveme *(n -rozměrným) náhodným vektorem*.

Poznámka. Zřejmě, pokud je X n -rozměrný náhodný vektor a $\Phi : \mathbb{R}^n \rightarrow \mathbb{R}^m$ měřitelné zobrazení, pak je $\Phi(X)$ m -rozměrný náhodný vektor.

Definice 3.2 Rozdělením náhodného vektoru $X : (\Omega, \mathcal{A}) \rightarrow (\mathbb{R}^n, \mathcal{B}^n)$ nazveme indukovanou pravděpodobnostní míru P_X na $(\mathbb{R}^n, \mathcal{B}^n)$ definovanou jako $P_X(B) = P(\{\omega \in \Omega : X(\omega) \in B\})$, $B \in \mathcal{B}^n$.

Poznámka.

- Takže opět P_X je obraz míry P v zobrazení X . Vše je obdobné jako u náhodných veličin.
- Platí, že $X = (X_1, \dots, X_n)^T$ je náhodný vektor právě tehdy, když X_i je náhodná veličina pro každé $i \in \{1, \dots, n\}$ (zopakujte si proč).
- Náhodné vektory budeme uvažovat vždy sloupcové.

I u náhodných vektorů můžeme zavést jednodušší charakterizaci rozdělení P_X náhodného vektoru X . Jak uvidíme dále, všechna informace definující míru P_X je obsažena i v distribuční funkci F_X náhodného vektoru X , tedy funkci z \mathbb{R}^n do \mathbb{R} .

Definice 3.3 (*Sdružená*) distribuční funkce náhodného vektoru X je definována jako

$$F_X(\mathbf{x}) = P\left(\bigcap_{i=1}^n \{X_i \leq x_i\}\right), \quad \mathbf{x} = (x_1, \dots, x_n)^T \in \mathbb{R}^n.$$

Zopakujme si:

Poznámka. Z teorie míry víme, že \mathcal{B}^n je generována:

- (i) měřitelnými obdélníky, tj. systémem $\mathcal{S} = \{B_1 \times \dots \times B_n : B_i \in \mathcal{B}, i = 1, \dots, n\}$,
- (ii) otevřenými intervaly $(\mathbf{a}, \mathbf{b}) = (a_1, b_1) \times \dots \times (a_n, b_n)$, $\mathbf{a}, \mathbf{b} \in \mathbb{R}^n$

- (iii) uzavřenými intervaly $[\mathbf{a}, \mathbf{b}] = [a_1, b_1] \times \cdots \times [a_n, b_n]$, $\mathbf{a}, \mathbf{b} \in \mathbb{R}^n$
- (iv) polouzavřenými intervaly $(\mathbf{a}, \mathbf{b}] = (a_1, b_1] \times \cdots \times (a_n, b_n]$, $\mathbf{a}, \mathbf{b} \in \mathbb{R}^n$
- (v) systémem $\mathcal{S} = \{(-\infty, \mathbf{a}] : \mathbf{a} \in \mathbb{R}^n\}$ (značíme $(-\infty, \mathbf{a}] = (-\infty, a_1] \times \cdots \times (-\infty, a_n]$).

Poznámka. Systém $\mathcal{S} = \{(-\infty, \mathbf{a}] : \mathbf{a} \in \mathbb{R}^n\}$ je uzavřený na konečné průniky a tak z věty o jednoznačnosti míry a předchozí poznámky máme, že $F_{\mathbf{X}}$ jednoznačně určuje $P_{\mathbf{X}}$. Tedy $F_{\mathbf{X}}$ opravdu je ekvivalentní charakterizace rozdělení \mathbf{X} .

A jak z $P_{\mathbf{X}}$ nebo $F_{\mathbf{X}}$ získat rozdělení podvektorů, nebo jednotlivých v \mathbf{X} zahrnutých náhodných veličin?

Věta 3.1 (o marginální distribuční funkci) Buď \mathbf{X} n -rozměrný náhodný vektor s distribuční funkcí $F_{\mathbf{X}}$. Pak pro každé $\mathbf{x} \in \mathbb{R}^n$ platí

$$\lim_{x_n \rightarrow \infty} F_{\mathbf{X}}(x_1, \dots, x_n) = F_{(X_1, \dots, X_{n-1})^T}(x_1, \dots, x_{n-1}),$$

kde $F_{(X_1, \dots, X_{n-1})^T}$ je distribuční funkce $(n-1)$ -rozměrného náhodného vektoru $(X_1, \dots, X_{n-1})^T$.

Důkaz: Využijeme Heineho větu. Buď $\{y_k\}_{k=1}^{\infty}$ libovolná posloupnost pro kterou $\lim_{k \rightarrow \infty} y_k = \infty$. Označme si

$$B = \bigcap_{i=1}^{n-1} \{X_i \leq x_i\}, \quad B_k = \left(\bigcap_{i=1}^{n-1} \{X_i \leq x_i\} \right) \cap \{X_n \leq y_k\}, \quad D_k = \left(\bigcup_{l=k}^{\infty} B_l^c \right)^c, \quad \text{pro } k \in \mathbb{N}.$$

Zřejmě $D_k \subseteq B_k \subseteq B = \bigcup_{k=1}^{\infty} B_k$ a $D_k \nearrow B$. Ze spojitosti míry P tedy máme $\lim_{k \rightarrow \infty} P(D_k) = P(B)$. A z monotonie P máme $P(D_k) \leq P(B_k) \leq P(B)$, tedy i $\lim_{k \rightarrow \infty} P(B_k) = P(B)$, což jsme měli dokázat. \square

Poznámka. Z definice distribuční funkce zřejmě

$$F_{\mathbf{X}}(x_1, \dots, x_n) = F_{(X_{\pi(1)}, \dots, X_{\pi(n)})^T}(x_{\pi(1)}, \dots, x_{\pi(n)}), \quad \mathbf{x} \in \mathbb{R}^n,$$

pro každou permutaci $\pi : \{1, \dots, n\} \rightarrow \{1, \dots, n\}$. Takže z předchozí věty získám limitními přechody z $F_{\mathbf{X}}$ distribuční funkci libovolného podvektoru \mathbf{X} .

Takže opravdu umíme z $F_{\mathbf{X}}$ získat distribuční funkci libovolného podvektoru \mathbf{X} .

Poznámka. Uvědomme si, že platí $\sigma((X_1, \dots, X_{n-1})^T) \subseteq \sigma(\mathbf{X}) \subseteq \mathcal{A}$.

A jak získat z $P_{\mathbf{X}}$ marginální rozdělení podvektoru?

Poznámka. Rozdělení $P_{\mathbf{Y}}$ podvektoru $\mathbf{Y} = (X_j)_{j \in J}$, $J \subset \{1, \dots, n\} = I$, se nazývá *marginální rozdělení*. Často $|J| = 1$, pak je Y náhodná veličina. Odvození $P_{\mathbf{Y}}$ z $P_{\mathbf{X}}$ je zřejmé, neboť

$$P_{\mathbf{Y}}(B) = P(\{\omega \in \Omega : \mathbf{Y}(\omega) \in B, (X_k)_{k \in I \setminus J}(\omega) \in \mathbb{R}^{n-|J|}\}), \quad B \in \mathcal{B}^{|J|}.$$

Speciálně $P_{\mathbf{Y}}(B) = P_{\mathbf{X}}(B \times \mathbb{R}^{n-m})$, pro $J = \{1, \dots, m\}$.

Bud' $\mathbf{X} = (X_1, X_2)^T$ náhodný vektor. Z předchozí poznámky vidíme, že $P_{\mathbf{X}}$ určuje P_{X_1} a P_{X_2} . Naopak to ale neplatí! Uvažme následující příklad:

Příklad. Uvažujme situaci hodu třemi spravedlivými kostkami a sledujme paritu výsledků – počet sudých čísel na kostkách. Označme X náhodnou veličinu určující počet sudých čísel na kostce číslo 1, Y náhodnou veličinu určující počet sudých čísel dohromady na kostkách číslo 2 a 3. Pak $P_{(X,Y)^T}$ je diskrétní míra s nosičem $\{0, 1\} \times \{0, 1, 2\}$ a můžeme ji popsat tabulkou pravděpodobností pro dané hodnoty:

$X \setminus Y$	0	1	2	Σ
0	$\frac{1}{8}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{2}$
1	$\frac{1}{8}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{2}$
Σ	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$	1

Zřejmě marginální rozdělení X je $P_X = \frac{1}{2}\delta_0 + \frac{1}{2}\delta_1$ a pro Y je $P_Y = \frac{1}{4}\delta_0 + \frac{1}{2}\delta_1 + \frac{1}{4}\delta_2$.

Uvažujme teď X stejné, ale Y bud' náhodná veličina popisující počet sudých čísel dohromady na kostkách číslo 1 a 2. Opět je $P_{(X,Y)^T}$ diskrétní míra s nosičem $\{0, 1\} \times \{0, 1, 2\}$ a můžeme ji popsat tabulkou pravděpodobností pro dané hodnoty:

$X \setminus Y$	0	1	2	Σ
0	$\frac{1}{4}$	$\frac{1}{4}$	0	$\frac{1}{2}$
1	0	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{2}$
Σ	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$	1

Tedy marginální rozdělení jsou stejná jako v předchozím případě, ovšem sdružené rozdělení je různé.

Pokud známe marginální rozdělení veličin zahrnutých v náhodném vektoru \mathbf{X} , známe vlastně projekce míry $P_{\mathbf{X}}$ do jednotlivých souřadnic. Víme z teorie míry, že tyto projekce určují celou $P_{\mathbf{X}}$ jen v některých velmi speciálních případech (viz kapitola 3.2), obecně ale ne.

Vraťme se ale nyní k otázce, jestli dokážeme poznat, kdy je funkce F distribuční funkcí nějakého náhodného vektoru. A zda existuje sada požadavků na F , které už zaručí, že musí existovat náhodný vektor \mathbf{X} s touto distribuční funkcí. To jest, jestli máme obdobnou situaci jako u náhodných veličin.

Značení. Mějme $(\mathbf{a}, \mathbf{b}]$ neprázdný interval v \mathbb{R}^n , značíme $\Delta_{n,k}$ množinu bodů $\mathbf{c} \in \mathbb{R}^n$ takových, že $c_i \in \{a_i, b_i\} \forall i \in \{1, \dots, n\}$ a $|\{i : c_i = a_i\}| = k$ (tedy \mathbf{c} se rovná \mathbf{a} přesně v k souřadnicích).

Věta 3.2 (o vlastnostech sdružené distribuční funkce) Distribuční funkce $F_{\mathbf{X}}$ náhodného vektoru \mathbf{X} splňuje:

- (i) $\lim_{x_i \rightarrow \infty, 1 \leq i \leq n} F_{\mathbf{X}}(x_1, \dots, x_n) = 1;$
- (ii) $\lim_{x_j \rightarrow -\infty} F_{\mathbf{X}}(x_1, \dots, x_n) = 0, \forall j \in \{1, \dots, n\} \forall x_1, \dots, x_{j-1}, x_{j+1}, \dots, x_n \in \mathbb{R};$
- (iii) $F_{\mathbf{X}}$ je zprava spojitá v každé proměnné;
- (iv) pro $(\mathbf{a}, \mathbf{b}]$ neprázdný interval v \mathbb{R}^n platí

$$\sum_{k=0}^n (-1)^k \sum_{\mathbf{c} \in \Delta_{n,k}} F_{\mathbf{X}}(\mathbf{c}) \geq 0.$$

Důkaz: Body (i)–(iii) jsou zřejmé z definice $F_{\mathbf{X}}$ a ze spojitosti míry P .

Abychom dokázali (iv), je třeba dokázat, že levá strana nerovnosti je rovna $P(\mathbf{X} \in (\mathbf{a}, \mathbf{b}])$, což je nutně nezáporné. Podrobný důkaz lze najít např. v [Dupač and Hušková \(2013\)](#) str. 42 (je vcelku technický). My si zde odvodíme rovnost jen pro $n = 2$ (může být užitečné si namalovat obrázek). V tom případě

$$\begin{aligned} \sum_{k=0}^2 (-1)^k \sum_{\mathbf{c} \in \Delta_{2,k}} F_{\mathbf{X}}(\mathbf{c}) &= F_{\mathbf{X}}(b_1, b_2) - [F_{\mathbf{X}}(b_1, a_2) + F_{\mathbf{X}}(a_1, b_2)] + F_{\mathbf{X}}(a_1, a_2) \\ &= [F_{\mathbf{X}}(b_1, b_2) - F_{\mathbf{X}}(b_1, a_2)] - [F_{\mathbf{X}}(a_1, b_2) - F_{\mathbf{X}}(a_1, a_2)] \\ &= P(X_1 \leq b_1, a_2 < X_2 \leq b_2) - P(X_1 \leq a_1, a_2 < X_2 \leq b_2) \\ &= P(a_1 < X_1 \leq b_1, a_2 < X_2 \leq b_2) = P(\mathbf{X} \in (\mathbf{a}, \mathbf{b}]). \end{aligned}$$

□

Poznámka. Mějme funkci $F : \mathbb{R}^n \rightarrow [0, 1]$ splňující body (i)–(iv) z věty 3.2. Pro každý neprázdný interval $(\mathbf{a}, \mathbf{b}]$ v \mathbb{R}^n definujeme

$$\mu_F((\mathbf{a}, \mathbf{b}]) = \sum_{k=0}^n (-1)^k \sum_{\mathbf{c} \in \Delta_{n,k}} F(\mathbf{c}).$$

Potom lze množinovou funkci μ_F na intervalech jednoznačně rozšířit na pravděpodobnostní borelovskou míru μ_F na \mathbb{R}^n . Tato míra se nazývá Lebesgue-Stieltjesova míra příslušná funkci F . (Bez důkazu, viz teorie míry).

Pokud byla $F = F_{\mathbf{X}}$ distribuční funkce nějakého náhodného vektoru, pak nutně Lebesgue-Stieltjesova míra $\mu_{F_{\mathbf{X}}}$ je shodná s rozdělením $P_{\mathbf{X}}$ vektoru \mathbf{X} .

Věta 3.3 Nechť $F : \mathbb{R}^n \rightarrow [0, 1]$ splňuje body (i)–(iv) z věty 3.2. Pak existuje pravděpodobnostní prostor (Ω, \mathcal{A}, P) a náhodný vektor $\mathbf{X} : \Omega \rightarrow \mathbb{R}^n$ takový, že $F_{\mathbf{X}} = F$.

Důkaz: Položme $\Omega = \mathbb{R}^n$, $\mathcal{A} = \mathcal{B}^n$, $P = \mu_F$ a $\mathbf{X}(\omega) = \omega$, $\omega \in \Omega$. Pak \mathbf{X} je zřejmě měřitelné a platí

$$F_{\mathbf{X}}(\mathbf{x}) = P(\{\omega : \omega_i \leq x_i, i \in \{1, \dots, n\}\}) = \mu_F((-\infty, \mathbf{x}]) = F(\mathbf{x}), \quad \forall \mathbf{x} \in \mathbb{R}^n$$

a tedy $\mu_F = P_{\mathbf{X}}$. □

Poznámka. Důsledkem věty (resp. jejího důkazu) je, že pokud máme určeno rozdělení náhodného vektoru, ať už Lebesgueovou-Stieltjesovou mírou μ_F nebo distribuční funkcí F splňující body (i)–(iv) z věty 3.2, tak vždy víme, že existuje nějaký pravděpodobnostní prostor (Ω, \mathcal{A}, P) a nějaký náhodný vektor $\mathbf{X} : \Omega \rightarrow \mathbb{R}^n$, který má přesně toto rozdělení.

I pro náhodné vektory rozlišíme dva speciální případy $P_{\mathbf{X}}$:

Definice 3.4 Náhodný vektor \mathbf{X} má *diskrétní rozdělení*, pokud existuje (konečná nebo spočetná) posloupnost bodů $\{\mathbf{x}_j\}_{j \in I} \subset \mathbb{R}^n$ a posloupnost čísel $\{p_j\}_{j \in I} \subset (0, 1)$, splňujících $\sum_{j \in I} p_j = 1$, takových, že

$$P_{\mathbf{X}} = \sum_{j \in I} p_j \delta_{\mathbf{x}_j}.$$

Odpovídající distribuční funkce $F_{\mathbf{X}}$ se pak nazývá *diskrétní distribuční funkce*.

Poznámka. Každé marginální rozdělení podvektoru vektoru \mathbf{X} je také diskrétní (ověřte).

Definice 3.5 Náhodný vektor \mathbf{X} má *absolutně spojitě rozdělení*, pokud existuje nezáporná měřitelná funkce $f_{\mathbf{X}}$ splňující

$$F_{\mathbf{X}}(\mathbf{x}) = \int_{-\infty}^{x_1} \dots \int_{-\infty}^{x_n} f_{\mathbf{X}}(t_1, \dots, t_n) dt_n \dots dt_1, \quad \forall \mathbf{x} \in \mathbb{R}^n.$$

Funkce $f_{\mathbf{X}}$ se nazývá *hustota rozdělení \mathbf{X}* . Odpovídající distribuční funkce $F_{\mathbf{X}}$ se pak nazývá *absolutně spojitá distribuční funkce*.

Poznámka. \mathbf{X} má *absolutně spojitě rozdělení* právě tehdy když $P_{\mathbf{X}} \ll \lambda^n$ (ověřte). Potom $f_{\mathbf{X}} = \frac{\partial^n}{\partial x_1 \dots \partial x_n} F_{\mathbf{X}} \lambda^n$ -s.v.

Poznámka. Každé marginální rozdělení podvektoru absolutně spojitěho náhodného vektoru \mathbf{X} je také absolutně spojitě (ověřte).

Vlastnost diskrétního rozdělení, resp. absolutně spojitěho rozdělení, pro marginální rozdělení nevynucuje takovou vlastnost pro sdružené rozdělení - umíte najít (proti)příklad?

Otázka. Umíte najít náhodný vektor (X_1, X_2) s absolutně spojitými marginálními rozděleními, který není absolutně spojitý?

Umíte najít náhodný vektor (X_1, X_2) s diskrétními marginálními rozděleními, který není diskrétní?

Rozdělení P_X náhodného vektoru (pravděpodobnostní míra na \mathbb{R}^n) může být tedy o dost složitější než rozdělení P_X náhodné veličiny (pravděpodobnostní míra na \mathbb{R}) – u té jsme měli úplný popis (viz poznámka za větou 2.4).

Jak jinak tedy ještě může vypadat P_X náhodného vektoru? – Různě. Dobře popsatelný je ještě případ, kdy P_X je absolutně spojitá vzhledem k nějaké součinové míře na \mathbb{R}^n .

Věta 3.4 (o hustotě P_X vzhledem k součinové referenční míře) Buď P_X rozdělení n -rozměrného náhodného vektoru \mathbf{X} . Nechť $P_X \ll \nu_1 \otimes \cdots \otimes \nu_n$ (součin σ -konečných měr na \mathbb{R}). Pak $P_{X_i} \ll \nu_i$, $\forall i \in \{1, \dots, n\}$ a existují nezáporné měřitelné funkce $f_X : \mathbb{R}^n \rightarrow [0, \infty)$, $f_i : \mathbb{R} \rightarrow [0, \infty)$, $i \in \{1, \dots, n\}$ takové, že:

$$P_X((-\infty, \mathbf{x}]) = F_X(\mathbf{x}) = \int_{(-\infty, \mathbf{x}]} f_X(t_1, \dots, t_n) d\nu_1 \otimes \cdots \otimes \nu_n(\mathbf{t}), \quad \forall \mathbf{x} \in \mathbb{R}^n,$$

a

$$P_X(B) = \int_B f_X(\mathbf{t}) d\nu_1 \otimes \cdots \otimes \nu_n(\mathbf{t}), \quad \forall B \in \mathcal{B}^n.$$

Navíc pro každé X_i , $i \in \{1, \dots, n\}$ platí

$$F_{X_i}(x_i) = \int_{-\infty}^{x_i} f_i(t) d\nu_i(t), \quad \forall x_i \in \mathbb{R},$$

kde

$$f_i(y_i) = \int_{\mathbb{R}^{n-1}} f_X(y_1, \dots, y_n) d\nu_1 \otimes \cdots \otimes \nu_{i-1} \otimes \nu_{i+1} \otimes \cdots \otimes \nu_n(y_1, \dots, y_{i-1}, y_{i+1}, \dots, y_n),$$

pro ν_i -s.v. $y_i \in \mathbb{R}$.

Důkaz: Existence a tvrzení o f_X plyne z Radon-Nikodýmovy věty.

Existence a tvrzení o f_i plyne z Fubiniovy věty a věty 3.1 o marginální distribuční funkci. \square

Poznámka. Zřejmě je funkce f_i rovna $\frac{dP_{X_i}}{d\nu_i}$ a je to marginální hustota P_{X_i} vzhledem k ν_i .

Věta dává návod, jak odvodit marginální hustotu, resp. marginální pravděpodobnosti pro speciální případ $\nu_1 \otimes \cdots \otimes \nu_n = \lambda^n$, resp. součin čítacích měr.

Otázka. Umíte najít nějaký náhodný vektor \mathbf{X} , aby nebylo možné $P_X \ll \nu_1 \otimes \cdots \otimes \nu_n$?

*Zde končí
předn. 8
(13.3.)*

3.2 NEZÁVISLÉ NÁHODNÉ VELIČINY

V této kapitole podrobněji prozkoumáme situaci, kdy je rozdělení P_X náhodného vektoru \mathbf{X} ve speciálním součinném tvaru – tj. $P_X = P_{X_1} \otimes \cdots \otimes P_{X_n}$. Z teorie míry víme, že v takovém případě P_X přesně odpovídá součinu svých projekcí. A v pravděpodobnosti

tento případ odpovídá situaci, kdy rozdělení jednotlivých marginálních veličin X_i z \mathbf{X} nenese v sobě žádnou informaci o rozdělení ostatních marginálů X_j , $j \neq i$ z \mathbf{X} . V pravděpodobnosti to tedy odpovídá případu, kdy $\{X_i\}$ jsou (stochasticky) nezávislé. Je to ten nejjednodušší případ závislosti náhodných veličin – nezávislost. A jak ji budeme definovat? Zatím jsme měli jen nezávislost pro náhodné jevy. A samozřejmě chceme, aby byla naše nová definice kompatibilní s tou starou.

Definice 3.6 Bud' $\{X_i, i \in I\}$ systém náhodných veličin na (Ω, \mathcal{A}, P) , kde $I \neq \emptyset$ je libovolná indexová množina. $\{X_i, i \in I\}$ nazveme (vzájemně) nezávislé, pokud pro každou konečnou $J \subset I$ platí

$$P\left(\bigcap_{i \in J} \{X_i \in B_i\}\right) = \prod_{i \in J} P(X_i \in B_i), \quad \forall B_i \in \mathcal{B}, i \in J.$$

Poznámka. Pro nezávislé náhodné veličiny jsou jevy $\{X_i \in B_i\}$ vzájemně nezávislé ve smyslu definice 1.4 nezávislosti pro náhodné jevy.

Poznámka. Pokud je $|I| < \infty$, pak stačí ověřit $P(\bigcap_{i \in I} \{X_i \in B_i\}) = \prod_{i \in I} P(X_i \in B_i)$, $\forall B_i \in \mathcal{B}, i \in I$, neboť platnost ostatních podmínek z definice 3.6 pro $J \subset I$ dostaneme dosazením \mathbb{R} za B_i pro $i \in I \setminus J$.

Odpovídá naše definice nezávislých náhodných veličin tomu, co jsme chtěli? Tedy součinné míře pro $P_{\mathbf{X}}$? Ano:

Věta 3.5 (o rozdělení vektoru s nezávislými složkami) Bud' $\mathbf{X} = (X_1, \dots, X_n)^T$ náhodný vektor. Pak $\{X_i\}_{i=1}^n$ jsou nezávislé náhodné veličiny právě tehdy, když

$$P_{\mathbf{X}} = P_{X_1} \otimes \dots \otimes P_{X_n}.$$

Důkaz: Pro všechny množiny $B_1 \times \dots \times B_n$, $B_i \in \mathcal{B}, i \in \{1, \dots, n\}$ platí

$$P_{\mathbf{X}}(B_1 \times \dots \times B_n) = \prod_{i=1}^n P_{X_i}(B_i)$$

z definice nezávislosti. Tedy $P_{\mathbf{X}}$ se rovná součinné míře na měřitelných obdélnících, což je systém množin uzavřený na průniky a generující \mathcal{B}^n . Protože je $P_{\mathbf{X}}(\mathbb{R}^n) = 1$ konečná, dostáváme z věty o jednoznačnosti míry rovnost obou měr na celém \mathcal{B}^n . \square

A jak poznat nezávislost složek náhodného vektoru pomocí distribuční funkce $F_{\mathbf{X}}$?

Věta 3.6 (o distribuční funkci vektoru s nezávislými složkami) Bud' $\mathbf{X} = (X_1, \dots, X_n)^T$ náhodný vektor. Pak $\{X_i\}_{i=1}^n$ jsou nezávislé náhodné veličiny právě tehdy, když

$$F_{\mathbf{X}}(\mathbf{x}) = \prod_{i=1}^n F_{X_i}(x_i), \quad \forall \mathbf{x} \in \mathbb{R}^n. \quad (3.1)$$

Důkaz: Pokud jsou veličiny $\{X_i\}_{i=1}^n$ nezávislé, pak

$$F_{\mathbf{X}}(x_1, \dots, x_n) = P\left(\bigcap_{i=1}^n \{X_i \in (-\infty, x_i]\}\right) = \prod_{i=1}^n P(X_i \in (-\infty, x_i]) = \prod_{i=1}^n F_{X_i}(x_i),$$

kde druhá rovnost plyne z definice 3.6.

Obráceně, pokud platí (3.1), tak $P_{\mathbf{X}}$ se rovná součinové míře $P_{X_1} \otimes \dots \otimes P_{X_n}$ na množinách ze systému $\mathcal{S} = \{(-\infty, \mathbf{x}] : \mathbf{x} \in \mathbb{R}^n\}$. \mathcal{S} je uzavřený na průniky a generuje \mathcal{B}^n . Protože je $P_{\mathbf{X}}(\mathbb{R}^n) = 1$ konečná, dostáváme z věty o jednoznačnosti míry rovnost obou měr na celém \mathcal{B}^n a tedy z věty 3.5 nezávislost $\{X_i\}_{i=1}^n$. \square

Poznámka. Tedy pro $\{X_i\}_{i=1}^n$ nezávislé už marginální rozdělení (resp. marginální distribuční funkce) jednoznačně určují sdružené rozdělení $P_{\mathbf{X}}$, resp. sdruženou distribuční funkci $F_{\mathbf{X}}$.

V případě, že je náhodný vektor \mathbf{X} absolutně spojitý, lze vzájemnou nezávislost jeho složek poznat i na hustotě $f_{\mathbf{X}}$:

Věta 3.7 (o hustotě náhodného vektoru s nezávislými složkami) Bud' $P_{\mathbf{X}}$ rozdělení n -rozměrného náhodného vektoru \mathbf{X} splňující $P_{\mathbf{X}} \ll \nu_1 \otimes \dots \otimes \nu_n$ (součin σ -konečných měr na \mathbb{R}), a $f_{\mathbf{X}}$ jeho hustota. Pak náhodné veličiny $\{X_i\}_{i=1}^n$ jsou nezávislé právě tehdy, když

$$f_{\mathbf{X}}(x_1, \dots, x_n) = \prod_{i=1}^n f_{X_i}(x_i), \quad \text{pro } \nu_1 \otimes \dots \otimes \nu_n - \text{s.v. } \mathbf{x} \in \mathbb{R}^n,$$

kde $f_{X_i} = \frac{dP_{X_i}}{d\nu_i}$, $i \in \{1, \dots, n\}$.

Důkaz: " \Rightarrow " Z věty 3.4 o hustotě $P_{\mathbf{X}}$ vzhledem k součinové referenční míře máme

$$F_{\mathbf{X}}(\mathbf{x}) = \int_{(-\infty, \mathbf{x}]} f_{\mathbf{X}}(t_1, \dots, t_n) d\nu_1 \otimes \dots \otimes \nu_n(\mathbf{t}), \quad \forall \mathbf{x} \in \mathbb{R}^n.$$

Z nezávislosti $\{X_i\}_{i=1}^n$ plyne podle věty 3.6 $F_{\mathbf{X}}(\mathbf{x}) = \prod_{i=1}^n F_{X_i}(x_i)$ pro všechna $\mathbf{x} \in \mathbb{R}^n$, a tedy

$$\begin{aligned} F_{\mathbf{X}}(\mathbf{x}) &= \prod_{i=1}^n F_{X_i}(x_i) = \prod_{i=1}^n \int_{-\infty}^{x_i} f_{X_i}(t_i) d\nu_i(t_i) = \int_{-\infty}^{x_1} \dots \int_{-\infty}^{x_n} \prod_{i=1}^n f_{X_i}(t_i) d\nu_n(t_n) \dots d\nu_1(t_1) \\ &= \int_{(-\infty, \mathbf{x}]} \prod_{i=1}^n f_{X_i}(t_i) d\nu_1 \otimes \dots \otimes \nu_n(\mathbf{t}), \end{aligned}$$

kde ve třetí rovnosti jsme použili linearitu integrálu a ve čtvrté rovnosti Fubiniovu větu. Kombinací obou vyjádření $F_{\mathbf{X}}$ dostaneme, že $f_{\mathbf{X}}(t_1, \dots, t_n) = \prod_{i=1}^n f_{X_i}(t_i)$ pro $\nu_1 \otimes \dots \otimes \nu_n$ -s.v. $\mathbf{t} \in \mathbb{R}^n$.

" \Leftarrow " Dokážeme obdobně obráceným postupem. \square

Poznámka. Pro \mathbf{X} absolutně spojitý, resp. diskrétní, $\nu_1 \otimes \cdots \otimes \nu_n$ -s.v. znamená λ^n -s.v. resp. s.v. vzhledem k součinu čítacích měr (konkrétní aplikace viz cvičení).

A co se stane, když máme náhodný vektor \mathbf{X} s nezávislými marginálními rozděleními a ztransformujeme ho na jiný náhodný vektor? Musí se zachovat nezávislost marginálních rozdělení? Typicky ne, ale v jednom speciálním (ovšem vcelku častém) případě ano:

Věta 3.8 Buď $\{X_i, i \in I\}$ systém nezávislých náhodných veličin a $g_i : \mathbb{R} \rightarrow \mathbb{R}, i \in I$ měřitelné funkce. Pak $\{g_i(X_i), i \in I\}$ je opět systém nezávislých náhodných veličin.

Důkaz: Buď $J \subset I$ konečná. Pak pro libovolné $B_i \in \mathcal{B}, i \in J$ platí

$$P\left(\bigcap_{i \in J} \{g_i(X_i) \in B_i\}\right) = P\left(\bigcap_{i \in J} \{X_i \in g_i^{-1}(B_i)\}\right) = \prod_{i \in J} P(X_i \in g_i^{-1}(B_i)) = \prod_{i \in J} P(g_i(X_i) \in B_i),$$

kde v druhé rovnosti jsme použili nezávislost $\{X_i, i \in I\}$. Tedy i $\{g_i(X_i), i \in I\}$ je systém nezávislých náhodných veličin. \square

Poznámka. Věta mluví o tom, že zobrazení náhodného vektoru „po složkách“ zachovává nezávislost.

Příklad. Buď $(X, Y)^T$ náhodný vektor s nezávislými složkami. Pak je i $(X^2, Y^2)^T$ náhodný vektor s nezávislými složkami. Ale např. $(X + Y, X - Y)^T$ už nemusí mít nezávislé marginály - viz kapitola 3.4.

3.3 MOMENTY NÁHODNÉHO VEKTORU

V této kapitole prozkoumáme momenty náhodného vektoru. Využijeme to, co už víme pro náhodné veličiny, ale přibude hodně informací o tzv. smíšených momentech, tedy středních hodnotách ze součinů různých marginálních veličin vektoru \mathbf{X} , a jejich vlastnostech a významu.

Střední hodnota pro náhodný vektor \mathbf{X} není nic nového, neboť ji definujeme po složkách:

Značení. Budeme značit $E \mathbf{X} = (E X_1, \dots, E X_n)^T$.

A jak to vypadá se střední hodnotou náhodného vektoru ve vztahu k transformaci náhodného vektoru?

Tvrzení 3.9 Buď \mathbf{X} náhodný vektor a $g : \mathbb{R}^n \rightarrow \mathbb{R}$ měřitelná funkce. Pak je $g(\mathbf{X})$ náhodná veličina a pro $P_{\mathbf{X}} \ll \nu_1 \otimes \cdots \otimes \nu_n$ platí

$$E g(\mathbf{X}) = \int_{\mathbb{R}} y dP_{g(\mathbf{X})} = \int_{\mathbb{R}^n} g(\mathbf{x}) dP_{\mathbf{X}}(\mathbf{x}) = \int_{\mathbb{R}^n} g(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x}) d\nu_1 \otimes \cdots \otimes \nu_n(\mathbf{x}).$$

Důkaz: Složení měřitelných funkcí je zřejmě měřitelné, první a druhá rovnost plynou z věty o přenosu integrace, třetí z Radon-Nikodýmovy věty. \square

Poznámka. Důležité na předchozím pozorování je, že k výpočtu $EY = E g(\mathbf{X})$ není potřeba znát rozdělení náhodné veličiny Y , stačí znát $P_{\mathbf{X}}$.

A nyní postoupíme k definici momentů pro kombinace různých náhodných veličin. Začneme dvěma náhodnými veličinami a kovariancí:

Definice 3.7 Buďte X, Y náhodné veličiny na tomtéž pravděpodobnostním prostoru. Pak *kovariance* X a Y je definována jako $\text{cov}(X, Y) = E[(X - EX)(Y - EY)]$.

Koeficient korelace je definován jako

$$\text{cor}(X, Y) = \frac{\text{cov}(X, Y)}{\sqrt{\text{var } X \text{var } Y}},$$

pokud $\text{var } X > 0$ a $\text{var } Y > 0$.

Poznámka. Ekvivalentní předpoklad k tomu, že X, Y jsou náhodné veličiny na tomtéž pravděpodobnostním prostoru, je, že $(X, Y)^T$ je náhodný vektor.

Poznámka. Kovariance splňuje $\text{cov}(aX + b, cY + d) = ac \text{cov}(X, Y)$, $\forall a, b, c, d \in \mathbb{R}$ (ověřte).

Abychom byli schopni odvodit vztahy mezi různými momenty náhodného vektoru, bude se nám hodit z teorie míry Hölderova nerovnost. Pro jistotu ji zopakujeme, už ve znění pro náhodné veličiny:

Věta 3.10 (Hölderova nerovnost) Buďte X_1, X_2 náhodné veličiny na tomtéž pravděpodobnostním prostoru, a necht' $E|X_1|^p \leq \infty$, $E|X_2|^q \leq \infty$, kde $p, q \geq 1$, $\frac{1}{p} + \frac{1}{q} = 1$. Potom

$$E|X_1 X_2| \leq (E|X_1|^p)^{\frac{1}{p}} (E|X_2|^q)^{\frac{1}{q}},$$

a rovnost nastává právě tehdy, když existují $a, b \in [0, \infty)$ (alespoň jedno z nich nenulové) takové, že $a|X_1|^p = b|X_2|^q$ P-skoro jistě.

Důkaz: Viz teorie míry a integrálu. \square

A co nám Hölderova nerovnost říká o kovarianci?

Poznámka. Speciálně pro $p = q = 2$ dostaneme

$$E|X_1 X_2| \leq \sqrt{E X_1^2 E X_2^2},$$

takže

$$|\text{cov}(X, Y)| \leq E|(X - EX)(Y - EY)| \leq \sqrt{E(X - EX)^2 E(Y - EY)^2} = \sqrt{\text{var}(X)\text{var}(Y)},$$

a $\text{cor}(X, Y) \in [-1, 1]$.

Navíc $|\text{cor}(X, Y)| = 1$ právě tehdy, když $X = aY + b$ P-skoro jistě, pro nějaké $a \neq 0$, $b \in \mathbb{R}$ (ověřte).

Poznámka. Rovněž z Hölderovy nerovnosti dostaneme, že pokud $E X^2 < \infty$ a $E Y^2 < \infty$, pak $\text{cov}(X, Y) \in \mathbb{R}$ (ověřte).

Poznámka. Náhodnou veličinu X nazveme *degenerovanou*, pokud existuje $c \in \mathbb{R}$ takové, že $X = c$ P-skoro jistě. Neboli $P_X = \delta_c$ a X je tedy skoro jistě „nenáhodná“. Pokud je X nebo Y degenerovaná náhodná veličina, tak také platí $\text{cov}(X, Y) = \sqrt{\text{var}(X)\text{var}(Y)}$, ale koeficient korelace není definován (ověřte).

A co nám říká kovariance o vlastnostech rozdělení vektoru $(X, Y)^T$? Heuristicky řečeno mluví o tom, jak moc jsou v průměru (váženo pomocí $P_{(X,Y)^T}$) podobné/jak se liší odchylky $(X - EX)(\omega)$ a $(Y - EY)(\omega)$ pro obě náhodné veličiny.

Poznámka. Zřejmě platí $\text{cov}(X, X) = \text{var } X$, $\text{cor}(X, X) = 1$, $\text{cor}(X, -X) = -1$. Kovariance je velká, pokud mají odchylky $(X - EX)(\omega)$ a $(Y - EY)(\omega)$ tendenci „jít stejným směrem“. Někdy se říká, že korelace $\text{cor}(X, Y)$ „měří míru lineární závislosti“. Rozmyslete si to lépe na následujícím příkladě.

Příklad. Buď $(X, Y)^T$ náhodný vektor s rovnoměrným rozdělením na jednotkovém kruhu v \mathbb{R}^2 . Pak $\text{cov}(X, Y) = 0$. Náhodný vektor $(X, X)^T$ má $\text{cor}(X, X) = 1$ a $P_{(X,X)^T}$ má nosič na hlavní diagonále v \mathbb{R}^2 . Náhodný vektor $(X, -X)^T$ má $\text{cor}(X, -X) = -1$ a $P_{(X,-X)^T}$ má nosič na vedlejší diagonále v \mathbb{R}^2 .

A teď si rozmyslíme, jak vypaří smíšené momenty náhodného vektoru, když má ten vektor nezávislé složky. Uvidíme, že mnoho věcí se zjednoduší. Začneme $E X_1 X_2$:

Věta 3.11 Buďte X_1, X_2 nezávislé náhodné veličiny a buď $E |X_i| < \infty, i = 1, 2$. Pak $E |X_1 X_2| < \infty$ a platí $E X_1 X_2 = E X_1 E X_2$.

Důkaz: Pokud je $E X_1 X_2$ definována, pak platí

$$\begin{aligned} E X_1 X_2 &= \int_{\mathbb{R}^2} x_1 x_2 dP_{\mathbf{X}}(\mathbf{x}) = \int_{\mathbb{R}^2} x_1 x_2 dP_{X_1} \otimes P_{X_2}(\mathbf{x}) = \int_{\mathbb{R}} \int_{\mathbb{R}} x_1 x_2 dP_{X_1}(x_1) dP_{X_2}(x_2) \\ &= \int_{\mathbb{R}} x_1 dP_{X_1}(x_1) \int_{\mathbb{R}} x_2 dP_{X_2}(x_2), \end{aligned}$$

kde ve druhé rovnosti jsme použili nezávislost X_1 a X_2 ($P_{\mathbf{X}} = P_{X_1} \otimes P_{X_2}$), ve třetí rovnosti Fubiniovu větu a ve čtvrté rovnosti linearitu integrálu.

Takže chybí ukázat, že za předpokladů věty je $E |X_1 X_2| < \infty$. Buď $n \in \mathbb{N}$, uvažujme funkci $g_n : \mathbb{R}^2 \rightarrow \mathbb{R}$ definovanou jako $g_n(x_1, x_2) = |x_1| \mathbb{1}_{[|x_1| \leq n]} |x_2| \mathbb{1}_{[|x_2| \leq n]}$. Tedy $g_n(X_1, X_2)$ je omezená náhodná veličina a $E g_n(X_1, X_2) \in \mathbb{R}$. Navíc platí

$$\begin{aligned} E g_n(X_1, X_2) &= \int_{\mathbb{R}^2} |x_1| \mathbb{1}_{[|x_1| \leq n]} |x_2| \mathbb{1}_{[|x_2| \leq n]} dP_{X_1} \otimes P_{X_2}(\mathbf{x}) \\ &= E |X_1| \mathbb{1}_{[|X_1| \leq n]} E |X_2| \mathbb{1}_{[|X_2| \leq n]} \leq E |X_1| E |X_2|, \end{aligned} \quad (3.2)$$

kde jsme opět použili předpoklad nezávislosti X_1 a X_2 , Fubiniovu větu a linearitu integrálu. Nezáporné funkce $g_n(x_1, x_2) \nearrow x_1 x_2$ na \mathbb{R}^2 , takže z Leviho věty $E g_n(X_1, X_2) \nearrow E |X_1 X_2|$ a $E |X_1 X_2| \leq E |X_1| E |X_2| < \infty$ z nerovnosti (3.2). \square

Poznámka. Uvědomte si, že k důkazu konečnosti momentu $E X_1 X_2$ z předpokladu konečnosti $E |X_1|$ a $E |X_2|$ nelze přímo využít Hölderovu nerovnost. Dokonce bez předpokladu nezávislosti X_1 a X_2 nemusí být $E X_1 X_2$ konečné (najděte (proti)příklad). Nezávislost je tedy důležitá nejen pro tvar $E X_1 X_2$, ale také pro to, aby pro konečnost $E X_1 X_2$ stačila jen konečnost prvních momentů X_1 a X_2 .

Nezávislost složek náhodného vektoru už určuje, jak vypadá kovariance mezi nimi, i jak počítat rozptyl lineární kombinace nezávislých veličin:

*Zde končí
předn. 9
(19.3.)*

Věta 3.12 (o cov a var pro nezávislé náhodné veličiny) Buďte X_1, \dots, X_n nezávislé náhodné veličiny a buď $E |X_i| < \infty$, $i = 1, \dots, n$. Pak $\text{cov}(X_i, X_j) = 0$, $\forall i \neq j$ a pro každé $\mathbf{a} \in \mathbb{R}^n$ platí

$$\text{var} \left(\sum_{i=1}^n a_i X_i \right) = \sum_{i=1}^n a_i^2 \text{var} X_i.$$

Důkaz: Počítejme:

$$\begin{aligned} \text{cov}(X_i, X_j) &= E[(X_i - E X_i)(X_j - E X_j)] = \\ &= E X_i X_j - E((E X_i)X_j) - E(X_i E X_j) + E X_i E X_j = E X_i X_j - E X_i E X_j. \end{aligned}$$

Použili jsme postupně definici kovariance, roznásobení a linearitu střední hodnoty. Z předpokladů věty a věty 3.11 plyne, že $E X_i X_j \in \mathbb{R}$ a $E X_i X_j = E X_i E X_j$, takže

$$\text{cov}(X_i, X_j) = E X_i E X_j - E X_i E X_j = 0.$$

Dále počítejme

$$\begin{aligned} \text{var} \left(\sum_{i=1}^n a_i X_i \right) &= E \left[\sum_{i=1}^n a_i X_i - E \left(\sum_{i=1}^n a_i X_i \right) \right]^2 = E \left[\sum_{i=1}^n (a_i (X_i - E X_i)) \right]^2 \\ &= E \left[\sum_{i=1}^n a_i^2 (X_i - E X_i)^2 + 2 \sum_{i=1}^{n-1} \sum_{j=i+1}^n a_i a_j (X_i - E X_i)(X_j - E X_j) \right] \\ &= \sum_{i=1}^n a_i^2 \text{var} X_i + 2 \sum_{i=1}^{n-1} \sum_{j=i+1}^n a_i a_j \text{cov}(X_i, X_j), \end{aligned}$$

kde jsme postupně sloučili sumy v mnohočlenu, roznásobili, a použili linearitu střední hodnoty. Až nakonec použijeme nezávislost X_i a X_j , kdy z prvního tvrzení věty máme $\text{cov}(X_i, X_j) = 0$ pro $i \neq j$. Tedy dvojná suma je rovna 0 a

$$\text{var} \left(\sum_{i=1}^n a_i X_i \right) = \sum_{i=1}^n a_i^2 \text{var} X_i.$$

□

Důsledek (důkazu věty 3.12). Buďte X_1, \dots, X_n náhodné veličiny a necht' $E X_i^2 < \infty$, $\forall i \in \{1, \dots, n\}$. Pak platí rovnost

$$\text{var} \left(\sum_{i=1}^n a_i X_i \right) = \sum_{i=1}^n a_i^2 \text{var} X_i + 2 \sum_{i=1}^{n-1} \sum_{j=i+1}^n a_i a_j \text{cov} (X_i, X_j). \quad (3.3)$$

Takže pro nezávislé náhodné veličiny je rozptyl jejich součtu roven součtu jejich rozptylů $\text{var} \left(\sum_{i=1}^n X_i \right) = \sum_{i=1}^n \text{var} X_i$. Ale obecně to neplatí!

Definice 3.8 Náhodné veličiny X a Y definované na tomtéž pravděpodobnostním prostoru nazveme *nekorelované*, pokud platí $\text{cov} (X, Y) = 0$.

Věta 3.12 říká, že z nezávislosti (a konečných prvních momentů) plyne nekorelovanost. Obráceně to ale neplatí!

Příklad. Uvažujme náhodné veličiny X = výsledek hodu korunou, a Y = výsledek hodu dvoukorunou. Vhodný model bude takový, ve kterém budou X a Y nezávislé a výsledky obou hodů nabývají hodnot z $\{0, 1\}$, každého s pravděpodobností $\frac{1}{2}$. Tedy

$$P_{(X,Y)^T} = \frac{1}{4}(\delta_{(0,0)} + \delta_{(0,1)} + \delta_{(1,0)} + \delta_{(1,1)}).$$

Označme $(U, V)^T = (X + Y, X - Y)^T$. Rozdělení náhodného vektoru (U, V) napočítáme po jednotlivých možnostech, kterých $(X, Y)^T$ nabývá s nenulovou pravděpodobností. Výsledek můžeme zapsat do tabulky

$X - Y \setminus X + Y$	0	1	2	marg V
-1	0	$\frac{1}{4}$	0	$\frac{1}{4}$
0	$\frac{1}{4}$	0	$\frac{1}{4}$	$\frac{1}{2}$
1	0	$\frac{1}{4}$	0	$\frac{1}{4}$
marg U	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$	

Zřejmě U a V nejsou nezávislé, např.:

$$P_{(U,V)^T}(0, 0) = \frac{1}{4} \neq P(U = 0)P(V = 0) = \frac{1}{4} \cdot \frac{1}{2} = \frac{1}{8}.$$

Ale U a V jsou nekorelované (neboť rozdělení obou vektorů je diskrétní s nosičem obsahujícím jen konečně mnoho bodů, existují momenty libovolně vysokého řádu, tedy i rozptyly a kovariance). Počítejme

$$\begin{aligned} \text{cov} (U, V) &= \text{cov} (X + Y, X - Y) = E (X^2 - Y^2) - E (X + Y)(X - Y) \\ &= E X^2 - E Y^2 - (E X)^2 + (E Y)^2 = \text{var} X - \text{var} Y. \end{aligned}$$

Nebo jinak počítejme

$$\begin{aligned}\operatorname{cov}(U, V) &= \operatorname{cov}(X + Y, X - Y) = \operatorname{cov}(X, X) + \operatorname{cov}(X, Y) + \operatorname{cov}(Y, X) - \operatorname{cov}(Y, Y) \\ &= \operatorname{cov}(X, X) + 0 + 0 - \operatorname{cov}(Y, Y) = \operatorname{var} X - \operatorname{var} Y.\end{aligned}$$

Ale $\operatorname{cov}(U, V) = \operatorname{var} X - \operatorname{var} Y = 0$, neboť X a Y mají stejné rozdělení a tedy i stejnou hodnotu rozptylu.

Zatím jsme definovali jen kovarianci pro náhodné veličiny. Ale pro náhodné vektory bychom chtěli také něco analogického k rozptylu pro náhodné veličiny.

Definice 3.9 *Varianční matice* náhodného vektoru \mathbf{X} je matice $n \times n$ s prvky $a_{i,j} = \operatorname{cov}(X_i, X_j)$, $i, j = 1, \dots, n$, tj.

$$\operatorname{Var} \mathbf{X} = \mathbb{E}(\mathbf{X} - \mathbb{E} \mathbf{X})(\mathbf{X} - \mathbb{E} \mathbf{X})^T.$$

Korelační matice náhodného vektoru \mathbf{X} je matice $n \times n$ s prvky $a_{i,j} = \operatorname{cor}(X_i, X_j)$, $i, j = 1, \dots, n$.

A jaké má varianční matice vlastnosti?

Věta 3.13 (o vlastnostech varianční matice) Buď $\mathbf{X} = (X_1, \dots, X_n)^T$ náhodný vektor takový, že $\mathbb{E} X_i^2 < \infty \forall i = 1, \dots, n$. Pak

(i) $\operatorname{Var} \mathbf{X}$ je symetrická a pozitivně semidefinitní.

(ii) Pro libovolná $\mathbf{a} \in \mathbb{R}^m$ a matici B typu $m \times n$ je

$$\operatorname{Var}(\mathbf{a} + B\mathbf{X}) = B \operatorname{Var} \mathbf{X} B^T.$$

(iii) $|\operatorname{cov}(X_i, X_j)| \leq \sqrt{\operatorname{var} X_i \operatorname{var} X_j}$ a rovnost nastává právě tehdy, když existují $a, b \in \mathbb{R}$ takové, že $X_i = a + bX_j$ skoro jistě, nebo $X_j = a + bX_i$ skoro jistě.

(iv) Jsou-li X_1, \dots, X_n vzájemně nezávislé, pak je $\operatorname{Var} \mathbf{X}$ diagonální.

(v) $\operatorname{Var} \mathbf{X}$ je singulární právě tehdy, když existují $a_1, \dots, a_n \in \mathbb{R}$, alespoň jedno z nich nenulové, takové, že $\sum_{i=1}^n a_i X_i = k$ skoro jistě, kde k je nějaká konstanta.

Důkaz: (i) Symetrie $\operatorname{Var} \mathbf{X}$ je zřejmá. Pro pozitivní semidefinitnost potřebujeme ověřit, že $\mathbf{a} \operatorname{Var} \mathbf{X} \mathbf{a}^T \geq 0$. Počítejme:

$$\mathbf{a} \operatorname{Var} \mathbf{X} \mathbf{a}^T = \sum_{i=1}^n \sum_{j=1}^n a_i \operatorname{cov}(X_i, X_j) a_j = \operatorname{var} \left(\sum_{i=1}^n a_i X_i \right) \geq 0,$$

druhá rovnost plyne z rozpisu (3.3).

(ii) Rozepsáním po složkách zjistíme, že platí $E B\mathbf{X} = B E \mathbf{X}$. Dále počítáme

$$\begin{aligned} \text{Var}(\mathbf{a} + B\mathbf{X}) &= E(\mathbf{a} + B\mathbf{X} - E(\mathbf{a} + B\mathbf{X}))(\mathbf{a} + B\mathbf{X} - E(\mathbf{a} + B\mathbf{X}))^T \\ &= E[B(\mathbf{X} - E\mathbf{X})(B(\mathbf{X} - E\mathbf{X}))^T] \\ &= E B(\mathbf{X} - E\mathbf{X})(\mathbf{X} - E\mathbf{X})^T B^T = B \text{Var} \mathbf{X} B^T. \end{aligned}$$

(iii) Už jsme ukázali jako důsledek Hölderovy nerovnosti 3.10.

(iv) Je zřejmé z věty 3.12.

(v) $\text{Var} \mathbf{X}$ je singulární právě tehdy, když existuje $\mathbf{a} \in \mathbb{R}^n$, $\mathbf{a} \neq \mathbf{0}$ tak, že $\mathbf{a} \text{Var}(\mathbf{X}) \mathbf{a}^T = 0$. Ale

$$\mathbf{a} \text{Var}(\mathbf{X}) \mathbf{a}^T = \text{var}(\mathbf{a}\mathbf{X}) = E(\mathbf{a}\mathbf{X} - E \mathbf{a}\mathbf{X})^2 = 0,$$

právě když $\mathbf{a}\mathbf{X} - E \mathbf{a}\mathbf{X} = 0$ skoro jistě. \square

Na závěr uvedeme ještě jedno tvrzení o průměru nezávislých náhodných veličin, které budeme potřebovat později.

Věta 3.14 (o momentech výběrového průměru) Buďte X_1, \dots, X_n nezávislé (a nebo jen nekorelované) náhodné veličiny a buďte $E X_i = \mu$, $\text{var} X_i = \sigma^2$, $\forall i = 1, \dots, n$, $\mu \in \mathbb{R}$, $\sigma^2 > 0$. Pak pro $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$ platí $E \bar{X}_n = \mu$, $\text{var} \bar{X}_n = \frac{\sigma^2}{n}$.

Důkaz: Dostáváme snadno rozepsáním z linearitě E a věty 3.13 o vlastnostech varianční matice, bod (ii). \square

*Zde končí
předn. 10
(20.3.)*

3.4 ROZDĚLENÍ TRANSFORMOVANÉHO NÁHODNÉHO VEKTORU

Na konci kapitoly 3.2 jsme si rozmysleli, že máme-li náhodný vektor $(X, Y)^T$ s nezávislými složkami a $g_i : \mathbb{R} \rightarrow \mathbb{R}$, $i = 1, 2$, měřitelné funkce, pak $(g_1(X), g_2(Y))^T$ je také náhodný vektor s nezávislými složkami, a rozdělení $P_{g_1(X)}$ snadno získáme jako rozdělení prvního marginálu. Ale co když zobrazení není “po složkách”? Jak potom získat rozdělení ztransformovaného vektoru?

Uvažujme náhodný vektor $(X, Y)^T$ s nezávislými složkami a měřitelné zobrazení $\Psi : \mathbb{R}^2 \rightarrow \mathbb{R}$. Buď $U = \Psi(X, Y)$. Rozdělení P_U je zřejmě $\Psi(P_X \otimes P_Y)$, obraz součinné míry $P_X \otimes P_Y$. Ale jak tahle míra vypadá? Bude jednodušší odvodit F_U , ta také charakterizuje rozdělení náhodné veličiny U .

Věta 3.15 Buďte X, Y nezávislé náhodné veličiny a $\Psi : \mathbb{R}^2 \rightarrow \mathbb{R}$ měřitelné zobrazení. Pak náhodná veličina $U = \Psi(X, Y)$ má distribuční funkci

$$\begin{aligned} G_U(u) &= \int_{\mathbb{R}} \int_{y: \Psi(x,y) \leq u} dF_Y(y) dF_X(x) = \int_{\mathbb{R}} \int_{y: \Psi(x,y) \leq u} dP_Y(y) dP_X(x) \\ &= \int_{\mathbb{R}} \int_{x: \Psi(x,y) \leq u} dF_X(x) dF_Y(y) = \int_{\mathbb{R}} \int_{x: \Psi(x,y) \leq u} dP_X(x) dP_Y(y), \quad u \in \mathbb{R}, \end{aligned}$$

a

$$E U = \int_{\mathbb{R}} \int_{\mathbb{R}} \Psi(x, y) dP_Y(y) dP_X(x),$$

pokud má $E U$ smysl.

Důkaz: Rovnost $E U = \int_{\mathbb{R}^2} \Psi(x, y) dP_X \otimes P_Y(x, y)$ platí z věty o obrazu míry. Pokud má $E U$ smysl, pak můžeme použít Fubiniovu větu a přepsat

$$E U = \int_{\mathbb{R}^2} \Psi(x, y) dP_X \otimes P_Y(x, y) = \int_{\mathbb{R}} \int_{\mathbb{R}} \Psi(x, y) dP_Y(y) dP_X(x).$$

Speciálně při použití na $\tilde{U} = \mathbb{1}_{(-\infty, u]}(U)$, která je integrovatelná, dostaneme

$$\begin{aligned} G_U(u) &= P(U \leq u) = E \mathbb{1}_{(-\infty, u]}(U) = \int_{\mathbb{R}} \int_{\mathbb{R}} \mathbb{1}(\Psi(x, y) \leq u) dP_Y(y) dP_X(x) \\ &= \int_{\mathbb{R}} \int_{y: \Psi(x, y) \leq u} dF_Y(y) dF_X(x), \quad u \in \mathbb{R}. \end{aligned}$$

Verze integrálů s prohozenými X a Y vyjde při použití opačného pořadí integrování ve Fubiniově větě. \square

Nejčastěji se vyskytující transformace je součet $\Psi(x, y) = x + y$.

Věta 3.16 (o rozdělení součtu nezávislých náhodných veličin) Buďte X, Y nezávislé náhodné veličiny. Pak jejich součet $U = X + Y$ má distribuční funkci

$$F_U(u) = \int_{\mathbb{R}} F_X(u - y) dF_Y(y) = \int_{\mathbb{R}} F_Y(u - x) dF_X(x), \quad u \in \mathbb{R}. \quad (3.4)$$

Důkaz: Dosazením do předchozí věty máme

$$F_U(u) = \int_{\mathbb{R}} \int_{x+y \leq u} dF_Y(y) dF_X(x) = \int_{\mathbb{R}} F_Y(u - x) dF_X(x) = \int_{\mathbb{R}} F_X(u - y) dF_Y(y), \quad u \in \mathbb{R}.$$

\square

Rozdělení součtu nezávislých náhodných veličin má i svůj speciální název.

Definice 3.10 Buď $\Psi : (x, y) \rightarrow (x + y)$. Pak $\Psi(P_X \otimes P_Y)$ se nazývá *konvoluce pravděpodobnostních rozdělení* P_X a P_Y a značíme ji $P_U = P_X * P_Y$.

Buďte F_X a F_Y distribuční funkce. Pak F_U definovaná pomocí vztahu (3.4) se nazývá *konvoluce distribučních funkcí* a značíme $F_U = F_X * F_Y$.

Důsledek (věty o rozdělení součtu). Buďte X a Y nezávislé absolutně spojitě náhodné veličiny. Pak náhodná veličina $U = X + Y$ je také absolutně spojitá s hustotou

$$f_U(u) = \int_{\mathbb{R}} f_X(u - y) f_Y(y) dy = \int_{\mathbb{R}} f_Y(u - x) f_X(x) dx, \quad u \in \mathbb{R}.$$

Důkaz: Dosazením do (3.4) máme pro každé $u \in \mathbb{R}$

$$\begin{aligned} F_U(u) &= \int_{\mathbb{R}} \int_{-\infty}^{u-x} f_Y(y) f_X(x) dy dx = \int_{\mathbb{R}} \int_{-\infty}^u f_Y(z-x) f_X(x) dz dx \\ &= \int_{-\infty}^u \left[\int_{\mathbb{R}} f_Y(z-x) f_X(x) dx \right] dz, \end{aligned}$$

kde v druhé rovnosti jsme použili substituci $z = y + x$. Neboť $F_U(u) = \int_{-\infty}^u f_U(z) dz$ pro každé $u \in \mathbb{R}$, splňuje $\int_{\mathbb{R}} f_Y(z-x) f_X(x) dx$ definici hustoty náhodné veličiny U a U je absolutně spojitá.

Výraz pro f_U s prohozenými X a Y získáme výměnou X a Y – součet je komutativní. \square

Uvědomme si, že jsou-li X a Y diskrétní, tak také umíme popsat rozdělení $X + Y$ jednoduchým způsobem.

Věta 3.17 Buďte náhodné veličiny X a Y nezávislé a čítací (to jest $P(X \in \mathbb{N}_0) = 1 = P(Y \in \mathbb{N}_0)$). Pak $U = X + Y$ je také čítací a

$$P(U = u) = \sum_{n=0}^u P(X = n)P(Y = u - n), \quad u \in \mathbb{N}_0.$$

Důkaz: Snadno z věty o úplné pravděpodobnosti. \square

Pro příklady použití všech výše uvedených vět odkazujeme na cvičení.

A co když nejsou složky náhodného vektoru \mathbf{X} nezávislé? Co můžeme dělat pak? Jak určit $\Psi(P_{\mathbf{X}})$?

Pokud má \mathbf{X} diskrétní rozdělení $P_{\mathbf{X}} = \sum_{j \in I} p_j \delta_{\mathbf{x}_j}$, I spočetná indexová množina, pak prostě přepočítáme

$$P_{\Psi(\mathbf{X})} = \Psi(P_{\mathbf{X}}) = \sum_{j \in I} p_j \delta_{\Psi(\mathbf{x}_j)}.$$

Samozřejmě Dirakovy míry v sumě nemusí být různé, pokud Ψ není prostá na $S_{\mathbf{X}} = \{\mathbf{x}_j\}_{j \in I}$, nosiči rozdělení \mathbf{X} .

Pokud má \mathbf{X} absolutně spojitě rozdělení a Ψ je difeomorfismus, můžeme využít větu o substituci z teorie míry. Zopakujme si ji:

Věta 3.18 (o substituci) Buď $M \subseteq \mathbb{R}^n$ otevřená a $\phi : M \rightarrow \mathbb{R}^n$ difeomorfismus. Buď $h : \phi(M) \rightarrow \mathbb{R}$ Lebesgueovsky měřitelná funkce a $N \subset \phi(M)$ Lebesgueovsky měřitelná množina. Pak

$$\int_N h(\mathbf{x}) d\mathbf{x} = \int_{\phi^{-1}(N)} h(\phi(\mathbf{t})) |J\phi(\mathbf{t})| d\mathbf{t},$$

pokud má alespoň jedna strana smysl.

Využijeme ji v důkazu následující věty.

Věta 3.19 (o transformaci hustot) Buď \mathbf{X} n -rozměrný absolutně spojitý náhodný vektor s hustotou $f_{\mathbf{X}}$ vzhledem k λ^n . Buď $S_{\mathbf{X}}$ otevřená taková, že $P(\mathbf{X} \in S_{\mathbf{X}}) = 1$ a $g : S_{\mathbf{X}} \rightarrow \mathbb{R}^n$ difeomorfismus. Pak rozdělení náhodného vektoru $\mathbf{Y} = g(\mathbf{X})$ má vzhledem k λ^n hustotu

$$f_{\mathbf{Y}}(\mathbf{y}) = f_{\mathbf{X}}(g^{-1}(\mathbf{y})) |Jg^{-1}(\mathbf{y})| \mathbb{1}_{g(S_{\mathbf{X}})}(\mathbf{y}), \quad \mathbf{y} \in \mathbb{R}^n. \quad (3.5)$$

Poznámka. Připomeňme si, že zobrazení g je difeomorfismus na otevřené $S_{\mathbf{X}}$, pokud je prosté, třídy C^1 a platí-li $Jg(\mathbf{x}) \neq 0$, $\mathbf{x} \in S_{\mathbf{X}}$. Potom je i inverzní zobrazení g^{-1} difeomorfismus na otevřené $g(S_{\mathbf{X}})$. Navíc platí

$$Jg^{-1}(\mathbf{y}) = \left(Jg(\mathbf{x}) \Big|_{\mathbf{x}=g^{-1}(\mathbf{y})} \right)^{-1} \quad \text{a} \quad |Jg^{-1}(\mathbf{y})| = \left| Jg(\mathbf{x}) \Big|_{\mathbf{x}=g^{-1}(\mathbf{y})} \right|^{-1},$$

pro $\mathbf{y} \in g(S_{\mathbf{X}})$.

Důkaz: Z věty o obrazu míry víme $P_{\mathbf{X}}(A) = P_{\mathbf{Y}}(g(A))$ pro každé $A \in \mathcal{B}^n$, resp. každé $g(A) \in \mathcal{B}^n$. Pokud existuje hustota $f_{\mathbf{Y}}$, pak také $P_{\mathbf{Y}}(g(A)) = \int_{g(A)} f_{\mathbf{Y}}(\mathbf{y}) d\mathbf{y}$. Z předpokladů věty a z poznámky máme, že g^{-1} je difeomorfismus na otevřené $g(S_{\mathbf{X}})$. Použitím věty o substituci pro $h = f_{\mathbf{X}}$, $\phi = g^{-1}$, $M = g(S_{\mathbf{X}})$ a $N = A$ dostaneme

$$P_{\mathbf{X}}(A) = \int_A f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x} = \int_{g(A)} f_{\mathbf{X}}(g^{-1}(\mathbf{y})) |Jg^{-1}(\mathbf{y})| d\mathbf{y},$$

pro každou $A \subseteq g^{-1}(g(S_{\mathbf{X}})) = S_{\mathbf{X}}$, resp. $g(A) \subseteq g(S_{\mathbf{X}})$. Levý integrál zřejmě existuje, tedy i pravý.

Z předpokladů víme

$$\int_{S_{\mathbf{X}}^c} f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x} = 0,$$

a

$$\int_{g(S_{\mathbf{X}})^c} f_{\mathbf{X}}(g^{-1}(\mathbf{y})) |Jg^{-1}(\mathbf{y})| \mathbb{1}_{g(S_{\mathbf{X}})}(\mathbf{y}) d\mathbf{y} = 0,$$

a integrály jsou nulové i pro podmnožiny $S_{\mathbf{X}}^c$, resp. $g(S_{\mathbf{X}})^c$.

Takže pro každou $B \in \mathcal{B}^n$ a funkci (3.5) dostáváme

$$\begin{aligned} P_{\mathbf{Y}}(B) &= P_{\mathbf{Y}}(B \cap g(S_{\mathbf{X}})) + P_{\mathbf{Y}}(B \setminus g(S_{\mathbf{X}})) = P_{\mathbf{X}}(g^{-1}(B \cap g(S_{\mathbf{X}}))) + 0 \\ &= \int_{B \cap g(S_{\mathbf{X}})} f_{\mathbf{Y}}(\mathbf{y}) d\mathbf{y} + 0 = \int_{B \cap g(S_{\mathbf{X}})} f_{\mathbf{Y}}(\mathbf{y}) d\mathbf{y} + \int_{B \setminus g(S_{\mathbf{X}})} f_{\mathbf{Y}}(\mathbf{y}) d\mathbf{y} = \int_B f_{\mathbf{Y}}(\mathbf{y}) d\mathbf{y}. \end{aligned}$$

Funkce $f_{\mathbf{Y}}$ je tedy opravdu hustota rozdělení $P_{\mathbf{Y}}$ vzhledem k Lebesgueově míře na \mathbb{R}^n . \square

Poznámka. To, co jsme v důkazu museli vyřešit navíc oproti větě o substituci, bylo, aby $f_{\mathbf{Y}}$ byla definovaná na celém \mathbb{R}^n , a využili jsme k tomu předpoklad, že množina $S_{\mathbf{X}}$ je nosič rozdělení $P_{\mathbf{X}}$ a zároveň je otevřená.

Jako příklad použití věty o transformaci hustot si ukážeme odvození rozdělení součinu dvou nezávislých náhodných veličin. Pro další příklady užití odkazujeme na cvičení.

Věta 3.20 (o rozdělení součinu a podílu dvou nezávislých náhodných veličin) Budťe X, Y , nezávislé náhodné veličiny s hustotami f_X, f_Y vzhledem k λ na \mathbb{R} . Pak $U = XY$ je také absolutně spojitá náhodná veličina s hustotou

$$f_U(u) = \int_{\mathbb{R}} f_X\left(\frac{u}{x}\right) f_Y(x) \frac{1}{|x|} \mathbb{1}(x \neq 0) dx. \quad (3.6)$$

Pokud navíc $P(Y > 0) = 1$, pak $V = \frac{X}{Y}$ je absolutně spojitá náhodná veličina s hustotou

$$f_V(v) = \int_0^\infty f_X(vx) f_Y(x) x dx.$$

Důkaz: Uvažujme $U = XY$. Tvrzení věty můžeme dokázat přes distribuční funkce a větu 3.15 jako skripta [Dupač and Hušková \(2013\)](#) str. 57 nebo použijeme větu o transformaci hustot.

Rozdělení $(X, Y)^T$ je absolutně spojitě a hustota vzhledem k λ^2 na \mathbb{R}^2 je z předpokladů $f_{(X,Y)^T}(x, y) = f_X(x) f_Y(y)$, $(x, y) \in \mathbb{R}^2$.

Bud' zobrazení g definované jako $g(x, y) = (xy, y)$, a označme $(U, V)^T = g((X, Y)^T)$. Pak g je difeomorfismus na otevřené množině $S_{(X,Y)^T} = \mathbb{R}^2 \setminus \mathbb{R} \times \{0\}$. Platí $P_{(X,Y)^T}((S_{(X,Y)^T})^c) = P(Y = 0) = 0$, neboť Y je z předpokladů absolutně spojitá náhodná veličina. Tedy $S_{(X,Y)^T}$ je opravdu nosič rozdělení $P_{(X,Y)^T}$.

Platí $g(S_{(X,Y)^T}) = \mathbb{R}^2 \setminus \mathbb{R} \times \{0\}$. Inverzní zobrazení $g^{-1}(u, v) = \left(\frac{u}{v}, v\right)$ na $\mathbb{R}^2 \setminus \mathbb{R} \times \{0\}$. Platí

$$|Jg(x, y)| = \left| \det \begin{pmatrix} y & 0 \\ x & 1 \end{pmatrix} \right| = |y| \neq 0, \quad (x, y) \in S_{(X,Y)^T}, \quad |Jg^{-1}(u, v)| = \frac{1}{|v|}, \quad (u, v) \in \mathbb{R}^2 \setminus \mathbb{R} \times \{0\}.$$

Předpoklady věty 3.19 jsou splněny a dostáváme

$$f_{(U,V)^T}(u, v) = f_X\left(\frac{u}{v}\right) f_Y(v) \frac{1}{|v|} \mathbb{1}(v \neq 0), \quad (u, v) \in \mathbb{R}^2.$$

Marginální hustotu $U = XY$ dostaneme vyintegrováním podle věty 3.4

$$f_U(u) = \int_{\mathbb{R}} f_X\left(\frac{u}{v}\right) f_Y(v) \frac{1}{|v|} \mathbb{1}(v \neq 0) dv.$$

Pro podíl dokážeme tvrzení věty analogicky za použití zobrazení $g : (x, y) \rightarrow \left(\frac{x}{y}, x\right)$. \square

Poznámka. Vidíme, že při použití věty o transformaci pro důkaz není nezávislost X a Y zásadním předpokladem. Pokud by X a Y nebyly nezávislé, ale $(X, Y)^T$ by byl absolutně spojitý náhodný vektor s hustotou $f_{(X,Y)^T}$, pak by $U = XY$ byla stále absolutně spojitá náhodná veličina ovšem s hustotou

$$f_U(u) = \int_{\mathbb{R}} f_{(X,Y)^T}\left(\frac{u}{v}, v\right) \frac{1}{|v|} \mathbb{1}(v \neq 0) dv.$$

*Zde končí
předn. II
(26.3.)*

3.5 PŘÍKLADY ROZDĚLENÍ NÁHODNÉHO VEKTORU

V této kapitole si ukážeme dvě nejčastější mnohorozměrná rozdělení, jedno diskrétní a jedno absolutně spojitě.

3.5.1 MULTINOMICKÉ ROZDĚLENÍ

Buď \mathbf{X} k -rozměrný náhodný vektor. Buď dále $n \in \mathbb{N}$ a konstanty $p_i > 0$, $i = 1, \dots, k$, takové, že $\sum_{i=1}^k p_i = 1$.

\mathbf{X} má *multinomické rozdělení* s parametry n , $\{p_i\}_{i=1}^k$, pokud platí

$$P_{\mathbf{X}}(\mathbf{x}) = P(X_1 = x_1, \dots, X_k = x_k) = \frac{n!}{x_1! \dots x_k!} p_1^{x_1} \dots p_k^{x_k}$$

pro každé $\mathbf{x} \in S_{\mathbf{X}} = \{\mathbf{x} \in \{0, \dots, n\}^k : \sum_{i=1}^k x_i = n\}$. Z multinomické věty dostaneme, že $\sum_{\mathbf{x} \in S_{\mathbf{X}}} P(\mathbf{X} = \mathbf{x}) = 1$, takže $P_{\mathbf{X}}$ je dobře definované diskrétní rozdělení náhodného vektoru.

Uvažujme situaci, kdy nezávisle n -krát opakujeme pokus, který může mít k různých výsledků s pravděpodobnostmi p_i , $i = 1, \dots, k$. Pokud označíme X_i počet pokusů s výsledkem i , pak $\mathbf{X} = (X_1, \dots, X_k)^T$ má multinomické rozdělení.

Platí, že marginální rozdělení X_i jsou binomická s parametry n a p_i (ověřte buď pomocí věty o úplné pravděpodobnosti nebo věty 3.4).

Platí

$$E X_i = n p_i \quad \text{var } X_i = n p_i (1 - p_i) \quad \text{cov}(X_i, X_j) = -n p_i p_j, \quad i \neq j.$$

První dvě rovnosti plynou z binomického marginálního rozdělení, třetí odvodíme takto:

$$\begin{aligned} E X_i X_j &= \sum_{\substack{0 \leq x_i + x_j \leq n \\ \sum_{i=1}^k x_i = n}} x_i x_j \frac{n!}{x_1! \dots x_k!} p_1^{x_1} \dots p_k^{x_k} \\ &= n(n-1) p_i p_j \sum_{\substack{0 \leq (x_i-1) + (x_j-1) \leq n-2 \\ \sum_{i=1}^k x_i - 2 = n-2}} \frac{(n-2)!}{x_1! \dots (x_i-1)! \dots (x_j-1)! \dots x_k!} p_1^{x_1} \dots p_i^{(x_i-1)} \dots p_j^{(x_j-1)} \dots p_k^{x_k} \\ &= n(n-1) p_i p_j, \end{aligned}$$

kde v poslední rovnosti jsme využili fakt, že v sumě byly všechny nenulové pravděpodobnosti odpovídající multinomickému rozdělení s parametry $n-2$, $\{p_i\}_{i=1}^k$, takže se vysčítaly na 1. Tedy

$$\text{cov}(X_i, X_j) = E X_i X_j - E X_i E X_j = n(n-1) p_i p_j - n^2 p_i p_j = -n p_i p_j.$$

Z nenulové kovariance je zřejmé, že složky multinomicky rozděleného vektoru nejsou nezávislé náhodné veličiny.

Ale součet nezávislých multinomicky rozdělených vektorů má opět multinomické rozdělení. Přesněji: Buďte $p_i > 0$, $i = 1, \dots, k$, takové, že $\sum_{i=1}^k p_i = 1$, a $n_1, \dots, n_m \in \mathbb{N}$, kde $m \in \mathbb{N}$. Nechť mají náhodné vektory \mathbf{Y}_j , $j = 1, \dots, m$ multinomická rozdělení s parametry $\{p_i\}_{i=1}^k$ a n_j , a nechť jsou vzájemně nezávislé. Pak $\sum_{j=1}^m \mathbf{Y}_j$ má také multinomické rozdělení s parametry $\{p_i\}_{i=1}^k$ a $n = \sum_{j=1}^m n_j$ (ověřte).

3.5.2 MNOHORozměRNÉ NORMÁLNÍ RozDĚLENÍ

V kapitole o náhodných veličinách jsme zdefinovali normované normální rozdělení $N(0, 1)$ s hustotou

$$f_Z(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}, \quad z \in \mathbb{R}$$

vzhledem k λ . Má-li Z rozdělení $N(0, 1)$ a je-li $X = \sigma Z + \mu$, pro nějaké $\sigma > 0$ a $\mu \in \mathbb{R}$, pak X má normální rozdělení $N(\mu, \sigma^2)$ s hustotou

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

a $E X = \mu$, $\text{var } X = \sigma^2$.

Nyní zdefinujeme mnohorozměrnou verzi normálního rozdělení.

Definice 3.11 Bud' $\mathbf{Z} = (Z_1, \dots, Z_r)^\top$, $r \in \mathbb{N}$, kde Z_i jsou vzájemně nezávislé a mají $N(0, 1)$ rozdělení. Bud' $A_{n \times r}$, $n \in \mathbb{N}$ matice a $\mu \in \mathbb{R}^n$ pevný vektor. Náhodný vektor definovaný jako $\mathbf{X} = A\mathbf{Z} + \mu$ má n -rozměrné normální rozdělení s parametry μ a $\Sigma = AA^\top$. Značíme $N_n(\mu, \Sigma)$.

Poznámka. Náhodný vektor \mathbf{Z} z definice má zřejmě $E \mathbf{Z} = \mathbf{0}$ a $\text{Var } \mathbf{Z} = \mathbb{I}_r$ jednotková matice.

Definice vypadá poněkud komplikovaně, ale je vymyšlená tak, aby zobecňovala rozdělení náhodné veličiny $N(\mu, \sigma^2)$, parametry byly opět shodné s prvními dvěma momenty, třída rozdělení $N(\mu, \sigma^2)$ byla stabilní vzhledem k lineárním transformacím, a ty se hezkým způsobem přenášely do změny parametrů, a aby umožňovala zahrnout do definice i degenerovaná rozdělení (extrémní případ je např. $N(\mu, \mathbf{0}) = \delta_\mu$ kde $\mu \in \mathbb{R}^n$ je pevný vektor).

Důsledky (definice mnohorozměrného normálního rozdělení). Bud' \mathbf{X} a \mathbf{Z} jako v definici 3.11.

- (i) Volíme-li v definici $A = \mathbb{I}_r$ jednotková matice a $\mu = \mathbf{0}$ nulový vektor, pak dostaneme, že \mathbf{Z} má rozdělení $N_r(\mathbf{0}, \mathbb{I}_r)$.
- (ii) Platí $E \mathbf{X} = \mu$ a $\text{Var } \mathbf{X} = \Sigma$.
- (iii) Pro $k \in \mathbb{N}$ a matici $B_{k \times n}$ platí $\mathbf{Y} = B\mathbf{X}$ má normální rozdělení $N_k(B\mu, B\Sigma B^\top)$.

(iv) Speciálně pro vektor $\mathbf{c} \in \mathbb{R}^n$ má \mathbf{cX} jednorozměrné normální rozdělení $N(\mathbf{c}\mu, \mathbf{c}\Sigma\mathbf{c}^T)$.

Důkaz: (ii) Podle definice $\mathbf{X} = A\mathbf{Z} + \mu$, tedy dosadíme a podle pravidel pro výpočet E a varianční matice (věta 3.13) počítáme

$$E\mathbf{X} = E(A\mathbf{Z} + \mu) = AE\mathbf{Z} + \mu = A\mathbf{0} + \mu = \mu,$$

a

$$\text{Var } \mathbf{X} = \text{Var}(A\mathbf{Z} + \mu) = A\text{Var } \mathbf{Z}A^T = A\mathbb{I}_rA^T = AA^T.$$

(iii) Podle definice $\mathbf{X} = A\mathbf{Z} + \mu$, tedy $Y = B\mathbf{X} = BA\mathbf{Z} + B\mu$ a podle definice má k -rozměrné normální rozdělení s parametry $B\mu$ a $BA(BA)^T = B\Sigma B^T$.

(iv) Speciální případ bodu (iii) pro matici $1 \times n$. □

Poznámka. Bod (iv) říká, že (vážený) součet normálně rozdělených marginálů

$$\sum_{i=1}^n c_i X_i$$

má opět normální rozdělení $N(\sum_{i=1}^n c_i \mu_i, \mathbf{c}^T \Sigma \mathbf{c})$. To je vcelku výjimečné – součet náhodných veličin s nějakým rozdělením nemá typicky to samé rozdělení. A dokonce ty marginály nemusí být nezávislé (pokud není Σ diagonální, tak nejsou nezávislé). Ovšem předpokládáme, že sdužené rozdělení \mathbf{X} je normální, nestačí předpokládat, že jednotlivé marginály X_i mají normální rozdělení.

A jak rozdělení $N_n(\mu, \Sigma)$ vypadá? Je absolutně spojitě? A jakou má hustotu? – Odpověď závisí na tom, jaká je Σ .

Víme z lineární algebry, že každou symetrickou pozitivně semidefinitní matici $\Sigma_{n \times n}$ lze rozepsat jako $\Sigma = AA^T$ pro nějakou $A_{n \times r}$, $r \leq n$. Navíc platí, že $r < n$ právě tehdy, když je Σ singulární. Takže pro singulární $\Sigma_{n \times n}$ existuje n -rozměrný normálně rozdělený náhodný vektor \mathbf{X} , který vznikl lineární transformací z r -rozměrného normálně rozděleného vektoru \mathbf{Z} s nezávislými složkami a $r < n$. Takže nelze očekávat, že by $P_{\mathbf{X}} \ll \lambda^n$. A ono to také neplatí:

Mějme n -rozměrný náhodný vektor \mathbf{X} s normálním rozdělením $N_n(\mu, \Sigma)$ a buď Σ singulární. Z důsledků definice (ii) plyne, že $\text{Var } \mathbf{X} = \Sigma$. Z věty 3.13 o vlastnostech varianční matice, bod (v) plyne, že existuje $\mathbf{a} \in \mathbb{R}^n$ takové, že $\mathbf{a}^T \mathbf{X} = k$ skoro jistě, kde $k \in \mathbb{R}$ je nějaká konstanta. Označme

$$E = \{x \in \mathbb{R}^n : \mathbf{a}^T \mathbf{x} = k\}.$$

Pak $P_{\mathbf{X}}(E) = 1$ ale $\lambda^n(E) = 0$. Takže opravdu nemůže platit $P_{\mathbf{X}} \ll \lambda^n$.

Extrémní případ nastává pro $\Sigma = \mathbb{0}_{n \times n}$ nulovou matici. Platí, že \mathbf{X} má rozdělení $N_n(\mu, \mathbb{0}_{n \times n})$ právě tehdy, když $P_{\mathbf{X}} = \delta_{\mu}$.

Pro Σ regulární je ale rozdělení $N_n(\mu, \Sigma)$ absolutně spojitě, jak ukazuje následující věta.

Věta 3.21 (o hustotě n -rozměrného normálního rozdělení) Buď \mathbf{X} náhodný vektor s rozdělením $N_n(\mu, \Sigma)$, kde Σ je regulární matice. Pak $P_{\mathbf{X}} \ll \lambda^n$ a

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{(2\pi)^{\frac{n}{2}} \sqrt{\det \Sigma}} e^{-\frac{1}{2}(\mathbf{x}-\mu)^T \Sigma^{-1}(\mathbf{x}-\mu)}, \quad \mathbf{x} \in \mathbb{R}^n. \quad (3.7)$$

Důkaz: Nejdříve uvažujme $N(\mathbf{0}, \mathbb{I}_n)$. $\mathbf{Z} = (Z_1, \dots, Z_n)^T$ s nezávislými složkami a $N(0, 1)$ rozdělenými marginály má toto rozdělení. Z věty o hustotě vektoru s nezávislými složkami je hustota \mathbf{Z} rovna součinu hustot

$$f_{\mathbf{Z}}(\mathbf{z}) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}} e^{-\frac{z_i^2}{2}} = \frac{1}{(2\pi)^{\frac{n}{2}} \sqrt{\det \mathbb{I}}} e^{-\frac{1}{2}(\mathbf{z}-\mu)^T \mathbb{I}^{-1}(\mathbf{z}-\mu)}, \quad \mathbf{z} \in \mathbb{R}^n,$$

tedy (3.7) platí.

Buď Σ pozitivně definitní a $\mu \in \mathbb{R}^n$. Pak existuje právě jedna matice $A_{n \times n}$ tak, že $\Sigma = AA^T$. Položme $\mathbf{X} = A\mathbf{Z} + \mu$ pro \mathbf{Z} z předchozího odstavce. Pak \mathbf{X} má rozdělení $N_n(\mu, \Sigma)$ z definice.

Zobrazení $g : \mathbb{R}^n \rightarrow \mathbb{R}^n$, $g(\mathbf{x}) = A\mathbf{x} + \mu$ je difeomorfismus na celém \mathbb{R}^n , $|Jg| = |\det A| \neq 0$ na celém \mathbb{R}^n . Z věty 3.19 o transformaci hustot dostaneme, že

$$\begin{aligned} f_{\mathbf{X}}(\mathbf{x}) &= \frac{1}{(2\pi)^{\frac{n}{2}} |\det A|} e^{-\frac{1}{2}(A^{-1}(\mathbf{z}-\mu))^T (A^{-1}(\mathbf{z}-\mu))} \\ &= \frac{1}{(2\pi)^{\frac{n}{2}} \sqrt{\det AA^T}} e^{-\frac{1}{2}(\mathbf{z}-\mu)^T (A^{-1})^T A^{-1}(\mathbf{z}-\mu)}, \quad \mathbf{x} \in \mathbb{R}^n, \end{aligned}$$

což je přesně (3.7), neboť $(A^{-1})^T A^{-1} = (AA^T)^{-1} = \Sigma^{-1}$. □

Značení. Symbolem \sim budeme značit, že náhodný vektor má nějaké rozdělení – např. $\mathbf{X} \sim N_2(\mu, \Sigma)$.

Příklad. Uvažujme \mathbf{Z} s dvourozměrným normálním rozdělením $N_2(\mu, \Sigma)$. Označme

$$\text{var } Z_1 = \sigma_1^2 \quad \text{var } Z_2 = \sigma_2^2 \quad \rho = \text{cor}(Z_1, Z_2)$$

pak

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$$

Pokud volíme $\mu = (0, 0)^T$ a $\sigma_1 = \sigma_2 = 1$ a $|\rho| \in [0, 1)$, dostaneme částečně znormované dvourozměrné normální rozdělení s hustotou

$$f_{(X,Y)^T}(x, y) = \frac{1}{2\pi\sqrt{1-\rho^2}} e^{-\frac{x^2+y^2-2\rho xy}{2(1-\rho^2)}}, \quad x, y \in \mathbb{R}.$$

Uvědomme si, že regulárnost matice Σ je zajištěna volbou $|\rho| \in [0, 1)$. Odintegrováním bychom zjistili, že marginální rozdělení X i Y je $N(0, 1)$ a X a Y jsou nezávislé právě když $\rho = 0$.

Platí

$$\mathbf{Z} = \mu + \begin{pmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{pmatrix} \begin{pmatrix} X \\ Y \end{pmatrix} \sim N_2(\mu, \Sigma),$$

a obráceně

$$\begin{pmatrix} \frac{Z_1 - \mu_1}{\sigma_1} & \frac{Z_2 - \mu_2}{\sigma_2} \end{pmatrix}^\top \sim N_2\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right)$$

Důsledek (o marginálních rozděleních v N_n). Buď \mathbf{X} náhodný vektor s rozdělením $N_n(\mu, \Sigma)$. Pak marginální rozdělení X_i je $N(\mu_i, \sigma_i^2)$, kde $\sigma_i^2 = \Sigma_{i,i} = \text{var } X_i$. Podvektor $(X_i, X_j)^\top$ má rozdělení

$$N_2\left(\begin{pmatrix} \mu_i \\ \mu_j \end{pmatrix}, \begin{pmatrix} \sigma_i^2 & \rho_{ij}\sigma_i\sigma_j \\ \rho_{ij}\sigma_i\sigma_j & \sigma_j^2 \end{pmatrix}\right),$$

kde $\rho_{ij} = \text{cor}(X_i, X_j)$.

Důkaz: Použijeme důsledek definice (iv) pro speciální volbu $\mathbf{c} = (0, \dots, 1, \dots, 0) = \mathbf{e}_i$, pak $X_i = \mathbf{c}\mathbf{X}$ má rozdělení $N(\mathbf{e}_i\mu, \mathbf{e}_i\Sigma\mathbf{e}_i^\top)$.

Obdobně pro (X_i, X_j) použijeme důsledek definice (iii) s maticí $B_{2 \times n}$ složenou z vektorů \mathbf{e}_i a \mathbf{e}_j . □

*Zde končí
předn. 12
(27.3.)*

Poznámka. Z příkladu pro N_2 respektive přímým dosazením do věty 3.21 pro $n = 2$ zjistíme, že hustota $f_{(X,Y)^\top}$ je v součinném tvaru právě tehdy, když $\rho = \text{cov}(X, Y) = 0$. Tedy hustota podvektoru $(X_i, X_j)^\top$ z předchozího důsledku je v součinném tvaru právě tehdy, když $\rho_{ij} = 0$. V tom případě je hustota rovna

$$f_{(X_i, X_j)^\top}(x_i, x_j) = \frac{1}{2\pi\sqrt{\sigma_i^2\sigma_j^2}} e^{-\frac{1}{2}\left(\frac{(x_i-\mu_i)^2}{\sigma_i^2} + \frac{(x_j-\mu_j)^2}{\sigma_j^2}\right)}, \quad x_i, x_j \in \mathbb{R}.$$

Ovšem z věty 3.7 je hustota $f_{(X_i, X_j)^\top}$ v součinném tvaru právě tehdy, když jsou X_i a X_j nezávislé. Z toho dostáváme následující tvrzení:

Tvrzení 3.22 Nechť má náhodný vektor $(X, Y)^\top$ dvourozměrné normální rozdělení. Pak mají i X a Y (jednorozměrné) normální rozdělení. Pokud je navíc $\text{cor}(X, Y) = 0$, pak jsou X a Y nezávislé.

Poznámka. Víme ze sekce 3.3, že nekorelovanost dvou náhodných veličin obecně neimplikuje nezávislost. Předchozí tvrzení říká, že ve speciálním případě, kdy mají tyto dvě náhodné veličiny sdružené normální rozdělení (to je velmi omezující předpoklad), tak nekorelovanost implikuje nezávislost.

Ale pozor! Pokud jen víme, že marginální rozdělení náhodné veličiny X a náhodné veličiny Y (každé zvlášť) je jednorozměrné normální a jsou nekorelované, neznamená to ještě, že jsou nezávislé (najděte protipříklad).

Pokud totiž víme jen, že marginální rozdělení X a Y jsou normální, sdružené rozdělení $(X, Y)^T$ ještě nemusí být dvourozměrné normální (najděte protipříklad).

Ale pokud víme, že marginální rozdělení X a Y jsou normální a ony jsou vzájemně nezávislé, pak už nutně $(X, Y)^T$ má sdružené normální rozdělení (a korelace $\text{cor}(X, Y) = 0$).

Až se dostaneme ke statistice, budeme používat ještě dvě rozdělení odvozená od normálního rozdělení. Představíme si je už zde.

Definice 3.12 Bud'te X_1, \dots, X_n , $n \in \mathbb{N}$ nezávislé náhodné veličiny s rozdělením $N(0, 1)$. Pak náhodná veličina $Y_n = \sum_{i=1}^n X_i^2$ má χ^2 -rozdělení o n stupních volnosti (značíme χ_n^2), s hustotou

$$g_n(y) = \frac{y^{\frac{n}{2}-1}}{2^{\frac{n}{2}} \Gamma\left(\frac{n}{2}\right)} e^{-\frac{y}{2}} \mathbb{1}(y > 0), \quad y \in \mathbb{R}.$$

χ^2 -rozdělení je speciální příklad Γ rozdělení (viz cvičení nebo proseminář). První dva momenty χ_n^2 -rozdělení snadno spočítáme z jeho definice:

$$\mathbb{E} Y_n = \mathbb{E} \left(\sum_{i=1}^n X_i^2 \right) = \sum_{i=1}^n \mathbb{E} X_i^2 = n \cdot 1 = n.$$

Použili jsme linearitu střední hodnoty a fakt, že pro $X_i \sim N(0, 1)$ je $\mathbb{E} X_i^2 = \text{var } X_i = 1$.

$$\text{var } Y_n = \text{var} \left(\sum_{i=1}^n X_i^2 \right) = \sum_{i=1}^n \text{var } X_i^2 = n(\mathbb{E} X_1^4 - (\mathbb{E} X_1^2)^2) = n(3 - 1) = 2n,$$

kde v druhé rovnosti jsme použili vzájemnou nezávislost X_i^2 a ve čtvrté znalost momentů $N(0, 1)$.

Definice 3.13 Bud'te $X \sim N(0, 1)$ a $Y \sim \chi_n^2$, $n \in \mathbb{N}$, nezávislé náhodné veličiny. Pak náhodná veličina $T = \frac{X}{\sqrt{\frac{Y}{n}}}$ má *Studentovo t_n -rozdělení* (neboli *t -rozdělení o n stupních volnosti*) s hustotou

$$h_n(t) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{\pi n} \Gamma\left(\frac{n}{2}\right)} \left(1 + \frac{t^2}{n}\right)^{-\frac{n+1}{2}}, \quad t \in \mathbb{R}.$$

Hustota rozdělení t_n je symetrická kolem 0, takže liché momenty (pokud existují) jsou rovny 0. Ony ale ne vždy existují.

Speciální případ t_n rozdělení pro $n = 1$ má hustotu

$$h_1(t) = \frac{1}{\pi(1+t^2)}, \quad t \in \mathbb{R},$$

a distribuční funkci

$$F_T(x) = \frac{1}{\pi} \left(\arctan(x) + \frac{\pi}{2} \right), \quad x \in \mathbb{R},$$

a jmenuje se *Cauchyho rozdělení*. Uvědomme si, že integrál

$$\int_{-\infty}^{\infty} \frac{t}{\pi(1+t^2)} dt$$

není definován, a tedy neexistuje střední hodnota $E T$ Cauchyho (resp. t_1) rozdělení. A ani žádné vyšší momenty.

Rozdělení t_2 už má střední hodnotu dobře definovanou, ale má nekonečný rozptyl. Obecně platí, že t_n -rozdělení má konečné momenty jen do řádu $n - 1$ (ověřte).

Také si můžeme povšimnout, že pro každé $t \in \mathbb{R}$

$$\lim_{n \rightarrow \infty} h_n(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}},$$

což je hustota $N(0, 1)$ rozdělení. K tomuto faktu se vrátíme, až budeme vědět, co je to konvergence náhodných veličin v distribuci.

Poznámka. Není potřeba si pamatovat hustoty χ_n^2 a t_n rozdělení, ale je třeba je umět odvodit jako příklad.

4 LIMITNÍ VĚTY

Zatím jsme si představili základní objekty, se kterými teorie pravděpodobnosti pracuje - pravděpodobnostní prostor, náhodné veličiny a náhodné vektory. Když máme nějakou (dostatečně jednoduchou) náhodnou situaci, umíme pro ní sestavit pravděpodobnostní model, který ji matematicky popisuje. V této kapitole se budeme zabývat situací, kdy opakujeme nějaký náhodný experiment mnohokrát (libovolně krát) a nezávisle. Co nám naše matematická teorie říká o takové situaci, a jak to odpovídá realitě?

Uvažujme nezávislé, stejně rozdělené náhodné veličiny X_1, \dots, X_n a jejich průměr

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i.$$

Co nám říká naše zkušenost o chování \bar{X}_n ? Třeba ve fyzice, když provádíme n nezávislých identických experimentů s naměřeným výsledkem X_i (pozor, je náhodný), pak jejich průměr \bar{X}_n (také náhodný) se pro velké n ustálí kolem nějakého čísla. Intuitivně bychom toto číslo nazvali očekávanou hodnotou výsledku. Tato zkušenost také umožňuje interpretaci pravděpodobností jako relativních četností

$$P(A) \approx \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{X_i \in A}$$

tedy „pravděpodobnost události A odpovídá relativní četnosti výskytu události A při provedení velkého množství nezávislých, stejně rozdělených pozorování X_i “.

V této kapitole budeme zkoumat tzv. zákony velkých čísel, které tvrdí

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{n \rightarrow \infty} E X_1,$$

budeme si ovšem muset vyjasnit, co znamená \rightarrow .

A také se dostaneme k nejdůležitější větě celé naší přednášky, které se říká *centrální limitní věta*. Ta tvrdí, že

$$\sqrt{n} (\bar{X}_n - E X_1) \xrightarrow{n \rightarrow \infty} N(0, \sigma^2),$$

kde to na pravé straně šipky je normální rozdělení se střední hodnotou 0 a nějakým (nenulovým) rozptylem. Mluví tedy o tom, jak moc se empirický průměr \bar{X}_n liší od střední hodnoty $E X_1$. I zde si budeme muset vyjasnit, co znamená \rightarrow .

4.1 CANTELLIHO A BORELOVA VĚTA

Začneme ovšem jednodušeji:

Příklad. Bud' $\{X_i\}_{i=1}^{\infty}$ posloupnost nezávislých stejně rozdělených náhodných veličin na tomtéž pravděpodobnostním prostoru. A necht' X_i mají alternativní rozdělení s pravděpodobností úspěchu $p = \frac{1}{2}$, tj.

$$P(X_i = 1) = P(X_i = 0) = \frac{1}{2}.$$

Takové posloupnosti $\{X_i\}_{i=1}^{\infty}$ se říká *posloupnost Bernoulliových pokusů* nebo jen *Bernoulliůvská posloupnost* (odpovídá to třeba opakovaným hodům spravedlivou mincí). Jaká je pravděpodobnost jevu

$$\{\text{pro nekonečně mnoho } i \text{ je } X_i = 1\} = \{\text{padne nekonečněkrát } 1\},$$

resp.

$$\{\text{jen pro konečně mnoho } i \text{ je } X_i = 1\} = \{\text{padne jen konečněkrát } 1\} ?$$

Naše fyzikální zkušenost by navrhovala 1, resp. 0. A co na to teorie pravděpodobnosti?

Poznámka. Některé zvědavé čtenáře tu mohlo napadnout — a jak vůbec víme, že Bernoulliůvská posloupnost existuje? Nebo obecněji, jak víme, že existuje pravděpodobnostní prostor (Ω, \mathcal{A}, P) takový, že z něj může vést nekonečně mnoho nezávislých náhodných veličin se stejným rozdělením?

Odpověď na tuto otázku si v této přednášce neukážeme (není na to prostor), takže zvědavé čtenáře musíme odkázat na kurz Teorie pravděpodobnosti I, který se učí ve třetím ročníku zaměřením stochastika. A nebo na Proseminář z pravděpodobnosti a matematické statistiky, kde si ukážeme alespoň existenci Bernoulliůvské posloupnosti. Nadále tedy budeme brát existenci nekonečné posloupnosti nezávislých náhodných veličin jako fakt bez důkazu.

Značení. Buďte A_1, A_2, \dots náhodné jevy (ze σ -algebry \mathcal{A}). Značíme

$$\limsup_{n \rightarrow \infty} A_n = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k, \quad \liminf_{n \rightarrow \infty} A_n = \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k,$$

což jsou také jevy ze σ -algebry \mathcal{A} .

Pokud je $\omega \in \limsup_{n \rightarrow \infty} A_n$, tak nutně musí být v každém $\bigcup_{k=n}^{\infty} A_k$, tedy opravdu musí být v nekonečně mnoha A_n . Takže $\limsup_{n \rightarrow \infty} A_n$ je přesně množina těch elementárních jevů $\omega \in \Omega$, ve kterých nastane nekonečně mnoho z jevů A_n .

Podobně, pokud je $\omega \in \liminf_{n \rightarrow \infty} A_n$, musí být v některém z $\bigcap_{k=n}^{\infty} A_k$, tedy může být jen konečně mnoho množin A_n , do kterých ω nepatří. Takže $\liminf_{n \rightarrow \infty} A_n$ je přesně množina těch elementárních jevů $\omega \in \Omega$, ve kterých nastanou všechny jevy A_n s výjimkou jen konečně mnoha z nich.

Poznámka. Všimněme si, že množinové \limsup a \liminf odpovídají \limsup a \liminf pro funkce v tom smyslu, že

$$\begin{aligned}\limsup_{n \rightarrow \infty} \mathbb{1}_{A_n}(\omega) &= \mathbb{1}_A(\omega) & \text{pro } A &= \limsup_{n \rightarrow \infty} A_n \\ \liminf_{n \rightarrow \infty} \mathbb{1}_{A_n}(\omega) &= \mathbb{1}_B(\omega) & \text{pro } B &= \liminf_{n \rightarrow \infty} A_n\end{aligned}$$

Poznámka. Ověřte, že platí

$$\left(\limsup_{n \rightarrow \infty} A_n \right)^c = \liminf_{n \rightarrow \infty} A_n^c.$$

Tato rovnost může být užitečná při výpočtech.

A jak tedy zjistit pravděpodobnost $\limsup_{n \rightarrow \infty} A_n$? Většinou je to příliš složitá množina na to, aby byla $P\left(\limsup_{n \rightarrow \infty} A_n\right)$ přímo vidět. Někdy ji ovšem můžeme odhadnout shora.

Věta 4.1 (Cantelliho) Buďte A_1, A_2, \dots náhodné jevy (ze σ -algebry \mathcal{A}). Pokud $\sum_{n=1}^{\infty} P(A_n) < \infty$, pak $P\left(\limsup_{n \rightarrow \infty} A_n\right) = 0$.

Důkaz: Počítejme:

$$P\left(\limsup_{n \rightarrow \infty} A_n\right) = P\left(\bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k\right) = \lim_{n \rightarrow \infty} P\left(\bigcup_{k=n}^{\infty} A_k\right) \leq \lim_{n \rightarrow \infty} \sum_{k=n}^{\infty} P(A_k) = 0.$$

V druhé rovnosti jsme využili spojitost pravděpodobnosti, neboť pro $B_n = \bigcup_{k=n}^{\infty} A_k$ platí $B_n \searrow \limsup_{n \rightarrow \infty} A_n$. Poslední rovnost plyne z předpokladu věty. \square

Důsledek. Za předpokladů věty platí $P\left[\left(\limsup_{n \rightarrow \infty} A_n\right)^c\right] = 1$, což je podle poznámky výše rovno $P\left(\liminf_{n \rightarrow \infty} A_n^c\right)$. Tedy s pravděpodobností 1 nastanou všechny jevy A_n^c až na konečně mnoho výjimek, tj. nastane jen konečně mnoho jevů A_n .

Pokud přidáme předpoklad nezávislosti jevů A_1, A_2, \dots , situace se zjednoduší.

Věta 4.2 (Borelova, Borelův 0-1 zákon) Buďte A_1, A_2, \dots nezávislé náhodné jevy (ze σ -algebry \mathcal{A}). Pak

$$\begin{aligned}\sum_{n=1}^{\infty} P(A_n) < \infty &\Leftrightarrow P\left(\limsup_{n \rightarrow \infty} A_n\right) = 0 \\ \sum_{n=1}^{\infty} P(A_n) = \infty &\Leftrightarrow P\left(\limsup_{n \rightarrow \infty} A_n\right) = 1.\end{aligned}$$

Poznámka. Implikace „ \Rightarrow “ v první řádce je Cantelliho věta, to už víme, ta platí i bez předpokladu nezávislosti. Nová je implikace „ \Rightarrow “ ve druhé řádce, k jejímu důkazu budeme předpoklad nezávislosti potřebovat.

Borelův 0-1 zákon nám tedy říká, že pokud jsou jevy A_1, A_2, \dots nezávislé, může být pravděpodobnost $\limsup_{n \rightarrow \infty} A_n$ jen 0 nebo 1, nic jiného nemůže nastat. Bez předpokladu

nezávislosti ovšem může být $P\left(\limsup_{n \rightarrow \infty} A_n\right)$ rovna čemukoli z $[0, 1]$ (ověřte).

Důkaz: Potřebujeme dokázat implikaci „ \Rightarrow “ ve druhé řádce. Platí

$$P\left(\limsup_{n \rightarrow \infty} A_n\right) = 1 - P\left(\liminf_{n \rightarrow \infty} A_n^c\right).$$

Bude snažší omezit seshora $P\left(\liminf_{n \rightarrow \infty} A_n^c\right)$ (a ukázat, že je rovno 0). Ze spojitosti pravděpodobnosti a z $B_n = \bigcap_{k=n}^{\infty} A_k^c \nearrow \liminf_{n \rightarrow \infty} A_n^c$ máme

$$P\left(\liminf_{n \rightarrow \infty} A_n^c\right) = P\left(\bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k^c\right) = \lim_{n \rightarrow \infty} P\left(\bigcap_{k=n}^{\infty} A_k^c\right) = \lim_{n \rightarrow \infty} \lim_{N \rightarrow \infty} P\left(\bigcap_{k=n}^N A_k^c\right).$$

V poslední rovnosti opět používáme spojitost pravděpodobnosti. Z předpokladu nezávislosti $\{A_k\}$, a dle věty 1.8 tedy i nezávislosti $\{A_k^c\}$, máme

$$P\left(\bigcap_{k=n}^N A_k^c\right) = \prod_{k=n}^N (1 - P(A_k)) \leq \prod_{k=n}^N e^{-P(A_k)} = e^{-\sum_{k=n}^N P(A_k)},$$

kde v nerovnosti jsme použili odhad $1 - x \leq e^{-x}$, který platí $\forall x \in \mathbb{R}$. Tedy dohromady

$$P\left(\liminf_{n \rightarrow \infty} A_n^c\right) = \lim_{n \rightarrow \infty} \lim_{N \rightarrow \infty} P\left(\bigcap_{k=n}^N A_k^c\right) \leq \lim_{n \rightarrow \infty} \lim_{N \rightarrow \infty} e^{-\sum_{k=n}^N P(A_k)} = \lim_{n \rightarrow \infty} e^{-\sum_{k=n}^{\infty} P(A_k)},$$

což je rovno 0 z předpokladu $\sum_{n=1}^{\infty} P(A_n) = \infty$.

Neboť nic jiného než $\sum_{n=1}^{\infty} P(A_n) < \infty$ nebo $\sum_{n=1}^{\infty} P(A_n) = \infty$ nastat nemůže, jsou obě implikace zároveň i ekvivalence. \square

A jak je to tedy s naší Bernoulliiovskou posloupností?

Příklad. Mějme posloupnost Bernoulliiovských pokusů a jevy $A_n = \{\omega \in \Omega : X_n(\omega) = 1\}$, tedy A_n je jev, že v n -tém pokuse padla 1. Jevy $\{A_n\}$ jsou vzájemně nezávislé a tedy z Borelovy věty je

$$P\left(\limsup_{n \rightarrow \infty} A_n\right) = 1 \quad \text{neboť} \quad \sum_{n=1}^{\infty} P(A_n) = \sum_{n=1}^{\infty} \frac{1}{2} = \infty.$$

Takže pravděpodobnost, že nekonečně krát padla 1, je, ve shodě s intuicí, rovna 1. A samozřejmě

$$P(\text{„padne jen konečně krát 1“}) = 1 - P\left(\limsup_{n \rightarrow \infty} A_n\right) = 0.$$

Pokud nemáme nezávislost a $\sum_{n=1}^{\infty} P(A_n) = \infty$ nevíme o $P\left(\limsup_{n \rightarrow \infty} A_n\right)$ obecně nic!

Příklad. Mějme pravděpodobnostní prostor s $\Omega = [0, 1]$, $\mathcal{A} = \mathcal{B}(\Omega)$ a $P = \lambda$. Buďte $A_i = (0, \frac{1}{i})$. Potom $P(A_i) = \frac{1}{i}$, takže $\sum_{i=1}^{\infty} P(A_i) = \infty$, ale jevy A_i zřejmě nejsou vzájemně nezávislé. Každé ω je jen v konečně mnoha jevech A_i , takže

$$\limsup_{n \rightarrow \infty} A_n = \emptyset \quad \text{a} \quad P\left(\limsup_{n \rightarrow \infty} A_n\right) = P(\emptyset) = 0.$$

V předchozím příkladu jsme znali jevy = množiny A_n přesně. To je výhodné pro výpočet pravděpodobností, ovšem v praxi tato situace typicky nenastává. Mnohem častěji neznáme přesně (Ω, \mathcal{A}, P) a neznáme náhodné veličiny X_n jako funkce z Ω . Známe jen něco o rozdělení X_n , jako třeba v následujícím příkladě:

Příklad. Buďte X_1, X_2, \dots nezávislé náhodné veličiny definované na tomtéž pravděpodobnostním prostoru, o kterých víme, že $E X_n = 0$ a $\text{var } X_n = 1$ pro každé $n \in \mathbb{N}$. Nemusejí mít stejné rozdělení. Jaká je pravděpodobnost jevu, že nastane jen konečně krát $|X_n| \geq n$? To jest kolik je

$$P\left(\liminf_{n \rightarrow \infty} \{|X_n| < n\}\right) = P\left(\liminf_{n \rightarrow \infty} \{\omega \in \Omega : |X_n(\omega)| < n\}\right)?$$

Z Čebyševovy nerovnosti 2.8 víme, že $P(|X_n| \geq n) \leq \frac{\text{var } X_n}{n^2}$, takže

$$\sum_{n=1}^{\infty} P(|X_n| \geq n) \leq \sum_{n=1}^{\infty} \frac{1}{n^2} < \infty.$$

Jevy $\{|X_n| \geq n\}$ jsou nezávislé, takže můžeme použít Borelovu větu (druhá rovnost).

$$P\left(\liminf_{n \rightarrow \infty} \{|X_n| < n\}\right) = 1 - P\left(\limsup_{n \rightarrow \infty} \{|X_n| \geq n\}\right) = 1 - 0 = 1.$$

Toto lze dokázat i za slabších předpokladů (viz cvičení).

Další příklady na použití Cantelliho a Borelovu věty budou na cvičení.

4.2 KONVERGENCE POSLOUPNOSTI NÁHODNÝCH VELIČIN

Vraťme se teď k našemu průměru \bar{X}_n z úvodu kapitoly 4 a jeho konvergenci k $E X_1$. V jakém smyslu by měla posloupnost náhodných veličin \bar{X}_n konvergovat (někam)?

Uvažujme opět naši Bernoulliiovskou posloupnost s pravděpodobností úspěchu $p = \frac{1}{2}$. Naše fyzikální zkušenost nám říká, že „i po velmi velkém množství pokusů má přesná shoda \bar{X}_n s $E X_1$ zanedbatelnou pravděpodobnost“. A co teorie pravděpodobnosti?

Připomeňme si Stirlingovu formuli, tj. tvrzení

$$n! \sim \sqrt{2\pi n} n^n e^{-n} \quad \text{pro } n \rightarrow \infty.$$

Trochu více formálně – platí:

$$\lim_{n \rightarrow \infty} \frac{n!}{\sqrt{2\pi n} n^n e^{-n}} = 1.$$

A s využitím Stirlingovy formule určíme pravděpodobnost, že se \bar{X}_n přesně rovná $E X_1 = \frac{1}{2}$. $\sum_{i=1}^{2n} X_i$ má binomické rozdělení $\text{Binom}(2n, \frac{1}{2})$, takže

$$P\left(\frac{1}{2n} \sum_{i=1}^{2n} X_i = \frac{1}{2}\right) = \binom{2n}{n} 2^{-2n} \sim \frac{(2n)^{2n} \sqrt{2\pi 2n}}{(n^n \sqrt{2\pi n})^2} 2^{-2n} = \frac{1}{\sqrt{\pi n}} \xrightarrow{n \rightarrow \infty} 0.$$

Tedy relativní četnost \bar{X}_n je přesně rovna $\frac{1}{2}$ jen se zanedbatelnou pravděpodobností i pro velká n . Zkušenost ukazuje, že relativní četnost \bar{X}_n se v nejlepším případě „ustálí blízko $\frac{1}{2}$ “. Požadujeme tedy radši místo přesné shody, aby

$$\forall \epsilon > 0 \quad P\left(|\bar{X}_n - E X_1| \leq \epsilon\right) \xrightarrow{n \rightarrow \infty} 1, \quad (4.1)$$

„průměr byl blízko střední hodnotě s velkou pravděpodobností pro velké n “.

Definice 4.1 Buďte Y_1, Y_2, \dots náhodné veličiny na stejném pravděpodobnostním prostoru (Ω, \mathcal{A}, P) . Řekneme, že posloupnost náhodných veličin $\{Y_n\}_{n=1}^{\infty}$ *konverguje v pravděpodobnosti* k náhodné veličině Y (značíme $Y_n \xrightarrow[n \rightarrow \infty]{P} Y$), pokud

$$\lim_{n \rightarrow \infty} P(|Y_n - Y| > \epsilon) = 0, \quad \forall \epsilon > 0.$$

Poznámka. Tuto konvergenci už známe z teorie míry – je to konvergence posloupnosti funkcí $Y_n : \Omega \rightarrow \mathbb{R}$ v míře P . A je to přesně ten způsob konvergence, který jsme požadovali po \bar{X}_n v (4.1).

Definice 4.2 Buďte Y_1, Y_2, \dots náhodné veličiny na stejném pravděpodobnostním prostoru (Ω, \mathcal{A}, P) . Řekneme, že posloupnost náhodných veličin $\{Y_n\}_{n=1}^{\infty}$ *konverguje P-skoro jistě* k náhodné veličině Y (značíme $Y_n \xrightarrow[n \rightarrow \infty]{sj} Y$), pokud

$$P\left(\left\{\omega \in \Omega : \lim_{n \rightarrow \infty} Y_n(\omega) = Y(\omega)\right\}\right) = 1.$$

Poznámka. I tuto konvergenci už známe z teorie míry – je to konvergence posloupnosti funkcí $Y_n : \Omega \rightarrow \mathbb{R}$ P-skoro všude.

Obě představené konvergence posloupnosti náhodných veličin jsou tedy opravdu konvergence posloupnosti funkcí na Ω . Všechny zúčastněné náhodné veličiny musí být proto definované na stejném pravděpodobnostním prostoru Ω , a k rozhodnutí o konvergenci nám nestačí znát jen rozdělení P_{Y_n} jednotlivých náhodných veličin Y_n , musíme je opravdu znát jako funkce z Ω . Alternativně musíme znát sdružené rozdělení celé posloupnosti $\{Y_n\}_{n=1}^{\infty}$ – viz sekce 4.3 a 4.4

Z teorie míry také známe některé vlastnosti konvergence v pravděpodobnosti a P-skoro jistě. Zopakujme si ty, které budeme dále potřebovat.

Tvrzení 4.3 Buďte Y_1, Y_2, \dots náhodné veličiny na stejném pravděpodobnostním prostoru. Pak $Y_n \xrightarrow[n \rightarrow \infty]{sj} Y \Rightarrow Y_n \xrightarrow[n \rightarrow \infty]{P} Y$.

Obráceně tvrzení neplatí!

Tvrzení 4.4 Buď $1 \leq p \leq \infty$ a buďte Y_1, Y_2, \dots a Y náhodné veličiny na stejném pravděpodobnostním prostoru, patřící do L^p . Pak $Y_n \xrightarrow[n \rightarrow \infty]{L^p} Y \Rightarrow Y_n \xrightarrow[n \rightarrow \infty]{P} Y$.

Konvergence P-skoro jistě ale z L^p konvergence neplyne.

Příklad. Buď $\Omega = [0, 1]$, $\mathcal{A} = \mathcal{B}(\Omega)$, $P = \lambda$ a náhodné veličiny $X_1 = 1$, $X_2 = \mathbb{1}_{[0, \frac{1}{2}]}$, $X_3 = \mathbb{1}_{[\frac{1}{2}, 1]}$, $X_4 = \mathbb{1}_{[0, \frac{1}{4}]}$, $X_5 = \mathbb{1}_{[\frac{1}{4}, \frac{1}{2}]}$, \dots . Platí

- $\forall \omega$ existuje nekonečně mnoho indexů takových, že $X_n(\omega) = 1$, takže X_n nekonečně často konverguje k 0 P-skoro jistě.
- Pro $\epsilon \in (0, 1)$ je $P(|X_n| > \epsilon) = P(X_n = 1) \xrightarrow[n \rightarrow \infty]{} 0$, takže $X_n \xrightarrow[n \rightarrow \infty]{P} 0$.
- Také $E|X_n - 0|^p \xrightarrow[n \rightarrow \infty]{} 0$, takže $X_n \xrightarrow[n \rightarrow \infty]{L^p} 0$ pro každé $p \in [1, \infty]$.

A při jakých operacích s náhodnými veličinami se konvergence zachovávají?

Věta 4.5 (o konvergenci v pravděpodobnosti a součtu) Buďte Y_1, Y_2, \dots a Z_1, Z_2, \dots náhodné veličiny na stejném pravděpodobnostním prostoru (Ω, \mathcal{A}, P) . Pak platí

$$Y_n \xrightarrow[n \rightarrow \infty]{P} 0 \quad \text{a} \quad Z_n \xrightarrow[n \rightarrow \infty]{P} 0 \Rightarrow (Y_n + Z_n) \xrightarrow[n \rightarrow \infty]{P} 0.$$

Buď navíc $\{a_n\}_{n=1}^{\infty}$ reálná omezená posloupnost. Pak platí

$$Y_n \xrightarrow[n \rightarrow \infty]{P} 0 \Rightarrow a_n Y_n \xrightarrow[n \rightarrow \infty]{P} 0.$$

Důkaz: Buď $\epsilon > 0$. Pak

$$P(|Y_n + Z_n| > \epsilon) \leq P\left(|Y_n| > \frac{\epsilon}{2}\right) + P\left(|Z_n| > \frac{\epsilon}{2}\right) \xrightarrow[n \rightarrow \infty]{} 0, \quad \text{z předpokladů } \forall \epsilon.$$

Buď $C > 0$ takové, že $|a_n| < C$, $\forall n \in \mathbb{N}$. Pak

$$P(|a_n Y_n| > \epsilon) \leq P\left(|Y_n| > \frac{\epsilon}{C}\right) \xrightarrow[n \rightarrow \infty]{} 0, \quad \text{z předpokladů } \forall \epsilon.$$

□

Poznámka. Pro konvergenci skoro jistě platí tvrzení věty také (ověřte).

Věta 4.6 (o spojitě transformaci) Buďte Y_1, Y_2, \dots náhodné veličiny na stejném pravděpodobnostním prostoru (Ω, \mathcal{A}, P) a $g : \mathbb{R} \rightarrow \mathbb{R}$ funkce spojitá na S_Y , nosiči rozdělení Y (tj. platí $P(Y \in S_Y) = 1$). Pak

$$(i) \quad Y_n \xrightarrow[n \rightarrow \infty]{P} Y \Rightarrow g(Y_n) \xrightarrow[n \rightarrow \infty]{P} g(Y)$$

$$(ii) \quad Y_n \xrightarrow[n \rightarrow \infty]{sj} Y \Rightarrow g(Y_n) \xrightarrow[n \rightarrow \infty]{sj} g(Y).$$

Důkaz: (i) Důkaz je technický, vynecháme z časových důvodů.

(ii) Buď N množina splňující $P(N) = 0$ a $Y_n(\omega) \xrightarrow[n \rightarrow \infty]{} Y(\omega)$ na $\Omega \setminus N$. Pak pro každé $\omega \in M = Y^{-1}(S_Y) \cap (\Omega \setminus N)$ uplatníme větu o limitě spojitě transformované posloupnosti a $P(M) = 1$ z předpokladů. \square

Poznámka. V předpokladech věty se implicitně požaduje, aby nosič S_Y byl otevřená množina – neboť požadují spojitou funkci g na celém S_Y . V případě, že $P(Y = c) = 1$, se tedy požaduje, aby g byla spojitá na S_Y , což je nějaké otevřené okolí bodu c .

Poznámka. Lze zadefinovat i vektorové verze konvergenčí (tj. pro náhodné vektory) v pravděpodobnosti a skoro jistě. Stačí v definici místo $|Y_n - Y|$ použít $\|Y_n - Y\|$, eukleidovskou normu na \mathbb{R}^k .

4.3 SLABÝ ZÁKON VELKÝCH ČÍSEL

A jak je to tedy s konvergencí průměrů \bar{X}_n ? Díky Čebyševově nerovnosti můžeme ukázat následující:

Věta 4.7 (Čebyševův slabý zákon velkých čísel) Buďte X_1, X_2, \dots vzájemně nezávislé náhodné veličiny splňující $E X_n^2 < \infty, \forall n \in \mathbb{N}$. Nechť

$$\lim_{n \rightarrow \infty} \left(\frac{1}{n^2} \sum_{i=1}^n \text{var } X_i \right) = 0. \quad (4.2)$$

Potom platí $|\bar{X}_n - E \bar{X}_n| \xrightarrow[n \rightarrow \infty]{P} 0$.

Důkaz: Použijeme postupně Čebyševovu nerovnost 2.8 a vzájemnou nezávislost $\{X_i\}$:

$$P(|\bar{X}_n - E \bar{X}_n| > \epsilon) \leq \frac{\text{var } \bar{X}_n}{\epsilon^2} = \frac{1}{\epsilon^2} \frac{1}{n^2} \sum_{i=1}^n \text{var } X_i,$$

což konverguje k 0, pro $n \rightarrow \infty$, pro každé $\epsilon > 0$ z předpokladu (4.2). \square

Poznámka. Ve větě stačí místo nezávislosti předpokládat jen nekorelovanost X_i a X_j pro každé $i \neq j$.

Slabý zákon velkých čísel (budeme zkracovat jako ZVČ) tedy říká, že při splnění podmínky na rozptyly, částečné průměry centrované posloupnosti $(X_n - E X_n)$ konvergují v pravděpodobnosti k 0. Posun všech X_n o střední hodnotu je potřeba, abychom průměrovali jen náhodné variace, nikoli systematické (nenáhodné) vychýlení $E X_n$.

A co nám ZVČ říká o Bernoulliiovské posloupnosti?

Příklad. Buď $\{X_n\}_{n=1}^{\infty}$ Bernoulliiovská posloupnost s pravděpodobností úspěchu p . Tedy X_n jsou vzájemně nezávislé, alternativně rozdělené, $E X_n^2 < \infty, \forall n \in \mathbb{N}$ a $E X_n = p, \text{var } X_n = p(1 - p)$. Podmínka (4.2) je splněna, neboť $\frac{1}{n^2}(np(1 - p)) \xrightarrow{n \rightarrow \infty} 0$.

$\{X_n\}_{n=1}^{\infty}$ tedy splňuje Čebyševův ZVČ a platí $|\bar{X}_n - p| \xrightarrow[n \rightarrow \infty]{P} 0$, neboli

$$P(|\bar{X}_n - p| \leq \epsilon) \xrightarrow[n \rightarrow \infty]{} 1, \quad \forall \epsilon > 0.$$

Takže opravdu platí, že „průměr \bar{X}_n je blízko střední hodnotě $E X_1$ s velkou pravděpodobností pro velké n “.

Poznámka. Pokud jsou náhodné veličiny X_n nezávislé, stejně rozdělené, s $\text{var } X_n < \infty$, pak splňují Čebyševův ZVČ. Dokonce by stačilo předpokládat jen $E X_n = \mu \in \mathbb{R}$ (nebudeme dokazovat).

Pokud je ale $E |X_n| = \infty$, pak tvrzení ZVČ platit nemusí. Uvažujme například $\{X_n\}_{n=1}^{\infty}$ nezávislé, stejně rozdělené s Cauchyho rozdělením (neboli t_1 -rozdělení, viz str.43). Pak $E X_n$ neexistuje a platí $\frac{1}{n} \sum_{i=1}^n X_i \sim \text{Cauchy}$ (ověřte, resp. viz cvičení). Tedy \bar{X}_n určitě nekonverguje ke konstantě.

*Zde končí
předn. 14
(3.4.)*

4.4 SILNÝ ZÁKON VELKÝCH ČÍSEL

V předchozí sekci jsme si dokázali, že pro Bernoulliiovskou posloupnost s $p = \frac{1}{2}$ platí, že četnost úspěchů \bar{X}_n bude pro velké n blízko $p = \frac{1}{2} = E X_1$ s velkou pravděpodobností. Ale my bychom chtěli ukázat ještě víc: když hodíme např. 100× mincí, tak se může stát, že s malou pravděpodobností se \bar{X}_n bude hodně lišit od $\frac{1}{2}$. Ale tato deviace by pro velké n měla postupně vymizet. A o tom mluví silný zákon velkých čísel.

Terminologie. Řekneme, že posloupnost náhodných veličin $\{X_n\}_{n=1}^{\infty}$ splňuje silný zákon velkých čísel (zkráceně SZVČ), pokud platí $|\bar{X}_n - E \bar{X}_n| \xrightarrow[n \rightarrow \infty]{\text{sj}} 0$.

Věta 4.8 (Silný zákon velkých čísel pro nestejně rozdělené náhodné veličiny) Buď $\{X_n\}_{n=1}^{\infty}$ posloupnost vzájemně nezávislých náhodných veličin, splňujících $E X_n^2 < \infty, \forall n \in \mathbb{N}$. Nechť

$$\sum_{n=1}^{\infty} \frac{\text{var } X_n}{n^2} < \infty. \quad (4.3)$$

Potom platí $|\bar{X}_n - E \bar{X}_n| \xrightarrow[n \rightarrow \infty]{\text{sj}} 0$.

K důkazu silného zákona velkých čísel budeme potřebovat následující zobecnění Čebyševovy nerovnosti.

Věta 4.9 (Kolmogorovova nerovnost) Buďte X_1, X_2, \dots vzájemně nezávislé náhodné veličiny, splňující $E X_n^2 < \infty, \forall n \in \mathbb{N}$. Označme $S_k = \sum_{j=1}^k X_j, k \in \mathbb{N}$. Pak

$$P\left(\max_{1 \leq k \leq n} |S_k - E S_k| \geq \epsilon\right) \leq \frac{\text{var } S_n}{\epsilon^2} = \frac{\sum_{k=1}^n \text{var } X_k}{\epsilon^2}, \quad \forall \epsilon > 0, \forall n \in \mathbb{N}. \quad (4.4)$$

Důkaz: Bez újmy na obecnosti předpokládejme $E X_j = 0, j \in \mathbb{N}$ (kdyby ne, aplikujeme následující na $(X_j - E X_j)$).

Pro $k \in \mathbb{N}$ označme

$$A_k = \{\max_{1 \leq j \leq k} |S_j| < \epsilon\} \quad B_k = \bigcap_{j=1}^{k-1} \{|S_j| < \epsilon\} \cap \{|S_k| \geq \epsilon\}.$$

Jevy B_k jsou vzájemně disjunktní a platí $A_n^c = \{\max_{1 \leq k \leq n} |S_k| \geq \epsilon\} = \bigcup_{k=1}^n B_k$. Jev B_k odpovídá dosažení hladiny ϵ v posloupnosti $\{|S_j|\}$ poprvé přesně v kroku k . Počítejme

$$\begin{aligned} \int_{B_k} S_n^2 dP &= E(S_n \mathbb{1}_{B_k})^2 = E((S_n - S_k) \mathbb{1}_{B_k} + S_k \mathbb{1}_{B_k})^2 \\ &= E(S_n - S_k)^2 \mathbb{1}_{B_k} + 2E(S_n - S_k)S_k \mathbb{1}_{B_k} + E S_k^2 \mathbb{1}_{B_k} \geq E S_k^2 \mathbb{1}_{B_k}, \end{aligned}$$

neboť $E(S_n - S_k)^2 \mathbb{1}_{B_k} \geq 0$ a

$$E(S_n - S_k)S_k \mathbb{1}_{B_k} = E(S_n - S_k) E S_k \mathbb{1}_{B_k} = 0 \cdot E S_k \mathbb{1}_{B_k} = 0.$$

V první rovnosti jsme využili, že $S_n - S_k = \sum_{j=k+1}^n X_j$ je funkce jen náhodných veličin $\{S_j\}_{j=k+1}^n$ a tedy je nezávislá s $S_k \mathbb{1}_{B_k}$, které závisí zase jen na $\{S_j\}_{j=1}^k$.

Platí $B_k \subseteq \{|S_k| \geq \epsilon\}$ a tedy

$$E S_k^2 \mathbb{1}_{B_k} \geq E \epsilon^2 \mathbb{1}_{B_k} = \epsilon^2 P(B_k).$$

Sečtením přes všechna $k = 1, \dots, n$ dostaneme

$$\int_{A_n^c} S_n^2 dP \geq \epsilon^2 P(A_n^c).$$

Zřejmě také platí

$$\int_{A_n^c} S_n^2 dP \leq E S_n^2 = \text{var } S_n = \sum_{k=1}^n \text{var } X_k,$$

protože $\{X_k\}$ jsou vzájemně nezávislé. Spojením obou nerovností dostáváme $\epsilon^2 P(A_n^c) \leq \text{var } S_n$, což jsme měli dokázat. \square

Poznámka. Kolmogorovova nerovnost není užitečná jen pro důkaz SZVČ. Mějme posloupnost nezávislých stejně rozdělených náhodných veličin $\{X_j\}_{j=1}^{\infty}$ s $E X_1 = 0$. Posloupnost částečných součtů $\{S_k\}_{k=0}^{\infty}$ (kde $S_k = \sum_{j=1}^k X_j$, $k \in \mathbb{N}_0$) pak definuje náhodný proces, kterému se říká náhodná procházka. Index k má roli času a X_k odpovídají přírůstkům náhodné procházky za jeden časový úsek. Kolmogorovova nerovnost pak dává horní mez na pravděpodobnost, že do času n dosáhne náhodná procházka (v absolutní hodnotě) nějaké hladiny ϵ .

Druhé důležité tvrzení, které potřebujeme k důkazu silného zákona velkých čísel, je Cantelliho věta, ale tu už známe.

Důkaz: (Silného zákona velkých čísel)

Posloupnost $\{X_n\}_{n=1}^{\infty}$ splňuje předpoklady Kolmogorovovy nerovnosti, a tak pro každé $n \in \mathbb{N}$, $\epsilon > 0$ platí

$$P\left(\max_{1 \leq k \leq n} |S_k - E S_k| \geq \epsilon n\right) \leq \frac{\text{var } S_n}{\epsilon^2 n^2}, \quad (4.5)$$

kde jsme do (4.4) místo ϵ použili $\tilde{\epsilon} = \epsilon n$. Posčítejme nerovnosti (4.5), ale přes 2^n místo přes n . Dostaneme:

$$\sum_{n=1}^{\infty} P\left(\max_{1 \leq k \leq 2^n} |S_k - E S_k| \geq \epsilon 2^n\right) \leq \sum_{n=1}^{\infty} \frac{\text{var } S_{2^n}}{\epsilon^2 2^{2n}}. \quad (4.6)$$

Pokud dokážeme, že suma vpravo je konečná, pak nám Cantelliho věta dá

$$P\left(\limsup_{n \rightarrow \infty} \left\{ \max_{1 \leq k \leq 2^n} |S_k - E S_k| \geq \epsilon 2^n \right\}\right) = 0, \quad \forall \epsilon > 0.$$

To jest $\max_{1 \leq k \leq 2^n} |S_k - E S_k| \leq \epsilon 2^n$ platí s pravděpodobností 1 až na konečně mnoho výjimek, a $\limsup_{n \rightarrow \infty} \frac{|S_{2^n} - E S_{2^n}|}{2^n} \leq \epsilon$ s pravděpodobností 1 pro libovolné $\epsilon > 0$.

A co zbylé indexy různé od 2^n ? Pro ty použijeme „sendvičový trik“. Pro každé $m \in \mathbb{N}$ existuje n tak, že $2^n \leq m \leq 2^{n+1}$. Potom $|S_m - E S_m| \leq \max_{1 \leq k \leq 2^{n+1}} |S_k - E S_k|$, tedy s pravděpodobností 1

$$|S_m - E S_m| \leq \epsilon 2^{n+1} \leq 2\epsilon m,$$

pro všechna dost velká m . Neboli existuje množina $N(\epsilon) \subset \Omega$, $P(N(\epsilon)) = 0$ taková, že $\forall \omega \notin N(\epsilon) \exists M(\omega) \in \mathbb{N} : \forall m \geq M(\omega) : \frac{|S_m(\omega) - E S_m|}{m} \leq 2\epsilon$. A tedy

$$\limsup_{m \rightarrow \infty} \frac{|S_m(\omega) - E S_m|}{m} \leq 2\epsilon \quad \forall \omega \notin N(\epsilon).$$

Ze σ -aditivity pravděpodobnosti pro $N = \bigcup_{l \in \mathbb{N}} N(\frac{1}{l})$ platí $P(N) \leq \sum_{l \in \mathbb{N}} P(N(\frac{1}{l})) = 0$, tedy pro každé $\omega \notin N$

$$\limsup_{m \rightarrow \infty} \frac{|S_m(\omega) - E S_m|}{m} \leq 2\frac{1}{l}, \quad \forall l \in \mathbb{N} \quad \Rightarrow \quad \limsup_{m \rightarrow \infty} \frac{|S_m(\omega) - E S_m|}{m} = 0,$$

a $\lim_{m \rightarrow \infty} \frac{|S_m(\omega) - \mathbb{E} S_m|}{m} = 0$ pro všechna $\omega \in N^c$, tedy skoro jistě, což jsme měli dokázat.

K dokončení důkazu nám chybí ukázat konvergenci řady z odhadu (4.6).

$$\begin{aligned} \sum_{n=1}^{\infty} \frac{\text{var } S_{2^n}}{2^{2n}} &= \sum_{n=1}^{\infty} \frac{1}{2^{2n}} \sum_{k=1}^{2^n} \text{var } X_k = \sum_{n=1}^{\infty} \frac{1}{2^{2n}} (\text{var } X_1 + \text{var } X_2) + \sum_{n=2}^{\infty} \frac{1}{2^{2n}} \sum_{k=3}^{2^n} \text{var } X_k \\ &= \frac{1}{3} (\text{var } X_1 + \text{var } X_2) + \sum_{n=2}^{\infty} \frac{1}{2^{2n}} \sum_{v=1}^{n-1} \sum_{k=2^v+1}^{2^{v+1}} \text{var } X_k, \end{aligned} \quad (4.7)$$

První člen je konečný, druhý budeme dále upravovat

$$\begin{aligned} \sum_{n=2}^{\infty} \frac{1}{2^{2n}} \sum_{v=1}^{n-1} \sum_{k=2^v+1}^{2^{v+1}} \text{var } X_k &= \sum_{v=1}^{\infty} \left(\sum_{n=v+1}^{\infty} \frac{1}{2^{2n}} \right) \sum_{k=2^v+1}^{2^{v+1}} \text{var } X_k \\ &\leq \sum_{v=1}^{\infty} \left(\sum_{n=v+1}^{\infty} \frac{1}{4^n} \right) \sum_{k=2^v+1}^{2^{v+1}} \text{var } X_k \left(\frac{2^{v+1}}{k} \right)^2 \\ &= \sum_{v=1}^{\infty} \left(\sum_{n=1}^{\infty} \frac{1}{4^n} \right) \sum_{k=2^v+1}^{2^{v+1}} \frac{4 \text{var } X_k}{k^2} = \frac{4}{3} \sum_{k=3}^{\infty} \frac{\text{var } X_k}{k^2} < \infty, \end{aligned}$$

z předpokladu (4.3). I v tomto odhadu jsme vlastně použili sendvičový postup - sumu jsme rozdělili na úseky s rostoucím počtem členů 2^v a v každém z těchto úseků můžeme použít že $1 \leq \frac{2^{v+1}}{k}$, což jsme udělali ve druhé řádce. \square

Poznámka. Z rozpisu (4.7) lze nahlédnout, že kdyby byly všechny $\text{var } X_k$ stejné, tak konvergenci řady z důkazu dokážeme snadno přímým dosazením. A také je vidět, proč potřebujeme úseky zvětšující se délky 2^n . Právě proto, aby řada rozptylů, která nám zajistí platnost Cantelliho věty, mohla konvergovat. Kdybychom sčítali přes n , a nikoli přes 2^n , konvergenci bychom nedokázali.

Poznámka. Podmínka $\sum_{n=1}^{\infty} \frac{\text{var } X_n}{n^2} < \infty$ je postačující pro SZVČ, nikoli nutná. Existuje posloupnost $\{X_n\}_{n=1}^{\infty}$ nezávislých náhodných veličin splňující $\sum_{n=1}^{\infty} \frac{\text{var } X_n}{n^2} = \infty$, $\mathbb{E} X_n = 0$,

a $\mathbb{P}(\lim_{n \rightarrow \infty} \bar{X}_n = 0) = 1$, tj. $\bar{X}_n \xrightarrow[n \rightarrow \infty]{\text{sj}} 0$.

Abychom ji zkonstruovali volme X_n nezávislé s

$$\mathbb{P}(X_n = -c_n) = d_n, \quad \mathbb{P}(X_n = 0) = 1 - 2d_n, \quad \mathbb{P}(X_n = c_n) = d_n, \quad n \in \mathbb{N},$$

kde

$$\sum_{n=1}^{\infty} d_n < \infty \quad \text{a} \quad 2c_n^2 d_n = \text{var } X_n.$$

Takové $\{c_n\}_{n=1}^{\infty}$ a $\{d_n\}_{n=1}^{\infty}$ vždy existují – např.

$$d_n = \frac{1}{n^2}, \quad \sigma_n^2 = n^2, \quad \text{pak} \quad c_n = \sqrt{\frac{n^2 \text{var } X_n}{2}} = \frac{n^2}{\sqrt{2}} \quad (\text{tedy docela velké}).$$

Potom z Cantelliho věty $\mathbb{P}(\limsup_{n \rightarrow \infty} (X_n \neq 0)) = 0$ a tedy $\mathbb{P}(\lim_{n \rightarrow \infty} \bar{X}_n = 0) = 1$.

A co nám SZVČ říká o Bernoulliiovské posloupnosti?

Příklad. Buď $\{X_n\}_{n=1}^{\infty}$ Bernoulliiovská posloupnost s pravděpodobností úspěchu p . Neboli X_n jsou vzájemně nezávislé, alternativně rozdělené, $E X_n^2 < \infty, \forall n \in \mathbb{N}$ a $E X_n = p, \text{var } X_n = p(1-p)$. Podmínka (4.3) je splněna, neboť

$$\sum_{n=1}^{\infty} \frac{p(1-p)}{n^2} = p(1-p) \sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{\pi^2 p(1-p)}{6} < \infty.$$

Tedy $\{X_n\}_{n=1}^{\infty}$ splňuje SZVČ 4.8 a platí $|\bar{X}_n - p| \xrightarrow[n \rightarrow \infty]{\text{sj}} 0$, neboli

$$P\left(\left\{\omega : \lim_{n \rightarrow \infty} \bar{X}_n(\omega) = p\right\}\right) = 1.$$

Takže až na množinu výjimečných ω , která má pravděpodobnost 0, pro všechna ostatní ω průměr $\bar{X}_n(\omega)$ konverguje ke střední hodnotě $E X_1$. Přesně ve shodě s naší fyzikální zkušeností tedy ty případy, ve kterých s malou pravděpodobností mohou být \bar{X}_n velmi daleko od p , s $n \rightarrow \infty$ opravdu „vymizí“.

Poznámka. ! Pozor ! SZVČ neříká, že

$$\text{počet úspěchů v } n \text{ pokusech} \longrightarrow \text{pravděpodobnost úspěchu} \times n,$$

neboť to na levé straně šipky je posloupnost náhodných veličin, která konverguje skoro jistě k ∞ , a to na pravé straně šipky je posloupnost čísel, které konvergují k ∞ . Tedy té šipce není možné přiřadit matematický smysl!

Zřejmě pro libovolnou posloupnost nezávislých, stejně rozdělených náhodných veličin $\{X_n\}_{n=1}^{\infty}$ na tomtež pravděpodobnostním prostoru, s konečnou střední hodnotou $E X_1 = \mu$ a konečným rozptylem, jsou předpoklady SZVČ 4.8 splněny a platí $\bar{X}_n \xrightarrow[n \rightarrow \infty]{\text{sj}} \mu$.

Nicméně v případě nezávislých, stejně rozdělených n.v. je možné předpoklady oslabit a ukázat, že $\{X_n\}_{n=1}^{\infty}$ splňuje SZVČ i v případě, kdy víme jen, že $X_1 \in L^1$.

Věta 4.10 (Silný zákon velkých čísel pro stejně rozdělené náhodné veličiny, L_1 verze) Buď $\{X_n\}_{n=1}^{\infty}$ posloupnost vzájemně nezávislých, stejně rozdělených náhodných veličin. Pak

$$\bar{X}_n \xrightarrow[n \rightarrow \infty]{\text{sj}} \mu \text{ pro nějaké } \mu \in \mathbb{R} \iff E |X_1| < \infty.$$

V takovém případě $E X_1 = \mu$.

K důkazu budeme potřebovat následující lemma.

Lemma 4.11 Pro libovolnou náhodnou veličinu X platí

$$\sum_{n=0}^{\infty} P(|X| \geq n) = \sum_{n=0}^{\infty} (n+1) P(n \leq |X| < n+1), \quad (4.8)$$

a

$$\sum_{n=0}^{\infty} P(|X| \geq n) - 1 \leq E|X| \leq \sum_{n=0}^{\infty} P(|X| \geq n). \quad (4.9)$$

Důkaz: $E|X|$ si můžeme přepsat jako integrál vzhledem k rozdělení P_X a ten rozdělit na integrály přes disjunktní intervaly $[n, n+1)$, na kterých odhadneme integrand $|x|$ shora

$$E|X| = \int_{\mathbb{R}} |x| dP_X(x) = \sum_{n=0}^{\infty} \int_{[n, n+1)} |x| dP_X(x) \leq \sum_{n=0}^{\infty} (n+1) P(n \leq |X| < n+1),$$

i zespoda

$$\sum_{n=0}^{\infty} \int_{[n, n+1)} |x| dP_X(x) \geq \sum_{n=0}^{\infty} n P(n \leq |X| < n+1) = \sum_{n=0}^{\infty} (n+1) P(n \leq |X| < n+1) - 1.$$

Zbývá dokázat první část lemmatu, což uděláme pomocí prohození pořadí sčítání ve dvojné sumě:

$$\begin{aligned} \sum_{n=0}^{\infty} P(|X| \geq n) &= \sum_{n=0}^{\infty} \sum_{k=n}^{\infty} P(k \leq |X| < k+1) \\ &= \sum_{k=0}^{\infty} \sum_{n=0}^k P(k \leq |X| < k+1) = \sum_{k=0}^{\infty} (k+1) P(k \leq |X| < k+1). \end{aligned}$$

□

Důkaz: (Silného zákona velkých čísel pro stejně rozdělené náhodné veličiny)

Nejdříve dokážeme implikaci " \Leftarrow ".

K tomu chceme použít SZVČ 4.8, chybí nám ale předpoklad konečného rozptylu. Proto si zavedeme omezené náhodné veličiny

$$Y_n = X_n \mathbb{1}_{\{|X_n| < n\}}, \quad n \in \mathbb{N}.$$

Ty budou také vzájemně nezávislé a zřejmě

$$P(X_n \neq Y_n) = P(|X_n| \geq n) = P(|X_1| \geq n).$$

Z konečnosti $E|X_1|$ a lemmatu 4.11 máme $\sum_{n=1}^{\infty} P(|X_1| \geq n) < \infty$, takže z Cantelliho věty plyne $P(\limsup_{n \rightarrow \infty} \{X_n \neq Y_n\}) = 0$. Tedy $X_n = Y_n$ s pravděpodobností 1 až na konečně

mnoho n . Proto $(\bar{X}_n - \bar{Y}_n) \xrightarrow[n \rightarrow \infty]{sj} 0$. Stačí tedy dokázat, že SZVČ splňují $\{Y_n\}_{n=1}^{\infty}$, a že $\lim_{n \rightarrow \infty} E \bar{Y}_n = \mu$. Potom už bude platit

$$\bar{X}_n - \mu = (\bar{X}_n - \bar{Y}_n) + (\bar{Y}_n - E \bar{Y}_n) + (E \bar{Y}_n - \mu) \xrightarrow[n \rightarrow \infty]{sj} 0 + 0 + 0 = 0.$$

Ovšem

$$E Y_n = E X_n \mathbb{1}_{\{|X_n| < n\}} = E X_1 \mathbb{1}_{\{|X_1| < n\}} \xrightarrow[n \rightarrow \infty]{} E X_1 = \mu,$$

z Lebesgueovy věty (konvergentní majoranta je $|X_1|$). Tedy také $\frac{1}{n} \sum_{j=1}^n E Y_j \xrightarrow[n \rightarrow \infty]{} \mu$.

Zbývá tedy ukázat, že $\{Y_n\}_{n=1}^{\infty}$ splňují předpoklady věty 4.8:

- Jsou vzájemně nezávislé, neboť $\{X_n\}_{n=1}^{\infty}$ jsou vzájemně nezávislé (věta 3.8).
- $E |Y_n|^2 < \infty$, neboť Y_n jsou omezené. Navíc platí $\text{var } Y_n \leq E |Y_n|^2$.
- Stačí tedy ukázat $\sum_{n=1}^{\infty} \frac{E Y_n^2}{n} < \infty$, neboli $\int \sum_{n=1}^{\infty} \frac{X_1^2 \mathbb{1}_{\{|X_1| < n\}}}{n^2} dP < \infty$.

Poslední integrál omezíme seshora následujícím způsobem:

$$\begin{aligned} \int \sum_{n=1}^{\infty} \frac{X_1^2 \mathbb{1}_{\{|X_1| < n\}}}{n^2} dP &= \sum_{j=0}^{\infty} \int \sum_{n=1}^{\infty} \frac{1}{n^2} X_1^2 \mathbb{1}_{\{|X_1| < n\}} \mathbb{1}_{\{j \leq |X_1| < j+1\}} dP \\ &= \sum_{j=0}^{\infty} \int \sum_{n=j+1}^{\infty} \frac{1}{n^2} X_1^2 \mathbb{1}_{\{|X_1| < n\}} \mathbb{1}_{\{j \leq |X_1| < j+1\}} dP \\ &\leq \sum_{j=0}^{\infty} (j+1)^2 \int \sum_{n=j+1}^{\infty} \frac{1}{n^2} \mathbb{1}_{\{|X_1| < n\}} \mathbb{1}_{\{j \leq |X_1| < j+1\}} dP \\ &\leq 2 \sum_{j=0}^{\infty} (j+1) P(j \leq |X_1| < j+1) = 2 \sum_{j=0}^{\infty} P(|X_1| \geq j) \\ &\leq 2(E |X_1| + 1) < \infty. \end{aligned}$$

Použili jsme nerovnosti

$$\sum_{n=1}^{\infty} \frac{1}{n^2} < 2, \quad \text{a} \quad \sum_{n=j+1}^{\infty} \frac{1}{n^2} < \int_j^{\infty} \frac{1}{x^2} dx = \frac{1}{j} \leq \frac{2}{j+1}, \quad j \in \mathbb{N},$$

a nakonec lemma 4.11. Tedy implikace " \Leftarrow " je dokázána.

Dokažme nyní " \Rightarrow ".

Můžeme psát $X_n = n\bar{X}_n - (n-1)\bar{X}_{n-1}$, takže

$$\frac{X_n}{n} = \bar{X}_n - \frac{n-1}{n} \bar{X}_{n-1} \xrightarrow[n \rightarrow \infty]{sj} \mu - \mu = 0,$$

z předpokladu věty. To ovšem implikuje

$$P\left(\frac{|X_n|}{n} \geq 1 \text{ pro nekonečně mnoho } n\right) = P(\limsup_{n \rightarrow \infty} \{|X_n| \geq n\}) = 0.$$

Jevy $\{|X_n| \geq n\}$ jsou nezávislé, neboť $\{X_n\}$ jsou vzájemně nezávislé a Borelova věta nám dá, že

$$\sum_{n=1}^{\infty} P(|X_n| \geq n) = \sum_{n=1}^{\infty} P(|X_1| \geq n) < \infty.$$

Rovnost sum plyne z toho, že všechny X_n jsou stejně rozdělené. Z lemmatu 4.11 máme $E|X_1| < \infty$, takže platí i implikace " \Leftarrow " a $\bar{X}_n \xrightarrow[n \rightarrow \infty]{sj} E X_1$ a z jednoznačnosti limity skoro jistě $E X_1 = \mu$. \square

Poznámka. Shrňme si, co jsme zatím zjistili: pro nezávislé, stejně rozdělené náhodné veličiny $\{X_n\}_{n=1}^{\infty}$ platí

- $E|X_1| < \infty \xRightarrow{SZVČ} \bar{X}_n \xrightarrow[n \rightarrow \infty]{sj} \mu$
- $\text{var } X_1 = \sigma^2 < \infty \xRightarrow{SZVČ} \overline{X_n^2} = \frac{1}{n} \sum_{i=1}^n X_i^2 \xrightarrow[n \rightarrow \infty]{sj} (\sigma^2 + \mu^2) = E X_1^2$

A tedy

$$\overline{X_n^2} - (\bar{X}_n)^2 \xrightarrow[n \rightarrow \infty]{sj} \sigma^2,$$

takže pokud bychom chtěli odhadnout rozptyl σ^2 z opakovaných pozorování, šlo by to udělat.

A také víme:

- (i) $(\bar{X}_n - \mu) \xrightarrow[n \rightarrow \infty]{sj} 0$ ze SZVČ.
- (ii) Z Čebyševovy nerovnosti $P(|\bar{X}_n - \mu| > \epsilon) \leq \frac{\text{var } \bar{X}_n}{\epsilon^2} = \frac{\sigma^2}{n\epsilon^2}$, neboli „pravděpodobnost velké odchylky je nejvýše řádu $\frac{1}{n}$ “.
- (iii) Prostým výpočtem $E\left(\sqrt{n} \frac{\bar{X}_n - \mu}{\sigma}\right) = 0$, a $\text{var}\left(\sqrt{n} \frac{\bar{X}_n - \mu}{\sigma}\right) = 1$.
- (iv) Ze cvičení, že nejen $(\bar{X}_n - \mu) \xrightarrow[n \rightarrow \infty]{sj} 0$, ale i $n^q(\bar{X}_n - \mu) \xrightarrow[n \rightarrow \infty]{P} 0$, pro $q \in (0, \frac{1}{2})$.

Takže $(\bar{X}_n - \mu)$ konverguje k 0 skoro jistě. Ale je možné $(\bar{X}_n - \mu)$ znormovat tak, aby nešlo ani k 0, ani neexplodovalo? Neboli je možné najít přesný řád konvergence $(\bar{X}_n - \mu)$?

Ze shrnutí výše je vidět, že \sqrt{n} je první řád, který může dát nekonstantní (tj. nenulovou) limitu. Neboli $\sqrt{n} \left(\frac{\bar{X}_n - \mu}{\sigma}\right)$ má šanci konvergovat pro $n \rightarrow \infty$ k nějaké náhodné veličině. Že skutečně konverguje, je tvrzením centrální limitní věty.

*Zde končí
předn. 16
(16.4.)*

4.5 CENTRÁLNÍ LIMITNÍ VĚTA

Centrální limitní věta (CLV) mluví o tom, že správně znormované odchylky výběrového průměru od střední hodnoty $\sqrt{n} \left(\frac{\bar{X}_n - \mu}{\sigma} \right)$ v posloupnosti nezávislých stejně rozdělených náhodných veličin konvergují k normovanému normálnímu rozdělení $N(0, 1)$. A to nezávisle na rozdělení X_i . Ukazuje tedy univerzálnost normálního rozdělení, a to, že normální rozdělení je dobrý model pro náhodné situace, ve kterých náhoda vznikla průměrováním velkého množství malých náhodných odchylek.

Abychom mohli formulovat centrální limitní větu, potřebujeme nový pojem konvergence náhodných veličin.

Definice 4.3 Buď $\{Y_n\}_{n=1}^\infty$ posloupnost náhodných veličin na libovolných pravděpodobnostních prostorech. Řekneme, že náhodné veličiny Y_n *konvergují v distribuci* k náhodné veličině Y , pokud $\lim_{n \rightarrow \infty} F_{Y_n}(x) = F_Y(x)$ v každém $x \in \mathbb{R}$, ve kterém je F_Y spojitá. Značíme $Y_n \xrightarrow[n \rightarrow \infty]{d} Y$.

Konvergence v distribuci je konvergence měr P_{Y_n} , nikoli funkcí Y_n . Proto mohou být Y_n definovány na různých Ω . Definice požaduje, aby míry P_{Y_n} množin $(-\infty, x]$ konvergovaly k míře množiny $(-\infty, x]$ měřené pomocí limitního rozdělení P_Y . Ovšem jen pro ta x , ve kterých je F_Y spojitá!

Pokud je rozdělení P_Y absolutně spojitě, pak konvergence v distribuci požaduje, aby $\lim_{n \rightarrow \infty} F_{Y_n}(x) = F_Y(x)$ pro všechna x . Dokonce budou F_{Y_n} konvergovat k F_Y stejnoměrně (neboť limitní funkce F_Y je spojitá a omezená a všechny F_{Y_n} jsou monotónní).

Pokud má limitní rozdělení P_Y nějaké atomy (tj. body, pro které $P_Y(\{x\}) > 0$), tak v těchto bodech nebude platit $\lim_{n \rightarrow \infty} F_{Y_n}(x) = F_Y(x)$.

Příklad. Buď $\{Y_n\}_{n=1}^\infty$ posloupnost náhodných veličin takových, že $Y_n \sim N(0, \frac{1}{n})$, $n \in \mathbb{N}$. Všechny Y_n mají absolutně spojitě rozdělení. A platí $Y_n \xrightarrow[n \rightarrow \infty]{d} Y$, pro Y s $P_Y = \delta_0$. Tedy $P(Y = 0) = 1$, Y má diskrétní rozdělení. Ověřte.

Příklad. Buď $\{a_n\}_{n=1}^\infty$ posloupnost čísel konvergujících k a seshora a $\{Y_n\}_{n=1}^\infty$ posloupnost náhodných veličin takových, že $P_{Y_n} = \delta_{a_n}$. Pak $Y_n \xrightarrow[n \rightarrow \infty]{d} Y$, kde $P_Y = \delta_a$, a v bodě a nebude platit $\lim_{n \rightarrow \infty} F_{Y_n}(a) = F_Y(a)$. Ověřte.

Poznámka. A pro jaké množiny $B \in \mathcal{B}$ konverguje $P_{Y_n}(B)$ k $P_Y(B)$, když $Y_n \xrightarrow[n \rightarrow \infty]{d} Y$? Pokud není Y absolutně spojitá náhodná veličina, tak ne nutně pro všechny. Lze ukázat, že $P_{Y_n}(B) \xrightarrow[n \rightarrow \infty]{} P_Y(B)$ pro všechny B s $P_Y(\partial B) = 0$, tedy s hranicí nulové míry P_Y (nebudeme dokazovat).

Poznámka. Konvergence $Y_n \xrightarrow[n \rightarrow \infty]{d} Y$ odpovídá takzvané slabé konvergenci měr $P_{Y_n} \xrightarrow[n \rightarrow \infty]{w} P_Y$, se kterou se setkáte ve funkcionální analýze.

Značení. Budeme psát například i $Y_n \xrightarrow[n \rightarrow \infty]{d} N(0, 1)$, a budeme tím mínit $Y_n \xrightarrow[n \rightarrow \infty]{d} Y$ a $Y \sim N(0, 1)$, to jest že Y_n konvergují v distribuci k náhodné veličině s rozdělením $N(0, 1)$.

K důkazu centrální limitní věty budeme ještě potřebovat ekvivalentní charakterizaci konvergence v distribuci.

Lemma 4.12 (charakterizace konvergence v distribuci) Buď $\{Y_n\}_{n=1}^{\infty}$ posloupnost náhodných veličin. Pak $Y_n \xrightarrow[n \rightarrow \infty]{d} Y$ právě tehdy, když $E h(Y_n) \xrightarrow[n \rightarrow \infty]{} E h(Y)$ pro každou spojitou a omezenou funkci $h : \mathbb{R} \rightarrow \mathbb{R}$.

Důkaz: Je technický, vynecháme z prostorových důvodů. Lze najít v [Georgii \(2013\)](#), kapitola 5.3. nebo [Lachout \(2004\)](#). \square

Poznámka. Na ověření $Y_n \xrightarrow[n \rightarrow \infty]{d} Y$ stačí ověřit $E h(Y_n) \xrightarrow[n \rightarrow \infty]{} E h(Y)$ i pro menší třídu funkcí než jsou spojitě a omezené. Stačí například h dvakrát spojitě diferencovatelné s omezenou a stejnoměrně spojitou první a druhou derivací (bez důkazu, viz např. [Georgii \(2013\)](#), kapitola 5.3).

Věta 4.13 (Centrální limitní věta) Buďte X_1, X_2, \dots vzájemně nezávislé, stejně rozdělené náhodné veličiny, $E X_i = \mu \in \mathbb{R}$, $\text{var } X_i = \sigma^2 \in (0, \infty)$, $\forall i \in \mathbb{N}$. Potom pro náhodné veličiny

$$Z_n = \frac{\sum_{i=1}^n (X_i - \mu)}{\sqrt{n\sigma^2}} \quad (4.10)$$

platí $Z_n \xrightarrow[n \rightarrow \infty]{d} Z$, kde $Z \sim N(0, 1)$.

Poznámky.

- Proč je v CLV v limitě právě $N(0, 1)$? Je to kvůli stabilitě normálního rozdělení vzhledem ke sčítání. Pokud jsou $X_i \sim N(0, 1)$ nezávislé, pak $Z_n = \frac{\sum_{i=1}^n X_i}{\sqrt{n}} \sim N(0, 1)$.
- Normální rozdělení je jediné rozdělení náhodné veličiny s konečným rozptylem, které má takovou vlastnost. Neboť, kdyby posloupnost náhodných veličin $\{X_i\}_{i=1}^{\infty}$ splňovala předpoklady CLV, každé X_i mělo rozdělení P_X a rozdělení P_X mělo tu vlastnost, že $Z_n = \frac{\sum_{i=1}^n X_i}{\sqrt{n}}$ má opět rozdělení P_X , tak nutně $Z_n \xrightarrow[n \rightarrow \infty]{d} Z$, která má rozdělení P_X a zároveň z CLV $Z \sim N(0, 1)$. Tedy P_X je rozdělení $N(0, 1)$.
- Předpoklad $X_i \in L^2$ (předpoklad konečného rozptylu) je nutný předpoklad pro platnost CLV. Extrémní protipříklad je náhodná posloupnost $\{X_i\}_{i=1}^{\infty}$, kde $X_i \sim \text{Cauchy}$, tj. $f_{X_i}(x) = \frac{1}{\pi(1+x^2)}$, $x \in \mathbb{R}$. Potom totiž $Z_n = \frac{\sum_{i=1}^n X_i}{\sqrt{n}}$ má rozdělení s hustotou

$$f_{Z_n}(z) = \frac{\sqrt{n}}{\pi(n+z^2)}, \quad z \in \mathbb{R},$$

tj. rozdělení stejné jako $\sqrt{n} X_1$, což zřejmě nekonverguje v distribuci k $N(0, 1)$.

- Předpoklad vzájemné nezávislosti v CLV není možno oslabit na předpoklad nekorelovanosti nebo nezávislosti po dvou.

Důkaz: (Centrální limitní věty)

Bez újmy na obecnosti můžeme předpokládat, že $E X_i = 0$ a $\text{var } X_i = 1$. Kdyby ne, aplikujeme následující na $\tilde{X}_i = \frac{X_i - \mu}{\sigma}$ — náhodné veličiny Z_n budou v obou případech stejné.

Z lemmatu 4.12 víme, že stačí ukázat $E h(Z_n) \xrightarrow{n \rightarrow \infty} E h(Z)$ pro každou omezenou spojitou funkci h . Podle poznámky za lemmatem to navíc stačí ukázat jen pro h dvakrát spojitě diferencovatelné s omezenou a stejnoměrně spojitou první a druhou derivací. Takže mějme takovou h .

A mějme navíc posloupnost vzájemně nezávislých náhodných veličin $\{Y_i\}_{i=1}^\infty$, které mají všechny rozdělení $N(0, 1)$. Pak $E Y_i = 0$, $\text{var } Y_i = 1$, a $T_n = \frac{\sum_{i=1}^n Y_i}{\sqrt{n}} \sim N(0, 1)$ (viz sekce 3.5.2). Takže chceme dokázat

$$|E(h(Z_n) - h(T_n))| \xrightarrow{n \rightarrow \infty} 0. \quad (4.11)$$

K tomu použijeme vyjádření $h(Z_n) - h(T_n)$ jako tzv. teleskopickou sumu. Nejdříve si zavedme označení

$$X_{i,n} = \frac{X_i}{\sqrt{n}}, \quad Y_{i,n} = \frac{Y_i}{\sqrt{n}}, \quad W_{i,n} = \sum_{j=1}^{i-1} Y_{j,n} + \sum_{j=i+1}^n X_{j,n}.$$

A nyní můžeme přepsat $h(Z_n) - h(T_n)$ jako součet malých rozdílů

$$h(Z_n) - h(T_n) = \sum_{i=1}^n (h(W_{i,n} + X_{i,n}) - h(W_{i,n} + Y_{i,n})), \quad (4.12)$$

neboť

$$W_{i,n} + X_{i,n} = W_{i-1,n} + Y_{i-1,n}, \quad 1 < i \leq n.$$

$X_{i,n}$ a $Y_{i,n}$ jsou malé a h je hladká. Použijeme Taylorův rozvoj k aproximaci h v bodě $W_{i,n} + X_{i,n}$ pomocí h v bodě $W_{i,n}$. Platí

$$h(W_{i,n} + X_{i,n}) = h(W_{i,n}) + h'(W_{i,n}) X_{i,n} + \frac{1}{2} h''(W_{i,n}) X_{i,n}^2 + R_{X_{i,n}}, \quad (4.13)$$

kde

$$R_{X_{i,n}}(\omega) = \frac{1}{2} X_{i,n}^2(\omega) [h''(W_{i,n}(\omega) + \zeta(\omega) X_{i,n}(\omega)) - h''(W_{i,n}(\omega))],$$

pro nějaké $\zeta(\omega) \in [0, 1]$, $\omega \in \Omega$.

Tedy jednak (pro větší názornost znovu použijeme zápis jako pro funkce $\omega \in \Omega$)

$$|R_{X_{i,n}}(\omega)| \leq X_{i,n}^2(\omega) \|h''\|_\infty, \quad \omega \in \Omega,$$

kde $\|\cdot\|_\infty$ je supremová norma funkce h'' . A druhak je h'' stejnoměrně spojitá, takže

$$\forall \epsilon > 0 \exists \delta > 0 : |R_{X_{i,n}}(\omega)| \leq X_{i,n}^2(\omega) \epsilon, \quad \text{pro } |X_{i,n}(\omega)| \leq \delta.$$

Dohromady tedy můžeme odhadnout shora

$$|R_{X_{i,n}}| \leq X_{i,n}^2 \left(\epsilon \mathbb{1}_{(|X_{i,n}| \leq \delta)} + \|h''\|_\infty \mathbb{1}_{(|X_{i,n}| > \delta)} \right). \quad (4.14)$$

Obdobně postupujeme pro $h(W_{i,n} + Y_{i,n})$. Dosadíme (4.13) i vyjádření pro $h(W_{i,n} + Y_{i,n})$ do (4.12) a spočteme střední hodnotu (o té chceme ukázat, že půjde k 0). Máme

$$\begin{aligned} E(h(Z_n) - h(T_n)) &= \sum_{i=1}^n \left[E(h(W_{i,n}) - h(W_{i,n})) + E(h'(W_{i,n})(X_{i,n} - Y_{i,n})) \right. \\ &\quad \left. + E\left(\frac{1}{2} h''(W_{i,n})(X_{i,n}^2 - Y_{i,n}^2)\right) + E R_{X_{i,n}} - E R_{Y_{i,n}} \right] = \sum_{i=1}^n [E R_{X_{i,n}} - E R_{Y_{i,n}}], \end{aligned}$$

neboť první tři střední hodnoty jsou nulové. Podrobně to ukážeme např. pro třetí z nich:

$$\begin{aligned} E\left(\frac{1}{2} h''(W_{i,n})(X_{i,n}^2 - Y_{i,n}^2)\right) &= \frac{1}{2} (E h''(W_{i,n})) (E(X_{i,n}^2 - Y_{i,n}^2)) \\ &= \frac{1}{2} (E h''(W_{i,n})) (E X_{i,n}^2 - E Y_{i,n}^2) = \frac{1}{2} (E h''(W_{i,n})) \cdot 0 = 0, \end{aligned}$$

neboť $W_{i,n}$ závisí jen na X_j, Y_j pro $j \neq i$ a tedy je nezávislá s $(X_i^2 - Y_i^2)$ a mohu použít Větu 3.11 pro druhou rovnost. Třetí rovnost plyne z rovnosti momentů X_i a Y_i .

Nyní použijeme (4.14), takže

$$\begin{aligned} |E(h(Z_n) - h(T_n))| &= \left| \sum_{i=1}^n (E R_{X_{i,n}} - E R_{Y_{i,n}}) \right| \leq \sum_{i=1}^n E(|R_{X_{i,n}}| + |R_{Y_{i,n}}|) \\ &\leq \sum_{i=1}^n \left[\epsilon E(X_{i,n}^2 - Y_{i,n}^2) + \|h''\|_\infty E(X_{i,n}^2 \mathbb{1}_{(|X_{i,n}| > \delta)} + Y_{i,n}^2 \mathbb{1}_{(|Y_{i,n}| > \delta)}) \right] \\ &= \left[\epsilon \sum_{i=1}^n \left(E \frac{X_1^2}{n} + E \frac{Y_1^2}{n} \right) \right] + \|h''\|_\infty \left[\sum_{i=1}^n E \left(\frac{X_1^2}{n} \mathbb{1}_{(|X_1| > \delta\sqrt{n})} + \frac{Y_1^2}{n} \mathbb{1}_{(|Y_1| > \delta\sqrt{n})} \right) \right] \\ &= 2\epsilon + \|h''\|_\infty E(X_1^2 \mathbb{1}_{(|X_1| > \delta\sqrt{n})} + Y_1^2 \mathbb{1}_{(|Y_1| > \delta\sqrt{n})}). \end{aligned}$$

Pro $E(X_1^2 \mathbb{1}_{(|X_1| > \delta\sqrt{n})})$ platí

$$E(X_1^2 \mathbb{1}_{(|X_1| > \delta\sqrt{n})}) = E X_1^2 - E(X_1^2 \mathbb{1}_{(|X_1| \leq \delta\sqrt{n})}) \xrightarrow{n \rightarrow \infty} E X_1^2 - E X_1^2 = 1 - 1 = 0,$$

z Lebesgueovy věty, neboť funkce $X_1^2 \mathbb{1}_{(|X_1| > \delta\sqrt{n})}$ konvergují monotónně k X_1^2 a $|X_1^2|$ je integrovatelná majoranta (předpoklad CLV je $X_1 \in L^2$). Obdobně postupujeme pro Y_1^2 a dohromady tedy dostaneme

$$\limsup_{n \rightarrow \infty} |E(h(Z_n) - h(T_n))| \leq 2\epsilon, \quad \forall \epsilon > 0,$$

což implikuje (4.11) a tím je věta dokázána. \square

Otázka. V předpokladech CLV je $\sigma^2 \in (0, \infty)$, tedy rozptyl nenulový. Co by se stalo, kdyby bylo $\sigma^2 = 0$?

Opět se můžeme zeptat, co nám říká CLV o posloupnosti Bernoulliiovských pokusů. V tomto případě si odpověď neformulujeme jako příklad, ale jako speciální centrální limitní větu.

Věta 4.14 (de Moivre–Laplaceova centrální limitní věta) Bud' $Y_n \sim \text{Binom}(n, p)$, $0 < p < 1$. Pak

$$\frac{Y_n - np}{\sqrt{np(1-p)}} \xrightarrow[n \rightarrow \infty]{d} N(0, 1).$$

Důkaz: Bud' $\{X_n\}_{n=1}^\infty$ Bernoulliiovská posloupnost s pravděpodobností úspěchu $p \in (0, 1)$. Neboli X_n jsou vzájemně nezávislé, alternativně rozdělené, $E X_n = p$, $\text{var } X_n = p(1-p) > 0$. Pak $\{X_n\}_{n=1}^\infty$ splňuje předpoklady centrální limitní věty a proto

$$Z_n = \frac{\sum_{i=1}^n (X_i - p)}{\sqrt{np(1-p)}} \xrightarrow[n \rightarrow \infty]{d} N(0, 1).$$

Ovšem $Y_n = \sum_{i=1}^n X_i \sim \text{Binom}(n, p)$, a tedy i $\frac{Y_n - np}{\sqrt{np(1-p)}} \xrightarrow[n \rightarrow \infty]{d} N(0, 1)$. \square

Poznámka. CLV pro binomické rozdělení, resp. pro Bernoulliiovskou posloupnost byla dokázána o mnoho dříve (A. de Moivre 1733, P.S. Laplace 1812) než obecná CLV (A.M. Ljapunov 1901, J.W. Lindenberga 1922, P. Lévy 1922). Je totiž možno upočítat limity pro distribuční funkci binomického rozdělení přímo – podrobnosti viz proseminář.

CLV se používá nejen jako teoretický limitní výsledek, ale i jako nástroj pro aproximaci rozdělení součtu většího (ovšem konečného) množství n nezávislých stejně rozdělených náhodných veličin $\sum_{i=1}^n X_i$. Přesné rozdělení takového součtu by šlo teoreticky odvodit pomocí konvoluce (věta 3.16), ale pro velké n je to výpočetně náročné a aproximace pomocí CLV je jednoduchá a ve většině případů dostačující (viz též cvičení). Kvalita aproximace samozřejmě závisí na velikosti n a na rozdílu ve tvaru rozdělení P_{X_1} a $N(0, 1)$, respektive na rozdílu tvaru jejich distribučních funkcí.

Poznámka. Pokud je $X_1 \in L^3$ (neboli má konečný třetí centrální moment), pak lze dokázat tzv. Berryho-Esseenovu nerovnost, která tvrdí (ve značení CLV 4.13)

$$\|F_{Z_n} - \Phi\|_\infty \leq 0.8 \frac{E(|X_1 - E(X_1)|^3)}{\sigma^3} \frac{1}{\sqrt{n}}.$$

Odhaduje tedy shora rychlost konvergence rozdílu distribučních funkcí zúčastněných v CLV a ukazuje, že řádově odpovídá $\frac{1}{\sqrt{n}}$ a závisí na znormovaném třetím momentu $\frac{E(|X_1 - E(X_1)|^3)}{\sigma^3}$, ve kterém se právě projevuje tvar rozdělení P_{X_1} .

Příklad. Uvažujme opakované hody spravedlivou mincí. Jaká je pravděpodobnost, že průměrná četnost hlav ve 100 pokusech se bude od $p = \frac{1}{2}$ lišit o více než 0.1?

Situaci můžeme modelovat posloupností $\{X_i\}_{i=1}^{100}$ nezávislých náhodných veličin s alternativním rozdělením s pravděpodobností úspěchu $p = \frac{1}{2}$ (mince je spravedlivá). Chceme tedy určit $P(|\frac{1}{100} \sum_{i=1}^{100} X_i - \frac{1}{2}| > \frac{1}{10})$. Ovšem $Y_n = \sum_{i=1}^n X_i$ má binomické rozdělení s parametry $p = \frac{1}{2}$ a n , tedy z CVL pro binomické rozdělení 4.14 platí

$$Z_n = \frac{Y_n - \frac{n}{2}}{\sqrt{\frac{n}{4}}} \xrightarrow[n \rightarrow \infty]{d} N(0, 1),$$

Neboli $\lim_{n \rightarrow \infty} F_{Z_n}(x) = \Phi(x)$. Přepíšme

$$P\left(\left|\frac{1}{100} \sum_{i=1}^{100} X_i - \frac{1}{2}\right| > \frac{1}{10}\right) = P\left(\frac{1}{\sqrt{4 \cdot 100}} \left|Y_{100} - \frac{100}{2}\right| > \frac{1}{10}\right) = P\left(\left|\frac{Y_{100} - \frac{100}{2}}{\sqrt{\frac{100}{4}}}\right| > 2\right).$$

Pokud nyní nahradíme distribuční funkci $\frac{Y_{100} - \frac{100}{2}}{\sqrt{\frac{100}{4}}}$ distribuční funkcí Φ normovaného normálního rozdělení, dostaneme aproximaci

$$P\left(\left|\frac{1}{100} \sum_{i=1}^{100} X_i - \frac{1}{2}\right| > \frac{1}{10}\right) \doteq 1 - \Phi(2) + \Phi(-2) = 2\Phi(-2) \doteq 0.046. \quad (4.15)$$

V druhé rovnosti jsme využili symetrii $\Phi(-x) = 1 - \Phi(x)$, $x \in \mathbb{R}$. Kdybychom si položili stejnou otázku pro případ $n = 1000$ pokusů, dostali bychom

$$P\left(\left|\frac{1}{1000} \sum_{i=1}^{1000} X_i - \frac{1}{2}\right| > \frac{1}{10}\right) \doteq 1 - \Phi(2\sqrt{10}) + \Phi(-2\sqrt{10}) = 2\Phi(-2\sqrt{10}) \doteq 2.5 \cdot 10^{-10},$$

tedy průměr z 1000 pokusů bezpečně odhaduje pravděpodobnost úspěchu s přesností na 1 desetinné místo. Pokud bychom se ptali na přesnost dvou desetinných míst, pak můžeme spočítat

$$P\left(\left|\frac{1}{1000} \sum_{i=1}^{1000} X_i - \frac{1}{2}\right| > \frac{1}{100}\right) \doteq 1 - \Phi\left(\sqrt{\frac{2}{5}}\right) + \Phi\left(-\sqrt{\frac{2}{5}}\right) = 2\Phi\left(-\sqrt{\frac{2}{5}}\right) \doteq 0.53,$$

tedy na tuto přesnost bychom potřebovali mnohem více pokusů než 1000. Pro $n = 10000$ pokusů bychom opět dostali $P(|\bar{X}_{10000} - \frac{1}{2}| > \frac{1}{100}) \doteq 0.046$, jako v (4.15). Neboli tento příklad také ilustruje, že odchylka empirického průměru od p , resp. $E X_1$ se opravdu zmenšuje s řádem $\frac{1}{\sqrt{n}}$.

Samozřejmě lze odpovídat i na otázky „v jakém rozmezí se nachází rozdíl mezi \bar{X}_n a p po $n = 100$ pokusech s pravděpodobností 0.99“ nebo „jak velké musí být n , aby odchylka mezi \bar{X}_n a p byla menší než 0.05 alespoň s pravděpodobností 0.95“. Jak je řešit – viz cvičení.

Poznámka. Mohli bychom se ovšem také ptát, jestli pro dané n je už rozumné použít aproximaci pomocí CLV, jako v předchozím příkladě. Tzv. pravidlo palce pro de Moivre-Laplaceovu CLV říká, že aproximace pro binomické rozdělení je prakticky použitelná pokud $np(1-p) \geq 9$ (viz např. [Dupač and Hušková \(2013\)](#), kapitola 4.3).

Pro obecnou CLV se udává jako mez použitelnosti $n \geq 30$, viz např. [Hogg \(2015\)](#), kapitola 5.6.

Ještě se na chvíli vraťme ke konvergenci v distribuci – zachovává se také při spojitě transformaci?

Lemma 4.15 (O spojitě transformaci a konvergenci v distribuci) Bud' $\{X_n\}_{n=1}^{\infty}$ posloupnost náhodných veličin taková, že $X_n \xrightarrow[n \rightarrow \infty]{d} X$, kde X je nějaká náhodná veličina. A bud' $g : \mathbb{R} \rightarrow \mathbb{R}$ spojitá funkce. Pak pro náhodné veličiny $Y_n = g(X_n)$, $n \in \mathbb{N}$, platí $Y_n \xrightarrow[n \rightarrow \infty]{d} Y$, kde $Y = g(X)$.

Důkaz: Použijeme charakterizaci konvergence v distribuci (Lemma 4.12). Tedy chceme dokázat, že $E h(Y_n) \xrightarrow[n \rightarrow \infty]{} E h(Y)$ pro každou spojitou, omezenou funkci $h : \mathbb{R} \rightarrow \mathbb{R}$. Ale $E h(Y_n) = E h(g(X_n))$ a složená funkce $h \circ g$ je zřejmě spojitá a omezená. Tedy $E h(g(X_n)) \xrightarrow[n \rightarrow \infty]{} E h(g(X)) = E h(Y)$ z předpokladu $X_n \xrightarrow[n \rightarrow \infty]{d} X$. \square

Poznámka. Lemma využijeme ve statistice pro konzistentní odhady.

Příklad. Bud' $\{X_n\}_{n=1}^{\infty}$ posloupnost nezávislých stejně rozdělených veličin (třeba s exponenciálním rozdělením, nebo Poissonovo, nebo rovnoměrným na $(-6, 13)$, nebo ...). Bud' $E X_1 = \mu$, $\text{var } X_1 = \sigma^2 \in (0, \infty)$, a necht' $E X_1^4 < \infty$. Potom z CLV platí

$$\frac{\sum_{i=1}^n (X_i - \mu)}{\sqrt{n} \sigma} \xrightarrow[n \rightarrow \infty]{d} N(0, 1). \quad (4.16)$$

A použitím CLV na $(X_i - \mu)^2$ dostaneme

$$\frac{\sum_{i=1}^n ((X_i - \mu)^2 - \sigma^2)}{\sqrt{n} \sqrt{\text{var}((X_1 - \mu)^2)}} = \sqrt{n} \frac{\overline{(X_n - \mu)^2} - \sigma^2}{\sqrt{\text{var}((X_1 - \mu)^2)}} \xrightarrow[n \rightarrow \infty]{d} N(0, 1).$$

A co víme o $\overline{(X_n - \mu)^2}$? Aplikací lemmatu o spojitě transformaci pro $\xrightarrow[n \rightarrow \infty]{d}$ na posloupnost z (4.16) a funkci $g(x) = x^2$ dostaneme

$$\left(\frac{\sum_{i=1}^n (X_i - \mu)}{\sqrt{n} \sigma} \right)^2 = \frac{\left(\sum_{i=1}^n (X_i - \mu) \right)^2}{n \sigma^2} \xrightarrow[n \rightarrow \infty]{d} V,$$

kde $V \sim \chi_1^2$, neboli V má stejné rozdělení jako Z^2 , kde Z má $N(0, 1)$ rozdělení. Toto bychom přímo pro $\frac{(\sum_{i=1}^n (X_i - \mu))^2}{n \sigma^2}$ nedokázali odvodit, po roznásobení bychom totiž obdrželi členy, které by už nebyly vzájemně nezávislé a nemohli bychom použít CLV.

A ještě je potřeba se zmínit o souvislosti konvergence v distribuci s předchozími konvergencemi.

Věta 4.16 Bud' $\{X_n\}_{n=1}^{\infty}$ posloupnost náhodných veličin definovaných na stejném (Ω, \mathcal{A}, P) . Pak $|X_n - X| \xrightarrow[n \rightarrow \infty]{P} 0 \Rightarrow X_n \xrightarrow[n \rightarrow \infty]{d} X$.

Důkaz: Neuvádíme, viz např. [Lachout \(2004\)](#). □

Důsledek. Speciálně tedy platí $|X_n| \xrightarrow[n \rightarrow \infty]{P} 0 \Rightarrow X_n \xrightarrow[n \rightarrow \infty]{d} 0$.

Poznámka. Obrácená implikace neplatí. Uvažujme $(\Omega, \mathcal{A}, P) = ([0, 1], \mathcal{B}([0, 1]), \lambda)$ a $X_n = \mathbb{1}_{[0, \frac{1}{2}]}$ pro n sudé, $X_n = \mathbb{1}_{[\frac{1}{2}, 1]}$ pro n liché. Platí $P(X_n = 0) = P(X_n = 1) = \frac{1}{2}$, $\forall n \in \mathbb{N}$, takže $X_n \xrightarrow[n \rightarrow \infty]{d} X$, kde $X \sim \text{Alt}(\frac{1}{2})$. Ale v pravděpodobnosti X_n nekonvergují.

Ve statistice se nám bude hodit ještě jedna věta, která velmi zvyšuje užitečnost CLV při konstrukci „pěkných“ statistických odhadů.

Věta 4.17 (Cramérova–Sluckého) Bud'te $\{X_n\}_{n=1}^{\infty}, \{Y_n\}_{n=1}^{\infty}, X$ náhodné veličiny a necht' $X_n \xrightarrow[n \rightarrow \infty]{d} X$ a $Y_n \xrightarrow[n \rightarrow \infty]{P} a \in \mathbb{R}$. Potom $(X_n + Y_n) \xrightarrow[n \rightarrow \infty]{d} (X + a)$ a $X_n \cdot Y_n \xrightarrow[n \rightarrow \infty]{d} a X$.

Důkaz: Větu dokážeme jen pro speciální případ $X \sim N(0, 1)$, kdy to budeme v dalším používat. Důkazy obou tvrzení jsou obdobné, ukážeme např. druhé tvrzení $X_n \cdot Y_n \xrightarrow[n \rightarrow \infty]{d} a X$.

Búno $a > 0$ a necht' $Y_n \xrightarrow[n \rightarrow \infty]{P} a \in \mathbb{R}$. Chceme zjistit kolik je $\lim_{n \rightarrow \infty} P(X_n \cdot Y_n \leq x)$, resp. ukázat, že je to rovno $P(Xa \leq x) = \Phi(\frac{x}{a})$. Bud' $x > 0$. Odhadneme množinu $\{\omega : X_n(\omega) \cdot Y_n(\omega) \leq x\}$ zespoda i seshora. Bud' $\epsilon > 0$, ale malé tak, aby $(a - \epsilon) > 0$. Pak platí:

$$\begin{aligned} & \{X_n(a - \epsilon) \leq x\} \cup \{|Y_n - a| > \epsilon\} \\ & \supseteq \{\omega : X_n(\omega) \cdot Y_n(\omega) \leq x\} = \{X_n \cdot Y_n \leq x, |Y_n - a| \leq \epsilon\} \cup \{X_n \cdot Y_n \leq x, |Y_n - a| > \epsilon\} \\ & \supseteq \{X_n \cdot Y_n \leq x, |Y_n - a| \leq \epsilon\} \supseteq \{X_n(a + \epsilon) \leq x, |Y_n - a| \leq \epsilon\} \\ & \supseteq \{X_n(a + \epsilon) \leq x\} \setminus \{|Y_n - a| > \epsilon\}. \end{aligned}$$

Z toho

$$P\left(X_n \leq \frac{x}{a + \epsilon}\right) - P(|Y_n - a| > \epsilon) \leq P(X_n \cdot Y_n \leq x) \leq P\left(X_n \leq \frac{x}{a - \epsilon}\right) + P(|Y_n - a| > \epsilon).$$

Spočítejme nyní limity pro $n \rightarrow \infty$. Dostaneme:

$$\Phi\left(\frac{x}{a + \epsilon}\right) - 0 \leq \liminf_{n \rightarrow \infty} P(X_n \cdot Y_n \leq x) \leq \limsup_{n \rightarrow \infty} P(X_n \cdot Y_n \leq x) \leq \Phi\left(\frac{x}{a - \epsilon}\right) + 0,$$

z předpokladů věty. Pokud teď spočítáme limitu pro $\epsilon \rightarrow 0$, dostaneme ze spojitosti Φ (je spojitá všude)

$$\lim_{n \rightarrow \infty} P(X_n \cdot Y_n \leq x) = \Phi\left(\frac{x}{a}\right) = P(aX \leq x). \quad (4.17)$$

Pro $x < 0$ postupujeme obdobně a dostaneme se k inkluzím

$$\{X_n(a + \epsilon) \leq x\} \cup \{|Y_n - a| > \epsilon\} \supseteq \{\omega : X_n(\omega) \cdot Y_n(\omega) \leq x\} \supseteq \{X_n(a - \epsilon) \leq x\} \setminus \{|Y_n - a| > \epsilon\},$$

a tedy nerovností

$$P\left(X_n \leq \frac{x}{a - \epsilon}\right) - P(|Y_n - a| > \epsilon) \leq P(X_n \cdot Y_n \leq x) \leq P\left(X_n \leq \frac{x}{a + \epsilon}\right) + P(|Y_n - a| > \epsilon).$$

Limitními předchody obdržíme opět (4.17). Pro $x = 0$ plyne konvergence z již dokázaného a monotonie distribuční funkce. \square

Poznámka. Pro F_X s body nespojitosti je důkaz složitější. Viz např. [Lachout \(2004\)](#).

A jaké je použití Cramérovo–Sluckého věty? Pro posloupnost nezávislých stejně rozdělených náhodných veličin $\{X_i\}_{i=1}^{\infty}$ máme z CLV

$$Z_n = \frac{\sum_{i=1}^n X_i - n\mu}{\sqrt{n\sigma^2}} \xrightarrow[n \rightarrow \infty]{d} N(0, 1).$$

Představme si, že neznámou střední hodnotu μ chci odhadnout právě z pozorování posloupnosti $\{X_i\}_{i=1}^{\infty}$. Pak mám problém, že na levé straně konvergence se vyskytuje také hodnota σ^2 , kterou rovněž neznám. Pokud ale najdu nějakou posloupnost náhodných veličin $\{Y_i\}_{i=1}^{\infty}$ takovou, že $Y_n \xrightarrow[n \rightarrow \infty]{P} \sigma^2$, pak z lemmatu o spojitě transformaci

pro konvergenci v pravděpodobnosti dostanu $\frac{Y_n}{\sigma^2} \xrightarrow[n \rightarrow \infty]{P} 1$ a také $U_n = \sqrt{\frac{\sigma^2}{Y_n}} \xrightarrow[n \rightarrow \infty]{P} 1$.

Takže $Z_n \xrightarrow[n \rightarrow \infty]{d} N(0, 1)$ a $U_n \xrightarrow[n \rightarrow \infty]{d} 1$, a z Cramérovo–Sluckého věty dostanu

$$Z_n \cdot U_n = \frac{\sum_{i=1}^n X_i - n\mu}{\sqrt{n} \sqrt{Y_n}} \xrightarrow[n \rightarrow \infty]{d} N(0, 1).$$

Nyní už je na levé straně neznámé jen to, co chci statisticky odhadnout.

*Zde končí
předn. 18
(23.4.)*

5 STATISTIKA

Statistika řeší, jak racionálně odpovídat na otázky, které nás zajímají, o reálné situaci, která zahrnuje náhodu. Například můžeme mít nový lék, a nás zajímá, jestli účinkuje. Nebo jestli účinkuje lépe než nějaký starší lék.

Aby bylo možno na takové otázky odpovědět, postupuje se ve statistice takto: pro náhodnou situaci vytvoříme pravděpodobnostní model („rozumný“, jehož předpoklady odpovídají charakteristickým rysům dané situace, anebo prostě jen dostatečně výpočetně zvládnutelný a ne zjevně ve sporu se situací). Na základě pozorování (empirických dat) pak můžeme odhadovat neznámé parametry toho modelu, testovat hypotézy o nich a nakonec i posoudit shodu modelu se skutečností (nebo přesněji, ověřit, že model není v rozporu se skutečností).

Definice 5.1 Necht' X_1, \dots, X_n jsou nezávislé a stejně rozdělené náhodné veličiny (nebo vektory) s rozdělením P_X , resp. s distribuční funkcí F . Pak X_1, \dots, X_n nazveme *náhodný výběr z rozdělení P_X , resp. náhodný výběr z rozdělení s distribuční funkcí F* . Číslo n se nazývá *rozsah výběru*.

U náhodného výběru P_X ani F neznáme. Chceme použít pozorování/data X_1, \dots, X_n k tomu, abychom se o P_X , resp. F něco dozvěděli. O P_X předpokládáme, že patří do nějaké množiny rozdělení, které říkáme model.

Definice 5.2 *Modelem pro pozorování X_1, \dots, X_n rozumíme předem stanovenou množinu rozdělení \mathcal{F} , do níž neznámé rozdělení P_X , resp. jeho distribuční funkce F patří.*

Model je tedy tvořen všemi potenciálními rozděleními P_X , z nichž by pozorování X_1, \dots, X_n mohla pocházet. Popis pomocí P_X resp. F je samozřejmě ekvivalentní. Model může být určen s různou přesností.

Pokud o P_X , resp. F , víme málo, nebo nechceme předpokládat moc, pak se používá tzv. „neparametrický“ přístup. \mathcal{F} může být např. třída všech rozdělení s konečnou střední hodnotou.

Příklad. Mějme náhodný výběr X_1, \dots, X_n , o němž nic nevíme, a chtěli bychom si udělat alespoň hrubou představu o rozdělení P_X , ze kterého pochází. Budeme tedy předpokládat velmi obecný model $\mathcal{F} = \{P_X \text{ taková, že } E X \in \mathbb{R}\}$. Pro pozorovaná data můžeme spočítat empirickou distribuční funkci \hat{F}_n (viz následující definice), a použít ji jako náš odhad neznámé distribuční funkce F . Díky Glivenko-Cantelliho větě (viz doplňky z týdne 8 – aplikace SZVČ) víme, že

$$\lim_{n \rightarrow \infty} \sup_{x \in \mathbb{R}} |\hat{F}_n(x) - F(x)| = 0, \quad \text{s.j.,}$$

tedy se zvětšujícím se rozsahem výběru bude náš odhad \hat{F}_n konvergovat k neznámé F skoro jistě. To je chování, které by rozumný odhad měl splňovat.

Definice 5.3 Buď X_1, \dots, X_n náhodný výběr. *Empirická distribuční funkce* je definována předpisem

$$\hat{F}_n(x, \omega) = \frac{1}{n} \sum_{k=1}^n \mathbb{1}_{(X_k \leq x)}(\omega) = \frac{1}{n} \sum_{k=1}^n \mathbb{1}_{(X_k(\omega) \leq x)}, \quad x \in \mathbb{R}, \omega \in \Omega.$$

Druhý možný přístup statistického modelování je parametrický. O neznámém rozdělení P_X , resp. distribuční funkci F , máme poměrně přesnou představu, neznámých je jen několik (málo) parametrů. Neboli neznámá $F = F_{\theta_0}$ patří do třídy distribučních funkcí $\mathcal{F} = \{F_\theta \in \Theta\}$, kde $\Theta \subseteq \mathbb{R}^d$, $d \in \mathbb{N}$, je borelovská podmnožina \mathbb{R}^d . Θ se nazývá *parametrický prostor* a θ_0 je neznámý (pro $d > 1$ vektorový) parametr. Tímto přístupem se budeme podrobněji zabývat v následujících kapitolách.

Příklad. Parametrický model může být např. $\mathcal{F} = \{N(\mu, \sigma^2) : \mu \in \mathbb{R}, \sigma^2 > 0\} = \{P_\theta : \theta \in \Theta\}$, s $\theta = (\mu, \sigma^2)$ a $\Theta = \mathbb{R} \times \mathbb{R}^+ -$ třída všech normálních rozdělení. Zde $d = 2$ a hledáme neznámé parametry: střední hodnotu μ a rozptyl σ^2 . Třeba když opakovaně měříme nějakou fyzikální veličinu s náhodnou chybou. Zajímá nás $\mu =$ skutečná hodnota fyzikální veličiny, kterou chceme změřit, a $\sigma^2 =$ přesnost přístroje, kterým ji měříme.

Nebo může být $\mathcal{F} = \{Alt(p) : p \in (0, 1)\} = \{P_\theta : \theta \in \Theta\}$ s $\theta = p$ a $\Theta = (0, 1) -$ třída všech alternativních rozdělení s pravděpodobností úspěchu p . Parametr je jedno-rozměrná pravděpodobnost úspěchu p . Například pozorujeme posloupnost Bernoulliových pokusů – výsledky opakovaného hodu (jednou, stejnou) mincí, a zajímá nás, jestli je mince spravedlivá (jestli je $p = \frac{1}{2}$).

Model \mathcal{F} i parametry θ volíme my. Model vyjadřuje naši apriorní (na datech nezávislou) představu o rozdělení pozorovaných veličin. Volba parametru závisí na otázce, kterou chceme zodpovědět.

*Zde končí
předn. 18
a půl
(24.4.)*

5.1 BODOVÝ ODHAD

Mějme tedy model – třídu parametrických rozdělení $\{P_\theta : \theta \in \Theta\}$, a X_1, \dots, X_n náhodný výběr z P_θ , kdy ovšem konkrétní hodnotu parametru θ neznáme, chceme ji určit na základě pozorování toho náhodného výběru. A nebo nechceme určit přímo θ , ale chceme určit nějakou jeho funkci $g(\theta)$ (například v normálním modelu nás nemusí zajímat celé (μ, σ^2) , ale chceme určit jen μ , nebo třeba $\mathbb{1}_{(\mu \in [1, 3])}$).

Definice 5.4 Borelovsky měřitelné zobrazení $g : \Theta \rightarrow \mathbb{R}$ nazveme *parametrickou funkcí*.

Na základě náhodného výběru X_1, \dots, X_n z P_θ tedy chceme odhadnout $g(\theta)$.

Definice 5.5 Bodový odhad φ_n parametrické funkce $g(\theta)$, je borelovsky měřitelné zobrazení $\varphi_n : \mathbb{R}^n \rightarrow \mathbb{R}$, jehož předpis nezávisí na θ (tedy ani na P_θ či F_θ), a jeho nosič je obor hodnot (X_1, \dots, X_n) .

Poznámka. X_1 je někdy náhodná veličina a někdy jedno pozorování. Jak to rozlišit? Pozorované $X_1(\omega)$ v nějakém konkrétním, avšak neznámém ω budeme značit malým písmenem, tj. $x_1 = X_1(\omega)$.

Příklad. Uvažujme posloupnost X_1, X_2, \dots výsledků hodu mincí. X_1, X_2, \dots lze tedy modelovat jako Bernoulliiovskou posloupnost. Pozorujeme počáteční úsek X_1, \dots, X_n a zajímá nás, jestli je mince spravedlivá. Pak X_1, \dots, X_n je náhodný výběr z alternativního rozdělení s neznámou pravděpodobností úspěchu p . Modelem tedy je $\mathcal{F} = \{Alt(p) : p \in [0, 1]\} = \{P_\theta : \theta \in \Theta\}$ s $\theta = p$ a $\Theta = [0, 1]$.

Chceme odhadnout přímo θ , tedy parametrická funkce $g(\theta) = \theta$ je identita. Na základě našich dosavadních znalostí z kurzu bychom jako bodový odhad $\theta = p$ navrhli $\varphi_n : \mathbb{R}^n \rightarrow \mathbb{R}$, definovanou jako

$$\varphi_n(x_1, \dots, x_n) = \frac{1}{n} \sum_{k=1}^n x_k = \bar{x}_n.$$

Uvědomme si, že pro $(x_1, \dots, x_n) \in \{0, 1\}^n$ je $\varphi_n(x_1, \dots, x_n) \in [0, 1]$, což chceme.

V příkladu je $\varphi_n : \mathbb{R}^n \rightarrow \mathbb{R}$ funkce, pro konkrétní pozorovaná x_1, \dots, x_n je tedy $\varphi_n(x_1, \dots, x_n) = \bar{x}_n$ číslo. Ovšem $\varphi_n(X_1, \dots, X_n) = \bar{X}_n$ je náhodná veličina. V češtině se vše nazývá odhad – ať je to funkce, náhodná veličina, nebo číslo. Ale je potřeba to rozlišovat! V anglické literatuře to rozlišené je – funkce φ_n a náhodná veličina $\varphi_n(X_1, \dots, X_n)$ se nazývá estimator, zatímco číslo $\varphi_n(x_1, \dots, x_n)$ se nazývá estimate.

Uvědomme si také, že funkce $\varphi_n(x_1, \dots, x_n) = \bar{x}_n$ (její předpis) nezávisí na hodnotě neznámého parametru $\theta = p$ (to nesmí z definice odhadu, a také, protože $\theta = p$ prostě neznáme). Ovšem rozdělení náhodné veličiny $\varphi_n(X_1, \dots, X_n) = \bar{X}_n$ na hodnotě parametru $\theta = p$ závisí. Konkrétně v příkladu výše má $\varphi_n(X_1, \dots, X_n)$ rozdělení jako $\frac{1}{n}Z$, kde $Z \sim Binom(n, p)$. Ideálně by rozdělení odhadu $\varphi_n(X_1, \dots, X_n)$ mělo na θ záviset velmi a pro různé hodnoty θ se podstatně odlišovat, protože právě tato odlišnost umožní neznámé θ (s určitou mírou nejistoty) identifikovat = odhadnout. To se ovšem dá zařídit jen někdy a do určité míry.

Značení. V dalším budeme někdy zkráceně značit odhad $\varphi_n(X_1, \dots, X_n)$ parametru θ na základě náhodného výběru X_1, \dots, X_n jen symbolem $\hat{\theta}_n$.

Zřejmě takových funkcí φ_n , které splňují definici bodového odhadu pro $g(\theta)$ může být mnoho. Ale které jsou ty “rozumné“, resp. „dobré“, nebo dokonce „nejlepší“?

Nejdříve si zadefinujeme některé žádoucí vlastnosti, které by „dobrý“ odhad měl splňovat.

Definice 5.6 Bodový odhad φ_n parametrické funkce $g(\theta)$ se nazývá *nestranný*, pokud pro každé $\theta \in \Theta$ platí

$$E_{\theta} \varphi_n(X_1, \dots, X_n) = g(\theta).$$

Použili jsme značení E_{θ} střední hodnoty počítané vzhledem k rozdělení $P_{\theta} \otimes \dots \otimes P_{\theta}$, tedy když je X_1, \dots, X_n náhodný výběr z rozdělení P_{θ} s distribuční funkcí F_{θ} .

Definice 5.7 Posloupnost bodových odhadů φ_n parametrické funkce $g(\theta)$ se nazývá *silně konzistentní*, pokud $\forall \theta \in \Theta$ platí

$$P(\lim_{n \rightarrow \infty} \varphi_n(X_1, \dots, X_n) = g(\theta)) = 1,$$

neboli když $\varphi_n(X_1, \dots, X_n)$ konverguje ke $g(\theta)$ skoro jistě.

Poznámka. Někdy se zkráceně říká, že „bodový odhad φ_n je konzistentní“, myslí se tím ovšem, že posloupnost bodových odhadů φ_n je konzistentní, tak jako v definici výše.

Poznámka. Posloupnost odhadů se nazývá *slabě konzistentní*, pokud $\varphi_n(X_1, \dots, X_n)$ konverguje ke $g(\theta)$ v pravděpodobnosti. Samozřejmě, ta pravděpodobnost, vzhledem ke které mají odhady konvergovat, je rozdělení celé posloupnosti nezávislých, stejně rozdělených $\{X_n\}_{n=1}^{\infty}$, které mají rozdělení P_{θ} .

Pokud řekneme jen „konzistentní“, budeme tím vždy myslet silně konzistentní.

Příklad. (pokračování) Odhad $\varphi_n(X_1, \dots, X_n) = \bar{X}_n$ je nestranný odhad pravděpodobnosti úspěchu p , neboť

$$E \bar{X}_n = \frac{1}{n} \sum_{k=1}^n E X_k = \frac{1}{n} \sum_{k=1}^n p = p.$$

A posloupnost odhadů φ_n je silně konzistentní, neboť $\bar{X}_n \xrightarrow[n \rightarrow \infty]{s_j} E X_1 = p$ ze SZVČ pro stejně rozdělené náhodné veličiny.

A co vlastně ty vlastnosti nestrannosti a konzistence znamenají? Nestrannost odpovídá tomu, že odhad $\hat{\theta}_n$ je zatížen jenom náhodnou, nikoli systematickou chybou – to jest není systematicky vychýlen. A konzistence, volně řečeno, znamená, že volbou dostatečně velkého rozsahu výběru n lze udělat chybu odhadu libovolně malou. Také ještě existuje vlastnost asymptotické nestrannosti, posloupnost odhadů φ_n ji splňuje, pokud platí

$$E_{\theta} \varphi_n(X_1, \dots, X_n) \xrightarrow[n \rightarrow \infty]{} g(\theta), \quad \forall \theta \in \Theta.$$

Neboli systematická chyba odhadu φ_n s $n \rightarrow \infty$ vymizí.

Konzistence je nutná vlastnost jakéhokoli rozumného odhadu, neboť bez ní ani při neomezeném počtu pozorování nejsme schopni určit $g(\theta)$. Neboli ani v principu neumíme určit $g(\theta)$ správně. Také asymptotickou nestrannost chceme. Nestrannost je

dobrá vlastnost, ale záleží i na jiných kritériích. Uvědomme si například, že nestrannost říká, že

$$E_{\theta}(\varphi_n(X_1, \dots, X_n) - g(\theta)) = 0, \quad \text{nikoli} \quad E_{\theta}|\varphi_n(X_1, \dots, X_n) - g(\theta)| = 0,$$

(otázka pro čtenáře - co by znamenalo, kdyby platila druhá rovnost?). Chybu mezi φ_n a $g(\theta)$ lze měřit např. i pomocí střední čtvercové chyby

$$E_{\theta}(\varphi_n(X_1, \dots, X_n) - g(\theta))^2. \quad (5.1)$$

A ptát se, jak rychle jde (5.1) s rostoucím n k 0. Pokud je odhad φ_n nestranný, pak (5.1) je rovno rozptylu $\text{var}(\varphi_n(X_1, \dots, X_n))$ odhadu φ_n , tedy (5.1) pro $n \rightarrow \infty$ popisuje, jak rychle s rostoucím n klesá variabilita (neboli nejistota) odhadu.

K diskusi vlastností bodových odhadů se ještě vrátíme. Teď začneme tím, že prozkoumáme 2 základní statistiky.

Definice 5.8 *Statistikou* nazveme libovolnou borelovsky měřitelnou funkci $T_n : \mathbb{R}^n \rightarrow \mathbb{R}$, jejíž nosič je obor hodnot náhodného výběru (X_1, \dots, X_n) .

Takže statistika $T_n(X_1, \dots, X_n)$ je náhodná veličina.

Definice 5.9 Statistika

$$\bar{X}_n = \frac{1}{n} \sum_{k=1}^n X_k$$

se nazývá *výběrový průměr*. A statistika

$$S_n^2 = \frac{1}{n-1} \sum_{k=1}^n (X_k - \bar{X}_n)^2$$

se nazývá *výběrový rozptyl*.

Věta 5.1 (o výběrovém průměru a rozptylu) Buď X_1, \dots, X_n náhodný výběr z rozdělení se střední hodnotou $\mu = EX \in \mathbb{R}$ a rozptylem $\sigma^2 = \text{var} X < \infty$. Pak výběrový průměr je nestranný a konzistentní odhad $\mu = EX$ a výběrový rozptyl je nestranný a konzistentní odhad rozptylu $\sigma^2 = \text{var} X$.

Důkaz: Z věty 3.14 máme nestrannost \bar{X}_n , a protože předpokládáme $\mu \in \mathbb{R}$, tak ze SZVČ pro stejně rozdělené náhodné veličiny máme $\bar{X}_n \xrightarrow[n \rightarrow \infty]{\text{sj}} EX = \mu$, tedy \bar{X}_n je silně konzistentní.

Pro S_n^2 si nejdříve upravíme

$$\begin{aligned} S_n^2 &= \frac{1}{n-1} \sum_{k=1}^n (X_k - \mu - (\bar{X}_n - \mu))^2 \\ &= \frac{1}{n-1} \sum_{k=1}^n (X_k - \mu)^2 - 2 \frac{1}{n-1} \sum_{k=1}^n (X_k - \mu)(\bar{X}_n - \mu) + \frac{n}{n-1} (\bar{X}_n - \mu)^2 \\ &= \frac{1}{n-1} \sum_{k=1}^n (X_k - \mu)^2 - \frac{n}{n-1} (\bar{X}_n - \mu)^2. \end{aligned} \quad (5.2)$$

Tedy

$$E_{(\mu, \sigma^2)} S_n^2 = \frac{1}{n-1} \sum_{k=1}^n E_{(\mu, \sigma^2)} (X_k - \mu)^2 - \frac{n}{n-1} E_{(\mu, \sigma^2)} (\bar{X}_n - \mu)^2 = \frac{n}{n-1} \sigma^2 - \frac{n}{n-1} \frac{1}{n} \sigma^2 = \sigma^2.$$

V předposlední rovnosti jsme využili, že $E_{(\mu, \sigma^2)} (\bar{X}_n - \mu)^2 = \text{var}_{(\mu, \sigma^2)} \bar{X}_n = \frac{1}{n} \text{var}_{(\mu, \sigma^2)} X_1$. Takže S_n^2 je nestranný odhad σ^2 .

Abychom ověřili konzistenci, musíme prozkoumat limity obou členů z (5.2). Neboť $\text{var } X_k \in \mathbb{R}$, a $(X_k - \mu)^2$ jsou vzájemně nezávislé náhodné veličiny, tak ze SZVČ pro stejně rozdělené náhodné veličiny máme

$$\frac{1}{n} \sum_{k=1}^n (X_k - \mu)^2 \xrightarrow[n \rightarrow \infty]{\text{sj}} E_{(\mu, \sigma^2)} (X_k - \mu)^2 = \sigma^2.$$

Protože $\frac{n}{n-1} \xrightarrow[n \rightarrow \infty]{} 1$, tak i $\frac{1}{n-1} \sum_{k=1}^n (X_k - \mu)^2 \xrightarrow[n \rightarrow \infty]{\text{sj}} \sigma^2$, tedy první člen konverguje, kam má. Nyní ukážeme, že druhý člen jde skoro jistě k 0, čímž dokážeme silnou konzistenci odhadu S_n^2 .

Ze SZVČ pro stejně rozdělené náhodné veličiny máme $(\bar{X}_n - \mu) \xrightarrow[n \rightarrow \infty]{\text{sj}} 0$, a z věty 4.6 o spojitě transformaci pro konvergenci skoro jistě máme $(\bar{X}_n - \mu)^2 \xrightarrow[n \rightarrow \infty]{\text{sj}} 0$. Protože $\frac{n}{n-1} \xrightarrow[n \rightarrow \infty]{} 1$, dostáváme dohromady $\frac{n}{n-1} (\bar{X}_n - \mu)^2 \xrightarrow[n \rightarrow \infty]{\text{sj}} 0$, což zbývalo dokázat. \square

Máme tedy rozumné odhady pro momenty $E X$ a $\text{var } X$. Ale jak najít rozumné (případně vůbec nějaké) odhady parametrů v modelu $\{P_\theta : \theta \in \Theta\}$? Ukážeme si dvě metody.

*Zde končí
předn. 19
a půl
(30.4.)*

5.1.1 METODA MOMENTŮ

Myšlenka je tato: umíme z pozorování náhodného výběru odhadnout momenty rozdělení, ze kterého náhodný výběr pochází (pro první dva momenty – střední hodnotu a rozptyl, viz předchozí větu). A momenty rozdělení z nějakého parametrického modelu $\mathcal{F} = \{P_\theta : \theta \in \Theta\}$ na hodnotě parametru typicky nějakým jednoduchým způsobem závisí. Pokud (pro několik prvních momentů) položíme do rovnosti teoretický výraz pro hodnotu momentu (tj. funkci parametru θ) a empirický odhad tohoto momentu z dat, dostaneme rovnici (resp. soustavu rovnic), po jejichž vyřešení obdržíme předpis pro $\hat{\theta}$, který by mohl sloužit jako rozumný bodový odhad θ . Obvykle potřebujeme tolik rovnic, kolik má θ složek.

Značení. Buď X_1, \dots, X_n náhodný výběr z rozdělení P_θ . Označme si $m_r(\theta) = E X^r$, $r \in \mathbb{N}$, r -tý moment náhodné veličiny X s rozdělením P_θ (pokud existuje). A označme si

$$\widehat{m_r(\theta)} = \frac{1}{n} \sum_{k=1}^n X_k^r,$$

r -tý výběrový moment spočítaný z náhodného výběru X_1, \dots, X_n .

Nechť pro $X \sim P_\theta$ platí $E|X|^r < \infty$. Náhodný výběr z P_θ jsou nezávislé, stejně rozdělené náhodné veličiny s rozdělením P_θ , tedy zřejmě platí

$$E \widehat{m_r(\theta)} = \frac{1}{n} \sum_{k=1}^n E X_k^r = \frac{n}{n} E X^r = E X^r,$$

neboť $E X^r \in \mathbb{R}$. Navíc, pokud jsou $\{X_i\}_{i=1}^\infty$ vzájemně nezávislé a stejně rozdělené náhodné veličiny s rozdělením P_θ , tak i $\{X_k^r\}_{k=1}^\infty$ jsou vzájemně nezávislé a stejně rozdělené náhodné veličiny s $E X_k^r \in \mathbb{R}$. Tedy splňují předpoklady SZVČ 4.10 pro stejně rozdělené náhodné veličiny a posloupnost odhadů $\widehat{m_r(\theta)}$ konverguje s rostoucím rozsahem výběru n skoro jistě k $m_r(\theta) = E X^r$. Takže jsme dokázali následující tvrzení:

Tvrzení 5.2 Pokud pro náhodnou veličinu X s rozdělením P_θ platí $E|X|^r < \infty$, pak r -tý výběrový moment $\widehat{m_r(\theta)}$ je nestranný a silně konzistentní odhad $m_r(\theta) = E X^r$.

Tedy výběrové momenty jsou „rozumnými“ bodovými odhady teoretických momentů rozdělení $X \sim P_\theta$.

Momentový odhad $\hat{\theta}$ parametru θ najdeme jako řešení soustavy momentových odhadovacích rovnic

$$m_r(\hat{\theta}) = \widehat{m_r(\theta)}, \quad r = 1, 2, \dots, d.$$

Zde d je typicky dimenze θ , a nebo kolik je potřeba, abychom mohli soustavu vyřešit.

Příklad. Bud' $\theta \in \mathbb{R}$, tedy neznámý parametr je jednorozměrný, a bud' funkce $m_1(\theta)$ prostá. Pak stačí použít jednu momentovou odhadovací rovnici $m_1(\hat{\theta}) = \widehat{m_1(\theta)}$, a $\hat{\theta} = m_1^{-1}(\widehat{m_1(\theta)})$ je momentový odhad parametru θ (pokud existuje).

Příklad. Bud' model $\mathcal{F} = \{P_\theta, \theta \in \Theta\} = \{Rovnom[0, a], a \in \mathbb{R}^+\}$, neboli $\theta = a$ a $\Theta = \mathbb{R}^+$. Výpočtem zjistíme, že pro $X \sim Rovnom[0, a]$ je $E X = \frac{a}{2}$, tedy momentová odhadovací rovnice je

$$\frac{\hat{a}}{2} = \bar{X}_n.$$

Jejím vyřešením zjistíme $\widehat{a}_n = 2\bar{X}_n$ (zde jsme přidali dolní index n k odhadu \hat{a} , jako indikaci rozsahu výběru). Protože odhad \widehat{a}_n je lineární funkcí výběrového průměru, snadno dostáváme, že je to nestranný a silně konzistentní odhad parametru $a = 2E X$ (což je ta samá lineární funkce střední hodnoty $E X$, kterou \bar{X}_n nestranně a konzistentně odhaduje (věta 5.1)).

Je to vždy tak jednoduché? Ne nutně.

Příklad. Bud' model $\mathcal{F} = \{P_\theta, \theta \in \Theta\} = \{Rovnom[-a, a], a \in \mathbb{R}^+\}$. Tedy opět $\theta = a$ a $\Theta = \mathbb{R}^+$.

Ovšem $E X = 0$ pro $X \sim Rovnom[-a, a]$, tedy první moment nelze pro momentovou odhadovací rovnici použít. Heuristicky: $E_a X$ na hodnotě parametru $\theta = a$ nezávisí,

a tedy ho nelze využít k odhadu θ . Ovšem pozor! neznamena to, že rozdělení \bar{X}_n nemůže na $\theta = a$ záviset. To na $\theta = a$ záviset může a nějakou informaci o $\theta = a$ může nést (a typicky ponese).

Zkusme tedy druhý moment:

$$E_a X^2 = \int_{-a}^a \frac{x^2}{2a} dx = 2 \int_0^a \frac{x^2}{2a} dx = \left[\frac{1}{a} \frac{x^3}{3} \right]_0^a = \frac{a^2}{3}. \quad (5.3)$$

Momentová odhadovací rovnice založená na druhém momentu tedy bude

$$\frac{(\widehat{a})^2}{3} = \overline{X_n^2} = \frac{1}{n} \sum_{i=1}^n X_i^2,$$

a řešení je $\widehat{a}_n = \sqrt{3 \overline{X_n^2}}$. A jak je to s vlastnostmi tohoto odhadu?

Ze SZVČ 4.10 pro stejně rozdělené náhodné veličiny aplikovaného na posloupnost náhodných veličin $\{X_i^2\}_{i=1}^\infty$ (čtenář si ověří, že předpoklady jsou splněny) a rovnice (5.3) dostáváme $\overline{X_n^2} \xrightarrow[n \rightarrow \infty]{\text{sj}} \frac{a^2}{3}, \forall a \in \mathbb{R}^+$. A z věty 4.6 o spojitě transformaci pro konvergenci skoro jistě dostaneme $\widehat{a}_n = \sqrt{3 \overline{X_n^2}} \xrightarrow[n \rightarrow \infty]{\text{sj}} a, \forall a \in \mathbb{R}^+$, tedy \widehat{a}_n je silně konzistentní odhad parametru a .

Z tvrzení 5.2 víme $E \overline{X_n^2} = \frac{a^2}{3}$, ovšem $E \sqrt{3 \overline{X_n^2}} \neq a$ (čtenář si dokáže z Hölderovy nerovnosti 3.10 pro $p = q = 2$ a náhodné veličiny $\sqrt{3 \overline{X_n^2}}$ a 1). Takže \widehat{a}_n není nestranný odhad parametru a . Mohli bychom dále zkoumat vychýlení odhadu $E(\sqrt{3 \overline{X_n^2}} - a)$, což ovšem na tomto místě dělat nebudeme.

Příklad. Buď model $\mathcal{F} = \{N(\mu, \sigma^2) : \mu \in \mathbb{R}, \sigma^2 > 0\}$, tedy $\theta = (\mu, \sigma^2)$ a $\Theta = \mathbb{R} \times \mathbb{R}^+$ – třída všech normálních rozdělení a $d = 2$. Víme, že $EX = \mu$ a $EX^2 = \mu^2 + \sigma^2$. Momentové odhadovací rovnice tedy jsou

$$\widehat{\mu} = \overline{X}_n \quad \widehat{\sigma^2} + (\widehat{\mu})^2 = \overline{X_n^2},$$

jejich vyřešením dostaneme

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

Už víme, že $\widehat{\mu}_n$ je silně konzistentní a nestranný odhad $\mu = EX$ (z tvrzení 5.2).

$\widehat{\sigma}_n^2$ je silně konzistentní odhad σ^2 (čtenář si dokáže buď z tvrzení 5.2 a věty o spojitě transformaci pro $\xrightarrow[n \rightarrow \infty]{\text{sj}}$ nebo z věty 5.1). Z věty 5.1 je také zřejmo, že $\widehat{\sigma}_n^2$ je asymptoticky nestranný, ovšem ne nestranný odhad σ^2 .

Jak je vidět z příkladů výše, vlastnosti nestrannosti a silné konzistence se z výběrových momentů nutně nemusí přenést na odhady parametrů odvozené metodou momentů. Pro odhady parametrů odvozené metodou momentů je tedy vždy třeba ověřit/dokázat jejich vlastnosti znovu přímo.

5.1.2 METODA MAXIMÁLNÍ VĚROHODNOSTI

Bud' X_1, \dots, X_n náhodný výběr z rozdělení P_θ . A necht' f_θ je hustota P_θ vzhledem k vhodné referenční míře ν . Pak sdružené rozdělení náhodného výběru má hustotu $f_\theta(x_1, \dots, x_n) = \prod_{i=1}^n f_\theta(x_i)$, $(x_1, \dots, x_n) \in \mathbb{R}^n$, vzhledem k ν^n (neboť složky náhodného výběru jsou vzájemně nezávislé náhodné veličiny). Z kapitoly 2 víme, že pro absolutně spojitě rozdělení P_θ je $\nu = \lambda$, pro diskrétní rozdělení P_θ je ν vhodná číselná míra. Hustota $f_\theta(\mathbf{x})$ na θ samozřejmě závisí, a pro stejné pozorování \mathbf{x} náhodného výběru \mathbf{X} z rozdělení P_θ má různou hodnotu. Toho využívá metoda maximální věrohodnosti. Nejdřív si ale zdefinujme, co je to věrohodnost.

Definice 5.10 Bud' $\mathcal{F} = \{P_\theta, \theta \in \Theta\}$ parametrický model, kde P_θ mají hustoty f_θ vzhledem ke stejné referenční míře ν . Pak hustotu $f_\theta(x_1, \dots, x_n)$ náhodného výběru X_1, \dots, X_n z P_θ , nahlíženou jako funkci θ , nazveme *věrohodností* $L(\theta; \mathbf{x})$ (případně věrohodností pro pozorování \mathbf{x}), neboli

$$L(\theta; \mathbf{x}) = f_\theta(x_1, \dots, x_n) = \prod_{i=1}^n f_\theta(x_i), \quad \mathbf{x} = (x_1, \dots, x_n) \in \mathbb{R}^n, \quad \theta \in \Theta.$$

$l(\theta; \mathbf{x}) = \log(L(\theta; \mathbf{x}))$ se nazývá *logaritmickou věrohodností*.

Bud' $\mathcal{F} = \{P_\theta, \theta \in \Theta\}$ parametrický model, X_1, \dots, X_n náhodný výběr z P_θ a $\mathbf{x} = \mathbf{X}(\omega)$ pozorování realizace náhodného výběru. Odhad parametru θ *metodou maximální věrohodnosti* je definován jako

$$\varphi_n(\mathbf{x}) = \arg \max_{\theta \in \Theta} L(\theta; \mathbf{x}).$$

Samozřejmě, pokud existuje. Existovat obecně nemusí, ani obecně nemusí být jednoznačný. Odvození probíhá pro \mathbf{x} , je možné, že pro většinu hodnot \mathbf{x} z oboru hodnot X_1, \dots, X_n je $\arg \max_{\theta \in \Theta} L(\theta; \mathbf{x})$ dobře definováno, ale pro nějaké výjimky ne. Pak je potřeba upravit předpis pro φ_n tak, aby byl dobře definován pro celý obor hodnot X_1, \dots, X_n , abychom dostali zobrazení φ_n splňující definici 5.6.

Věrohodnost $L(\theta; \mathbf{x})$ je funkce, která opravdu odpovídá naší intuitivní představě „věrohodnosti“, jak moc věrohodná v dané náhodné situaci konkrétní hodnota $\theta = \theta_0$ je. Pro jednu konkrétní hodnotu pozorování $\mathbf{x} = \mathbf{X}(\omega)$ (ovšem v $\omega \in \Omega$, které neznáme) vyšší hodnota „hustoty“ $L(\theta_0; \mathbf{x})$ pro hodnotu θ_0 než $L(\theta_1; \mathbf{x})$ v θ_1 vlastně odpovídá větší šanci „pozorovat přesně to pozorování \mathbf{x} , které jsme pozorovali“, když $X \sim P_{\theta_0}$ než když $X \sim P_{\theta_1}$. A odhad metodou maximální věrohodnosti vlastně hledá to θ , pro které byla „největší šance“ pozorovat přesně to pozorování \mathbf{x} , které jsme pozorovali.

K hledání $\arg \max_{\theta \in \Theta} L(\theta; \mathbf{x})$ můžeme samozřejmě využít našich znalostí matematické analýzy, a pokud je $L(\theta; \mathbf{x})$ diferencovatelná, tak hledat

$$\arg \max_{\theta \in \Theta} L(\theta; \mathbf{x}) = \arg \max_{\theta \in \Theta} l(\theta; \mathbf{x})$$

jako řešení soustavy věrohodnostních rovnic

$$\frac{\partial}{\partial \theta_j} \sum_{i=1}^n \log f_{\theta}(x_i) = 0, \quad j = 1, \dots, d, \quad \text{pro } \theta \in \Theta \subseteq \mathbb{R}^d,$$

kde d je dimenze (vektorového) parametru θ .

Příklad. Buď model $\mathcal{F} = \{N(\mu, \sigma^2) : \mu \in \mathbb{R}, \sigma^2 > 0\}$, tedy $\theta = (\mu, \sigma^2)$ a $\Theta = \mathbb{R} \times \mathbb{R}^+$ – třída všech normálních rozdělení a $d = 2$. Normální rozdělení je absolutně spojitě, tedy referenční míra $\nu = \lambda$ a

$$f_{\theta}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad x \in \mathbb{R},$$

takže

$$L(\theta; \mathbf{x}) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}$$

a

$$l(\theta; \mathbf{x}) = \sum_{i=1}^n \left(-\frac{1}{2} \log(2\pi\sigma^2) - \frac{(x_i - \mu)^2}{2\sigma^2} \right) = -\frac{n}{2} \log(2\pi) - \frac{n}{2} \log(\sigma^2) - \frac{\sum_{i=1}^n (x_i - \mu)^2}{2\sigma^2}.$$

Derivace jsou

$$\frac{\partial l(\theta; \mathbf{x})}{\partial \mu} = \frac{2 \sum_{i=1}^n (x_i - \mu)}{2\sigma^2}, \quad \frac{\partial l(\theta; \mathbf{x})}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{\sum_{i=1}^n (x_i - \mu)^2}{2(\sigma^2)^2},$$

takže věrohodnostní rovnice jsou

$$\sum_{i=1}^n x_i - n \mu = 0 \quad \sum_{i=1}^n (x_i - \mu)^2 - n \sigma^2 = 0.$$

Vyřešením získáme $\hat{\mu}_n = \bar{X}_n$, $\hat{\sigma}_n^2 = \frac{\sum_{i=1}^n (X_i - \bar{X}_n)^2}{n}$, stejné odhady jako metodou momentů. Takže oba odhady jsou konzistentní a $\hat{\mu}_n$ je také nestranný.

Odhady metodou maximální věrohodnosti a metodou momentů se v tom samém modelu mohou shodovat (jako v předchozím příkladě) a nebo lišit (jako v příkladě následujícím).

Příklad. Buď model $\mathcal{F} = \{P_{\theta}, \theta \in \Theta\} = \{Rovnom[0, a], a \in \mathbb{R}^+\}$, neboli $\theta = a$ a $\Theta = \mathbb{R}^+$. Rovnoměrné rozdělení je absolutně spojitě, tedy referenční míra $\nu = \lambda$ a

$$f_a(x) = \frac{1}{a} \mathbb{1}_{[0, a]}(x), \quad x \in \mathbb{R}.$$

Tedy věrohodnost je

$$L(a; \mathbf{x}) = \prod_{i=1}^n \frac{1}{a} \mathbb{1}_{[0,a]}(x_i) = \frac{1}{a^n} \mathbb{1} \left(\max_{i=1, \dots, n} x_i \leq a \right) \mathbb{1} \left(0 \leq \min_{i=1, \dots, n} x_i \right).$$

Ta je zřejmě maximalizována pro $\max_{i=1, \dots, n} x_i$, tedy $\hat{a}_n = \max_{i=1, \dots, n} X_i$. Pro další použití si tento odhad označme také $U_n = \max_{i=1, \dots, n} X_i$.

Vidíme, že odhad U_n je slabě konzistentní, neboť

$$\begin{aligned} P(|U_n - a| > \epsilon) &= P(U_n > a + \epsilon) + P(U_n < a - \epsilon) = P\left(\bigcap_{i=1}^n \{X_i < a - \epsilon\}\right) \\ &= \prod_{i=1}^n P(X_i < a - \epsilon) = \left(\frac{a - \epsilon}{a}\right)^n \xrightarrow{n \rightarrow \infty} 0, \quad \forall a \in \mathbb{R}^+. \end{aligned}$$

Využili jsme, že $P(\max_{i=1, \dots, n} X_i > a) = 0$ ve druhé rovnosti, a nezávislost X_i ve třetí rovnosti.

Abychom ukázali silnou konvergenci, musíme zjistit, jestli $P(\lim_{n \rightarrow \infty} \hat{a}_n = a) = 1$. Můžeme využít speciální vlastnost odhadu \hat{a}_n , že je neklesající v n , $\forall \omega \in \Omega, a \in [0, 1]$. Takže platí

$$P\left(\lim_{n \rightarrow \infty} \hat{a}_n < c\right) = P\left(\bigcap_{n=1}^{\infty} \{\hat{a}_n < c\}\right) = \lim_{n \rightarrow \infty} P(\hat{a}_n < c), \quad (5.4)$$

ze spojitosti pravděpodobnosti, neboť jevy $B_n = \{\hat{a}_n < c\}$ jsou také monotónní. Počítáme tedy

$$\begin{aligned} P\left(\lim_{n \rightarrow \infty} \hat{a}_n = a\right) &= 1 - P\left(\lim_{n \rightarrow \infty} \hat{a}_n < a\right) = 1 - P\left(\bigcup_{m=1}^{\infty} \left\{\lim_{n \rightarrow \infty} \hat{a}_n < a - \frac{1}{m}\right\}\right) \\ &= 1 - \lim_{m \rightarrow \infty} P\left(\lim_{n \rightarrow \infty} \hat{a}_n < a - \frac{1}{m}\right) = 1 - \lim_{m \rightarrow \infty} \lim_{n \rightarrow \infty} P\left(\hat{a}_n < a - \frac{1}{m}\right) = 1 - \lim_{m \rightarrow \infty} 0 = 1. \end{aligned}$$

Využili jsme, že když je nerovnost $\lim_{n \rightarrow \infty} \hat{a}_n(\omega) < a$ ostrá, musí existovat nějaké $m \in \mathbb{N}$ takové, že platí i $\lim_{n \rightarrow \infty} \hat{a}_n(\omega) < a - \frac{1}{m}$. Proto platí druhá rovnost. Jevy $\{\lim_{n \rightarrow \infty} \hat{a}_n < a - \frac{1}{m}\}$ jsou monotónní v m , takže ve třetí rovnosti jsme použili spojitost pravděpodobnosti. Ve čtvrté jsme použili (5.4). Dokázali jsme tedy, že $U_n = \hat{a}_n$ je i silně konzistentní odhad a .

Pro střední hodnotu můžeme spočítat

$$E_a U_n = E_a \max_{i=1, \dots, n} X_i = \int_0^a x n x^{n-1} a^{-n} dx = \frac{n}{n+1} a,$$

neboť snadno odvodíme hustotu $\max_{i=1, \dots, n} X_i$ pomocí metod z kapitoly 3. Odhad U_n je tedy jen asymptoticky nestranný odhad a , není nestranný. Zadefinujeme si ještě jiný odhad

$$V_n = \frac{n+1}{n} U_n = \frac{n+1}{n} \max_{i=1, \dots, n} X_i.$$

Ten je nestranný a konzistentní (čtenář si dokáže sám, pomocí toho, co ví o U_n).

V předchozí sekci 5.1.1 jsme odvodili ve stejném modelu momentový odhad $T_n = 2\bar{X}_n$ parametru a . Ten byl také nestranný a silně konzistentní. Tedy oba odhady jsou centrované kolem „správného“ a a pro $n \rightarrow \infty$ k němu konvergují skoro jistě. Nabízí se otázka, který odhad je lepší. Když spočítáme rozptyl obou odhadů, zjistíme, že

$$\text{var } V_n = \frac{a^2}{n(n+2)} \quad \text{var } T_n = \frac{a^2}{3n}.$$

Takže variabilita („nejistota odhadu“) odhadu V_n jde k 0 řádově rychleji, než pro T_n . Tedy V_n je lepší odhad parametru a .

Poznámka. Existuje věta o nejlepším nestranném odhadu (tj. s nejmenším rozptylem), který existuje za určitých předpokladů na \mathcal{F} a její prvky. Jsou ale i situace, ve kterých existuje vychýlený odhad se (značně) menší střední čtvercovou chybou $E_\theta(\hat{\theta}_n - \theta)^2$ (resp. viz (5.1) pro případ odhadu parametrické funkce θ), než má nejlepší nestranný odhad. Toto už přesahuje možnosti naší přednášky a bude podrobněji prozkoumáno v přednáškách Matematická statistika 1 a 2.

Střední čtvercová chyba je průměrná (přes všechna možná pozorování náhodného výběru) kvadratická odchylka mezi odhadem $\hat{\theta}_n$ a tím, co má odhadovat. Samozřejmě závisí na hodnotě neznámého θ . Můžeme si ji rozepsat

$$E_\theta(\hat{\theta}_n - \theta)^2 = E_\theta(\hat{\theta}_n - E_\theta \hat{\theta}_n)^2 + E_\theta(E_\theta \hat{\theta}_n - \theta)^2 = \text{var}_\theta \hat{\theta}_n + (E_\theta \hat{\theta}_n - \theta)^2.$$

Závisí tedy na obojím - rozptylu $\hat{\theta}_n$ i vychýlení $E_\theta \hat{\theta}_n - \theta$. Extrémní příklad toho, jak by to dopadlo, kdybychom brali ohled jen na rozptyl $\hat{\theta}_n$ je následující:

Příklad. Bud' X náhodný výběr z modelu $\mathcal{F} = \{Alt(p) : p \in [0, 1]\}$. A definujme $\hat{p}_n = \frac{1}{2}$. Pak zřejmě $\text{var}(\hat{p}_n) = 0$, tedy minimální možný. Ale vychýlení

$$E_p(\hat{p}_n - p) = \frac{1}{2} - p, \quad \forall p \in [0, 1],$$

tedy může být i velmi velké, podle toho, jaká je skutečná neznámá hodnota parametru p . A to nezávisle na rozsahu výběru n a i nezávisle na datech – odhad je totiž vůbec nevyužívá. Toto není dobrý odhad a zřejmě není konzistentní.

V předchozím jsme se zaměřili na metody odvození odhadu přímo parametru θ . Ale co když potřebuji odhadnout parametrickou funkci $g(\theta)$? Pak se jako rozumná strategie jeví definovat

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

Pokud byl $\hat{\theta}_n$ „rozumný“, je šance, že i $\widehat{g(\theta)}_n$ by mohl být „rozumný“. Například, pokud je g spojitá, tak se konzistence $\hat{\theta}_n$ přenesení na odhad $\widehat{g(\theta)}_n$. Další obecná diskuse této otázky už je ale mimo rozsah našeho kurzu a zájemce opět odkazujeme na přednášky Matematická statistika 1 a 2.

*Zde končí
předn. 20
a půl
(7.5.)*

Obecná nevýhoda bodových odhadů je, že nám nedávají žádnou představu o spolehlivosti našeho odhadu, resp. o velikosti chyby, které se při bodovém odhadu parametru z náhodného výběru dopouštíme. Proto je mnohem rozumnější používat intervalové odhady.

5.2 INTERVALOVÝ ODHAD

Definice 5.11 Bud' \mathbf{X} náhodný výběr o rozsahu n z P_θ a $\alpha \in (0, 1)$. *Intervalovým odhadem* parametrické funkce $g(\theta)$ o *spolehlivosti* $(1 - \alpha)$ nazveme dvojici borelovských funkcí $\eta_L : \mathbb{R}^n \rightarrow \mathbb{R}$ a $\eta_U : \mathbb{R}^n \rightarrow \mathbb{R}$, jejichž předpis nezávisí na θ , a $\forall \theta \in \Theta$ platí

$$P_\theta(\eta_L(\mathbf{X}) < g(\theta) < \eta_U(\mathbf{X})) \geq 1 - \alpha.$$

Poznámka. $\eta_L(\mathbf{X})$ a $\eta_U(\mathbf{X})$ jsou náhodné veličiny. Interval chceme co nejuzší (chceme získat o $g(\theta)$ maximum informace). Na druhou stranu, α také chceme co nejmenší (chceme co největší spolehlivost). Tyto dva požadavky jdou proti sobě, takže je potřeba volit vhodný kompromis.

Poznámka. α volíme malé, typicky 1%, 5% nebo 10%.

Spolehlivost $(1 - \alpha)$ NEZNAMENÁ, že θ padne do $(\eta_L(\mathbf{X}), \eta_U(\mathbf{X}))$ s šancí $(1 - \alpha)$! θ totiž není náhodné, jen neznámé. Spolehlivost např. 95% znamená, že v 95% všech pozorovaných náhodných výběrů z nich spočítaný intervalový odhad $(\eta_L(\mathbf{X}), \eta_U(\mathbf{X}))$ překryje neznámou (ale pevnou) hodnotu θ .

V definici výše byl představen tzv. oboustranný intervalový odhad. Někdy se používají i jednostranné intervalové odhady.

Definice 5.12 Bud' \mathbf{X} náhodný výběr o rozsahu n z P_θ a $\alpha \in (0, 1)$. *Dolním intervalovým odhadem* parametrické funkce $g(\theta)$ o *spolehlivosti* $(1 - \alpha)$ nazveme borelovskou funkci $\eta_D : \mathbb{R}^n \rightarrow \mathbb{R}$, jejíž předpis nezávisí na θ a $\forall \theta \in \Theta$ platí

$$P_\theta(\eta_D(\mathbf{X}) < g(\theta)) \geq 1 - \alpha.$$

Horním intervalovým odhadem parametrické funkce $g(\theta)$ o *spolehlivosti* $(1 - \alpha)$ nazveme borelovskou funkci $\eta_H : \mathbb{R}^n \rightarrow \mathbb{R}$, jejíž předpis nezávisí na θ a $\forall \theta \in \Theta$ platí

$$P_\theta(g(\theta) < \eta_H(\mathbf{X})) \geq 1 - \alpha.$$

A jak zkonstruovat intervalový odhad? Obecný postup je možno popsat následovně:

- Najdeme funkci $H(X_1, \dots, X_n; \theta)$ takovou, že rozdělení $H(X_1, \dots, X_n; \theta)$ nezávisí na θ . Označme distribuční funkci tohoto rozdělení F_H . A pro jednoduchost budeme předpokládat, že F_H je spojitá.
- Najdeme co nejkratší interval (q_L, q_U) takový, že $F_H(q_U) - F_H(q_L) \geq 1 - \alpha$. Pak (díky spojitosti F_H) bude také platit

$$P_\theta(q_L < H(X_1, \dots, X_n; \theta) < q_U) \geq 1 - \alpha, \quad \forall \theta \in \Theta.$$

- Nyní oddělíme v předpisu H od sebe $g(\theta)$ a náhodný výběr (X_1, \dots, X_n) tak, aby

$$q_L < H(X_1, \dots, X_n; \theta) < q_U \quad \Leftrightarrow \quad \eta_L(X_1, \dots, X_n) < g(\theta) < \eta_U(X_1, \dots, X_n).$$

- Pokud se to povedlo, máme intervalový odhad $(\eta_L(\mathbf{X}), \eta_U(\mathbf{X}))$ parametrické funkce $g(\theta)$ o spolehlivosti $(1 - \alpha)$.

Příklad. Uvažujme opět model $\mathcal{F} = \{\text{Rovnom}[0, a], a \in \mathbb{R}^+\}$, ale nyní chceme odvodit intervalový odhad pro parametr a .

Budeme se držet postupu výše a zkusíme za pomoci specifických vlastností rovnoměrného rozdělení najít vhodnou funkci H . Víme, že $X_i \in [0, a]$, $i = 1, \dots, n$, skoro jistě a tedy $\frac{X_i}{a} \in [0, 1]$, $i = 1, \dots, n$, skoro jistě. Snadno odvodíme

$$P_a\left(\frac{X_i}{a} \leq x\right) = P_a(X_i \leq ax) = \begin{cases} 0, & x < 0 \\ \frac{ax}{a} = x, & x \in [0, 1] \\ 1, & x > 1. \end{cases}$$

Tedy $Y_i = \frac{X_i}{a} \sim \text{Rovnom}[0, 1]$, a to nezávisí na hodnotě a . V předchozím se nám v tomto modelu osvědčilo použít maximum z pozorování v náhodném výběru, zkusíme ho použít i teď. Ovšem na Y_i .

Z předchozího víme, že pro $\max_{i=1, \dots, n} Y_i$ platí $P\left(\max_{i=1, \dots, n} Y_i \leq x\right) = \left(\frac{x}{1}\right)^n = x^n$, $x \in [0, 1]$.

Volíme

$$H(X_1, \dots, X_n; a) = \frac{1}{a} \max_{i=1, \dots, n} X_i,$$

takže máme

$$F_H(x) = \begin{cases} 0, & x < 0 \\ x^n, & x \in [0, 1] \\ 1, & x > 1, \end{cases}$$

což nezávisí na a . Když si načrtneme hustotu rozdělení $H(X_1, \dots, X_n; a)$, vidíme, že nejkratší (q_L, q_U) získáme volbou $q_U = 1$ a q_L tak, aby

$$P_a\left(q_L < \frac{\max_{i=1, \dots, n} X_i}{a} < 1\right) = P_a\left(\max_{i=1, \dots, n} X_i < a < \frac{\max_{i=1, \dots, n} X_i}{q_L}\right) = 1 - \alpha.$$

Tedy $1 - F_H(q_L) = 1 - \alpha$, z toho $\alpha = F_H(q_L) = q_L^n$, a vyřešíme $q_L = \alpha^{\frac{1}{n}}$. Dohromady dostaneme, že intervalový odhad pro a o spolehlivosti $(1 - \alpha)$ je

$$\left(\max_{i=1, \dots, n} X_i, \alpha^{-\frac{1}{n}} \max_{i=1, \dots, n} X_i\right).$$

V předchozím příkladě je hodnota q_L zřejmě hodnota kvantilové funkce $F_H^{-1}(\alpha)$. Zavedeme si ještě jednu související definici.

Definice 5.13 Buď F distribuční funkce, spojitá a ryze rostoucí na $F^{-1}(0, 1)$, a buď $\beta \in (0, 1)$. β -kvantilem rozdělení s distribuční funkcí F nazveme hodnotu $q_\beta = F^{-1}(\beta)$.

Poznámka. Pro spojitou a ryze rostoucí F je tedy β -kvantil určen jednoznačně. Což ovšem neplatí obecně. Vzpomeňme si na medián – to je $\frac{1}{2}$ -kvantil $q_{\frac{1}{2}}$. I β -kvantil lze definovat obecněji, jako hodnotu q_β splňující

$$P(X \geq q_\beta) \geq 1 - \beta \quad \text{a zároveň} \quad P(X \leq q_\beta) \geq \beta.$$

Ta obecně není určena jednoznačně, ale $F^{-1}(\beta)$ této definici vždy vyhovuje. V našem kurzu ovšem obecnou definici β -kvantilu nebudeme potřebovat, protože budeme vždy v situaci, kdy F je spojitá a ryze rostoucí na $F^{-1}(0, 1)$.

Značení. β -kvantil normovaného normálního rozdělení značíme u_β , β -kvantil χ^2 -rozdělení o n stupních volnosti značíme $\chi_{\beta, n}^2 = \chi_n^2(\beta)$, β -kvantil Studentova t -rozdělení o n stupních volnosti značíme $t_{\beta, n} = t_n(\beta)$. Pro všechna tato rozdělení platí, že jejich distribuční funkce je spojitá a ryze rostoucí na $F^{-1}(0, 1)$.

Poznámka. Pokud je absolutně spojitě rozdělení symetrické kolem 0, tj. jeho hustota je sudá funkce, pak víme, že $F(x) = 1 - F(-x)$, a tedy pro kvantily platí $q_\beta = -q_{1-\beta}$, $\beta \in (0, 1)$. Speciálně tedy tyto rovnosti platí pro normované normální a Studentovo rozdělení.

Poznámka. U intervalových odhadů je typickou volbou $q_L = q_{\frac{\alpha}{2}}$ a $q_U = q_{1-\frac{\alpha}{2}}$, protože $F_H(q_{1-\frac{\alpha}{2}}) - F_H(q_{\frac{\alpha}{2}}) = 1 - \frac{\alpha}{2} - \frac{\alpha}{2} = 1 - \alpha$. Tedy dostaneme intervalový odhad o správné spolehlivosti α . Tato volba nemusí být optimální, ve smyslu minimální délky intervalu (q_L, q_U) . Je optimální např. pro unimodální rozdělení s hustotou symetrickou kolem modu.

*Zde končí
předn. 21
(14.5.)*

5.2.1 INTERVALOVÉ ODHADY V NORMÁLNÍM MODELU

V této sekci si odvodíme intervalové odhady pro parametry normálního modelu. Normální model se nejčastěji vyskytuje v aplikacích a díky specifickým vlastnostem normálního rozdělení (viz sekce 3.5.2) je možno odvodit intervalové odhady s přesnou spolehlivostí poměrně snadno.

Začneme situací se známým rozptylem:

Věta 5.3 Buď (X_1, \dots, X_n) náhodný výběr z modelu $\mathcal{F} = \{N(\mu, \sigma^2), \mu \in \mathbb{R}\}$, kde $\sigma^2 > 0$ je známá konstanta. Pak

$$\left(\bar{X}_n - u_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \bar{X}_n + u_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \right) \quad (5.5)$$

je oboustranný intervalový odhad parametru μ o spolehlivosti $(1 - \alpha)$.

Důkaz: Volbou $\mathbf{c} = (1/n, \dots, 1/n)$ v důsledku (iv) definice 3.11 mnohorozměrného normálního rozdělení dostáváme, že pokud je (X_1, \dots, X_n) náhodný výběr z normálního rozdělení $N(\mu, \sigma^2)$, $\mu \in \mathbb{R}$, $\sigma^2 > 0$, pak $\bar{X}_n \sim N\left(\mu, \frac{\sigma^2}{n}\right)$. Pomocí věty 2.12 o monotónní transformaci pak následně dostaneme, že

$$\frac{\bar{X}_n - \mu}{\sigma} \sqrt{n} \sim N(0, 1)$$

(nebo viz příklad na str. 16).

Volme tedy v obecném postupu pro odvození intervalového odhadu $H(X_1, \dots, X_n; \mu) = \frac{\bar{X}_n - \mu}{\sigma} \sqrt{n}$. Její rozdělení je $N(0, 1)$ a nezávisí na hodnotě μ . Tedy platí

$$P_\mu \left(u_{\frac{\alpha}{2}} < H(\mathbf{X}, \mu) < u_{1-\frac{\alpha}{2}} \right) = 1 - \alpha, \quad \forall \mu \in \mathbb{R}.$$

Použijeme symetrii normovaného normálního rozdělení, abychom přepsali $u_{\frac{\alpha}{2}} = -u_{1-\frac{\alpha}{2}}$, a ekvivalentně přepíšeme nerovnosti v závorce:

$$\begin{aligned} 1 - \alpha &= P_\mu \left(-u_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} < \bar{X}_n - \mu < u_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \right) = P_\mu \left(|\bar{X}_n - \mu| < u_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \right) \\ &= P_\mu \left(\bar{X}_n - u_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} < \mu < \bar{X}_n + u_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \right), \quad \forall \mu \in \mathbb{R}. \end{aligned}$$

Tím je věta dokázána. \square

Poznámka. Délka intervalového odhadu pro μ z předchozí věty je $2u_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$ a je nenáhodná. S rostoucím rozsahem výběru n pak konverguje (pro pevné α) k 0.

Ale co když neznáme σ^2 ? Pokud by nás zajímal jen intervalový odhad pro μ , pak by šlo uvažovat o nahrazení σ^2 jeho konzistentním odhadem S_n^2 a použití Cramérový-Sluckého věty. Tím bychom ale dostali intervalový odhad o nikoli přesné, ale jen asymptotické spolehlivosti $(1 - \alpha)$ (podrobněji o tomto přístupu viz další sekci). Pokud chceme odhadnout oba parametry μ i σ^2 , pak bude potřeba využít dalších speciálních vlastností (mnohorozměrného) normálního rozdělení.

Věta 5.4 Bud' \mathbf{X} náhodný vektor s rozdělením $N_n(0, \sigma^2 \mathbb{I}_n)$, kde $\sigma^2 > 0$. Bud' C ortonormální matice rozměru $n \times n$ (tj. platí $CC^T = C^T C = \mathbb{I}_n$). Položme $\mathbf{Y} = C\mathbf{X}$. Pak $\mathbf{Y} \sim N_n(0, \sigma^2 \mathbb{I}_n)$. Tedy $\{Y_i\}_{i=1}^n$ jsou vzájemně nezávislé se stejným rozdělením $N(0, \sigma^2)$.

Důkaz: Dosadíme ortonormální matici C do důsledku (iii) definice 3.11 mnohorozměrného normálního rozdělení. \square

Věta 5.5 Bud' (X_1, \dots, X_n) náhodný výběr z rozdělení $N(\mu, \sigma^2)$, $\mu \in \mathbb{R}$, $\sigma^2 \in \mathbb{R}^+$. Pak platí:

- (i) \bar{X}_n a $S_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$ jsou nezávislé náhodné veličiny,
- (ii) náhodná veličina $\frac{n-1}{\sigma^2} S_n^2 = \sum_{i=1}^n \frac{(X_i - \bar{X}_n)^2}{\sigma^2}$ má χ_{n-1}^2 rozdělení,

(iii) náhodná veličina $T_n = \frac{\bar{X}_n - \mu}{S_n} \sqrt{n}$ má Studentovo t_{n-1} rozdělení.

Důkaz: Větu stačí dokázat pro případ $\mu = 0$ a pak ji aplikovat na náhodné veličiny $(X_i - \mu)$.

Pro důkaz použijeme předchozí větu pro vhodnou ortonormální matici C . Volíme $c_{1,i} = \frac{1}{\sqrt{n}}$, $i = 1, \dots, n$. Zbytek matice volíme libovolně tak, aby celá C byla ortonormální (to lze). Položme $\mathbf{Y} = C\mathbf{X}$. Potom

$$Y_1 = \frac{1}{\sqrt{n}} \sum_{i=1}^n X_i = \sqrt{n} \bar{X}_n \sim N(0, \sigma^2).$$

Také platí

$$\sum_{i=1}^n Y_i^2 = \mathbf{X}^T C^T C \mathbf{X} = \mathbf{X}^T \mathbf{X} = \sum_{i=1}^n X_i^2,$$

takže

$$\sum_{i=2}^n Y_i^2 = \sum_{i=1}^n X_i^2 - Y_1^2 = \sum_{i=1}^n X_i^2 - n(\bar{X}_n)^2 = \sum_{i=1}^n (X_i - \bar{X}_n)^2 = (n-1)S_n^2.$$

Tedy $\bar{X}_n = \frac{Y_1}{\sqrt{n}}$ a $S_n^2 = \frac{\sum_{i=2}^n Y_i^2}{n-1}$ jsou vzájemně nezávislé, neboť jsou funkcemi Y_1 a $\{Y_2, \dots, Y_n\}$, které jsou vzájemně nezávislé. Tím je dokázán bod (i).

Z věty 3.8 máme, že $\left\{\frac{Y_i}{\sigma}\right\}_{i=1}^n$ jsou vzájemně nezávislé a všechny jsou $N(0, 1)$ rozdělené. Tedy $\sum_{i=2}^n \frac{Y_i^2}{\sigma^2} = \frac{n-1}{\sigma^2} S_n^2 \sim \chi_{n-1}^2$ z definice χ^2 rozdělení. Tím je dokázán bod (ii).

Lineární transformací vzájemně nezávislých \bar{X}_n a S_n^2 dostaneme také vzájemně nezávislé náhodné veličiny U a V

$$U = \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} = \frac{Y_1}{\sigma} \sim N(0, 1) \quad V = \frac{(n-1)S_n^2}{\sigma^2} \sim \chi_{n-1}^2.$$

Tedy $T_n = \frac{U}{\sqrt{\frac{V}{n-1}}} \sim t_{n-1}$ z definice Studentova t -rozdělení. Tím je dokázán bod (iii). \square

Věta 5.6 Buď (X_1, \dots, X_n) náhodný výběr z modelu $\mathcal{F} = \{N(\mu, \sigma^2), \mu \in \mathbb{R}, \sigma^2 > 0\}$. Pak

(i)

$$\left(\bar{X}_n - t_{n-1}(1-\alpha/2) \frac{S_n}{\sqrt{n}}, \bar{X}_n + t_{n-1}(1-\alpha/2) \frac{S_n}{\sqrt{n}} \right) \quad (5.6)$$

je oboustranný intervalový odhad parametru μ o spolehlivosti $(1-\alpha)$.

(ii)

$$\left(\frac{(n-1)S_n^2}{\chi_{n-1}^2(1-\frac{\alpha}{2})}, \frac{(n-1)S_n^2}{\chi_{n-1}^2(\frac{\alpha}{2})} \right)$$

je oboustranný intervalový odhad parametru σ^2 o spolehlivosti $(1-\alpha)$.

Používáme značení kvantilů χ^2 a Studentova t -rozdělení ze strany 83.

Důkaz: (i) Postupujeme podle obecného postupu odvození intervalového odhadu a volíme $H(\mathbf{X}; \mu) = \frac{\bar{X}_n - \mu}{S_n} \sqrt{n}$. To má podle věty 5.5 bod (iii) t_{n-1} rozdělení, které nezávisí na parametrech (μ, σ^2) a je symetrické kolem 0. Proto

$$\begin{aligned} 1 - \alpha &= P_{\mu, \sigma^2} \left(t_{n-1}(\alpha/2) < \frac{\bar{X}_n - \mu}{S_n} \sqrt{n} < t_{n-1}(1 - \alpha/2) \right) \\ &= P_{\mu, \sigma^2} \left(\bar{X}_n - t_{n-1}(1 - \alpha/2) \frac{S_n}{\sqrt{n}} < \mu < \bar{X}_n + t_{n-1}(1 - \alpha/2) \frac{S_n}{\sqrt{n}} \right), \quad \forall \mu \in \mathbb{R}, \sigma^2 \in \mathbb{R}^+, \end{aligned}$$

čímž je (i) dokázáno.

(ii) Postupujeme podle obecného postupu odvození intervalového odhadu a volíme $H(\mathbf{X}; \sigma^2) = \frac{n-1}{\sigma^2} S_n^2$. To má podle věty 5.5 bod (ii) χ_{n-1}^2 rozdělení, které nezávisí na parametrech (μ, σ^2) . Proto

$$\begin{aligned} 1 - \alpha &= P_{\mu, \sigma^2} \left(\chi_{n-1}^2(\alpha/2) < \frac{n-1}{\sigma^2} S_n^2 < \chi_{n-1}^2(1 - \alpha/2) \right) \\ &= P_{\mu, \sigma^2} \left(\frac{(n-1)S_n^2}{\chi_{n-1}^2(1 - \frac{\alpha}{2})} < \sigma^2 < \frac{(n-1)S_n^2}{\chi_{n-1}^2(\frac{\alpha}{2})} \right), \quad \forall \mu \in \mathbb{R}, \sigma^2 \in \mathbb{R}^+, \end{aligned}$$

čímž je (ii) dokázáno. □

Horní a dolní intervalové odhady bychom odvodili obdobně, např. dolní intervalový odhad pro μ v normálním modelu je $\left(\bar{X}_n - t_{n-1}(1 - \alpha) \frac{S_n}{\sqrt{n}}, +\infty \right)$.

Poznámka. Délka intervalu (5.6) je $2t_{n-1}(1 - \alpha/2) \frac{S_n}{\sqrt{n}}$ a je náhodná. I kdyby se náhodou $S_n = \sigma$, tak je větší než $2u_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$, délka intervalového odhadu pro případ známého σ^2 , neboť platí $u_{1-\frac{\alpha}{2}} < t_{n-1}(1 - \alpha/2)$, $\forall 0 < \alpha < 1/2$, $n > 1$. To odpovídá větší nejistotě při nahrazení známého σ^2 jeho odhadem S_n^2 . Ale platí rovněž $t_{n-1}(1 - \alpha/2) \searrow u_{1-\frac{\alpha}{2}}$ pro $n \rightarrow \infty$, a tedy pro zvětšující se rozsah výběru n budou (náhodné) meze intervalového odhadu μ pro případ neznámého σ^2 konvergovat k mezím intervalového odhadu pro případ známého σ^2 .

Příklady použití intervalových odhadů pro konkrétní data budou na cvičení.

Zde ještě doplníme jednu informaci o distribuční funkci a kvantilech χ_n^2 a t_n -rozdělení. Kvantily těchto rozdělení totiž nejsou pro velká n tabelovány, a i statistický software je nahrazuje kvantily normovaného normálního rozdělení. Důvod ukazuje následující věta.

Věta 5.7 Označme H_n distribuční funkci Studentova t -rozdělení o n stupních volnosti a G_n distribuční funkci χ^2 rozdělení o n stupních volnosti. Pak platí

$$\lim_{n \rightarrow \infty} H_n(x) = \Phi(x), \quad \forall x \in \mathbb{R} \quad \text{a} \quad \lim_{n \rightarrow \infty} G_n(n + \sqrt{2n} x) = \Phi(x), \quad \forall x \in \mathbb{R}.$$

Důkaz: Budte $U, Y_1, \dots, Y_i, \dots$ vzájemně nezávislé stejně rozdělené náhodné veličiny s $N(0, 1)$ rozdělením. Pak $V_n = \sum_{i=1}^n Y_i^2$ má χ_n^2 rozdělení s distribuční funkcí G_n a platí $E V_n = n$ a $\text{var } V_n = 2n$. Rovněž platí

$$T_n = \frac{U}{\sqrt{\frac{V_n}{n}}} \sim t_n,$$

s distribuční funkcí H_n . Posloupnost $\{Y_i^2\}_{i=1}^\infty$ splňuje předpoklady Čebyševova slabého zákona velkých čísel a tedy $\frac{V_n}{n} \xrightarrow[n \rightarrow \infty]{P} 1$. Věta o spojitě transformaci pro konvergenci v pravděpodobnosti nám při volbě funkce $g(x) = 1/\sqrt{x}$, která je spojitá na okolí bodu 1, dá $1/\sqrt{\frac{V_n}{n}} \xrightarrow[n \rightarrow \infty]{P} 1$. Pro posloupnost $\{U_n\}_{n=1}^\infty$, kde $U(\omega) = U_n(\omega)$, $\omega \in \Omega$, $n \in \mathbb{N}$ zřejmě platí $U_n \xrightarrow[n \rightarrow \infty]{d} N(0, 1)$ a Cramérova-Sluckého věta pak dává

$$T_n = U_n \frac{1}{\sqrt{\frac{V_n}{n}}} \xrightarrow[n \rightarrow \infty]{d} N(0, 1),$$

což je ekvivalentní rovnosti

$$\lim_{n \rightarrow \infty} H_n(x) = \Phi(x), \quad \forall x \in \mathbb{R},$$

protože $\Phi(x)$ je spojitá všude.

Posloupnost $\{Y_i^2\}_{i=1}^\infty$ splňuje i předpoklady centrální limitní věty (neboť $\text{var } Y_i^2 = 2 \in \mathbb{R}^+$) a z té plyne, že

$$\frac{V_n - n}{\sqrt{2n}} \xrightarrow[n \rightarrow \infty]{d} N(0, 1). \quad (5.7)$$

Přepíšme si nyní postupně, za pomoci definice G_n a faktu, že V_n má χ_n^2 rozdělení

$$\lim_{n \rightarrow \infty} G_n(n + \sqrt{2n} x) = \lim_{n \rightarrow \infty} P\left(\sum_{i=1}^n Y_i^2 \leq n + \sqrt{2n} x\right) = \lim_{n \rightarrow \infty} P\left(\frac{V_n - n}{\sqrt{2n}} \leq x\right) = \Phi(x), \quad \forall x \in \mathbb{R},$$

z (5.7) a toho, že distribuční funkce $\Phi(x)$ je spojitá všude. \square

Důsledek. Platí

$$\lim_{n \rightarrow \infty} t_n(\alpha) = u_\alpha, \quad \forall \alpha \in (0, 1) \quad \text{a} \quad \lim_{n \rightarrow \infty} \frac{\chi_n^2(\alpha) - n}{\sqrt{2n}} = u_\alpha, \quad \forall \alpha \in (0, 1).$$

Důkaz: Si laskavý čtenář udělá sám za pomoci předchozí věty. \square

V praxi se nám také často stává, že spíše než určit intervalový odhad pro střední hodnotu jednoho náhodného výběru, potřebujeme intervalový odhad pro rozdíl středních hodnot dvou náhodných výběrů. Například, když porovnáváme dvě různé

*Zde končí
předn. 22
(15.5.)*

populace (výsledky testů na jedné a druhé škole). Nebo když chceme rozhodnout o účinnosti nějakého léčebného postupu, účinnost určujeme podle hodnot nějaké měřitelné charakteristiky (např. koncentrace čehosi v krvi), a naměříme hodnoty této charakteristiky před a po léčení. Pak bychom rádi věděli, jestli se ta střední hodnota dané charakteristiky v obou výběrech (měření před a po) liší (resp. zvýšila, nebo snížila). Odvodíme si proto ještě intervalový odhad pro rozdíl středních hodnot dvou výběrů z normálního rozdělení.

Mějme tedy (X_1, \dots, X_n) náhodný výběr z $N(\mu_1, \sigma_1^2)$ a (Y_1, \dots, Y_m) na něm nezávislý náhodný výběr z $N(\mu_2, \sigma_2^2)$. Zde $n, m \in \mathbb{N}$ jsou známé, $\mu_1, \mu_2 \in \mathbb{R}$, $\sigma_1^2, \sigma_2^2 \in \mathbb{R}^+$, jsou neznámé parametry. Chceme intervalový odhad $(\mu_1 - \mu_2)$. Podle obecného postupu potřebujeme vhodnou funkci $H(X_1, \dots, X_n, Y_1, \dots, Y_m; \mu_1, \mu_2, \sigma_1^2, \sigma_2^2)$, jejíž rozdělení nezávisí na odhadovaných parametrech. Abychom ji mohli definovat, musíme slevit z obecnosti a přidat si předpoklad stejných rozptylů v obou výběrech, tj.

$$\sigma_1^2 = \sigma_2^2 = \sigma^2 \in \mathbb{R}^+.$$

Zopakujme si, co už víme:

$$\bar{X}_n \sim N\left(\mu_1, \frac{\sigma^2}{n}\right) \quad \bar{Y}_m \sim N\left(\mu_2, \frac{\sigma^2}{m}\right).$$

Protože oba výběry jsou nezávislé, jsou nezávislé i náhodné veličiny \bar{X}_n a \bar{Y}_m a (\bar{X}_n, \bar{Y}_m) má dvojrozměrné normální rozdělení s vektorem středních hodnot $\mu = (\mu_1, \mu_2)^\top$ a varianční maticí $\sigma = (\sigma^2/n, \sigma^2/m) \mathbb{I}_2$. Tedy z důsledku (iv) definice 3.11 máme $(\bar{X}_n - \bar{Y}_m) \sim N(\mu_1 - \mu_2, \sigma^2/n + \sigma^2/m)$ a dále

$$\frac{\bar{X}_n - \bar{Y}_m - (\mu_1 - \mu_2)}{\sigma} \sqrt{\frac{mn}{m+n}} \sim N(0, 1), \quad (5.8)$$

které nezávisí na hodnotě parametrů. Ovšem ještě musíme vyřešit problém neznámého σ^2 . Uděláme to obdobně jako v případě jednoho výběru. Označme

$$S_X^2 = \frac{1}{n-1} \sum_{k=1}^n (X_k - \bar{X}_n)^2 \quad S_Y^2 = \frac{1}{m-1} \sum_{k=1}^m (Y_k - \bar{Y}_m)^2. \quad (5.9)$$

Víme

$$\frac{(n-1)S_X^2}{\sigma^2} \sim \chi_{n-1}^2 \quad \frac{(m-1)S_Y^2}{\sigma^2} \sim \chi_{m-1}^2,$$

a protože jsou to funkce nezávislých náhodných výběrů, jsou také nezávislé. Součet dvou nezávislých χ^2 -rozdělených náhodných veličin je také χ^2 -rozdělený, takže

$$\frac{(n-1)S_X^2 + (m-1)S_Y^2}{\sigma^2} \sim \chi_{m+n-2}^2. \quad (5.10)$$

Náhodné veličiny $\bar{X}_n, \bar{Y}_m, S_X^2, S_Y^2$ jsou vzájemně nezávislé, a tedy z (5.8), (5.10) a definice t -rozdělení dostaneme

$$H(X_1, \dots, X_n, Y_1, \dots, Y_m; \mu_1, \mu_2, \sigma^2) = \frac{\bar{X}_n - \bar{Y}_m - (\mu_1 - \mu_2)}{\sqrt{\frac{(n-1)S_X^2 + (m-1)S_Y^2}{n+m-2}}} \sqrt{\frac{mn}{m+n}} \sim t_{m+n-2}.$$

Nyní už snadno dokážeme větu:

Věta 5.8 Buďte (X_1, \dots, X_n) náhodný výběr z $N(\mu_X, \sigma^2)$ a (Y_1, \dots, Y_m) na něm nezávislý náhodný výběr z $N(\mu_Y, \sigma^2)$. Pak oboustranný intervalový odhad parametrické funkce $(\mu_X - \mu_Y)$ o spolehlivosti $(1 - \alpha)$ má tvar

$$\left(\bar{X}_n - \bar{Y}_m - t_{m+n-2} \left(1 - \frac{\alpha}{2}\right) S^* \sqrt{\frac{m+n}{mn}}, \bar{X}_n - \bar{Y}_m + t_{m+n-2} \left(1 - \frac{\alpha}{2}\right) S^* \sqrt{\frac{m+n}{mn}} \right),$$

kde S_X^2 a S_Y^2 jsou definovány v (5.9) a $S^* = \left(\frac{1}{m+n-2} \left((n-1)S_X^2 + (m-1)S_Y^2 \right) \right)^{\frac{1}{2}}$.

Důkaz: Zřejmý. □

Poznámka. Proč bylo potřeba předpokládat stejné σ^2 pro oba výběry? Abychom našli funkci výběrů \mathbf{X} a \mathbf{Y} , která má χ^2 -rozdělení, a ono neznámé σ se v (5.8) a (5.10) pokrátí tak, že předpis $H(X_1, \dots, X_n, Y_1, \dots, Y_m; \mu_1, \mu_2, \sigma^2)$ už na něm nezávisí.

A proč jsme odvodili intervalový odhad za použití S^* ? Nešlo by místo něj použít např. S_X (to je také konzistentní odhad σ)? Šlo by odvodit intervalový odhad $(\mu_1 - \mu_2)$ se spolehlivostí $(1 - \alpha)$, kde by na místě S^* bylo S_X (bude potřeba upravit ještě něco dalšího)? A byl by tento intervalový odhad lepší nebo horší než ten z věty 5.8? Rozmyslete.

Pro úplnost uvedeme i intervalový odhad pro případ známého σ^2 :

Věta 5.9 Buďte (X_1, \dots, X_n) náhodný výběr z $N(\mu_X, \sigma^2)$ a (Y_1, \dots, Y_m) na něm nezávislý náhodný výběr z $N(\mu_Y, \sigma^2)$, kde $\sigma^2 > 0$ je známé. Pak oboustranný intervalový odhad parametrické funkce $(\mu_X - \mu_Y)$ o spolehlivosti $(1 - \alpha)$ má tvar

$$\left(\bar{X}_n - \bar{Y}_m - u_{1-\frac{\alpha}{2}} \sigma \sqrt{\frac{m+n}{mn}}, \bar{X}_n - \bar{Y}_m + u_{1-\frac{\alpha}{2}} \sigma \sqrt{\frac{m+n}{mn}} \right).$$

Důkaz: Zřejmý. □

Ve větě 5.8 i 5.9 je v předpokladech nezávislost náhodných výběrů \mathbf{X} a \mathbf{Y} . To je předpoklad, který by v našem příkladě s ověřováním účinnosti léčebného postupu nešlo považovat za splněný (ti samí pacienti měření „před“ a „po“ léčebném postupu nebudou dávat nezávislé výsledky). Takže co dělat v případě, kdy náhodné výběry \mathbf{X} a \mathbf{Y} nezávislé nejsou? Pak musíme něco předpokládat o sdruženém rozdělení výběrů \mathbf{X} a \mathbf{Y} . Nejpřirozenější je předpokládat, že $(X_1, Y_1)^T \dots (X_n, Y_n)^T$ je náhodný výběr z dvojrozměrného normálního rozdělení. V našem příkladě pak (X_i, Y_i) jsou měření na i -tém pacientovi před a po léčebném postupu.

Věta 5.10 Buď $(X_1, Y_1)^T \dots (X_n, Y_n)^T$ náhodný výběr z dvojrozměrného normálního rozdělení $N_2\left((\mu_X, \mu_Y)^T, \begin{pmatrix} \sigma_X^2 & \rho\sigma_X\sigma_Y \\ \rho\sigma_X\sigma_Y & \sigma_Y^2 \end{pmatrix}\right)$, kde $\mu_X, \mu_Y \in \mathbb{R}$, $\sigma_X^2, \sigma_Y^2 \in \mathbb{R}^+$ a $\rho \in (-1, 1)$ jsou neznámé parametry. Pak

$$\left(\bar{X}_n - \bar{Y}_n - t_{n-1} \left(1 - \frac{\alpha}{2}\right) \frac{S_{D,n}}{\sqrt{n}}, \bar{X}_n - \bar{Y}_n + t_{n-1} \left(1 - \frac{\alpha}{2}\right) \frac{S_{D,n}}{\sqrt{n}}\right),$$

je intervalový odhad parametrické funkce $(\mu_X - \mu_Y)$ o spolehlivosti $(1 - \alpha)$. Zde $D_i = X_i - Y_i$, $i = 1, \dots, n$, a \bar{D}_n a $S_{D,n}^2$ jsou jejich výběrový průměr a výběrový rozptyl.

Důkaz: Z důsledku (iv) definice 3.11 máme $D_i \sim N(\mu_X - \mu_Y, \sigma_X^2 + \sigma_Y^2 - 2\rho\sigma_X\sigma_Y)$ a $\mathbf{D} = (D_1, \dots, D_n)$ je náhodný výběr z jednorozměrného normálního rozdělení. Stačí tedy dosadit do věty 5.6, bod (i). \square

Poznámka. V čem se liší intervalové odhady z věty 5.8 a věty 5.10 pro případ $m = n$ a $\sigma_X = \sigma_Y$? Rozptyl rozdílu $\bar{X}_n - \bar{Y}_n$ za předpokladů první věty 5.8 bude $\frac{2\sigma^2}{n}$, neboť výběry \mathbf{X} a \mathbf{Y} jsou nezávislé. Za předpokladů druhé věty 5.10 bude $\text{var}(\bar{X}_n - \bar{Y}_n) = \frac{2(1-\rho)\sigma^2}{n}$, tedy pro $\rho > 0$ je menší, a pro ρ blízko 1 značně menší než $\frac{2\sigma^2}{n}$. Tedy variabilita intervalového odhadu z druhé věty bude menší. Toho se využívá při plánování experimentů – v případě, že design experimentu můžeme předem ovlivnit, volíme ho tak, aby byly mezi X_i a Y_i kladné korelace.

V našem příkladě s měřením pacientů „před“ a „po“ léčebném postupu lze očekávat, že $\rho > 0$ skutečně bude. Podle věty 5.8 bychom museli postupovat v případě, že by někdo ztratil označení pacientů u druhého měření, a my bychom tedy nebyli schopni identifikovat, které Y_j patří k i -tému pacientovi a měření X_i . Měli bychom tedy jen dva náhodné výběry z populace pacientů „před“ a „po“ léčebném postupu, o kterých bychom potom předpokládali, že jsou nezávislé (s argumentem, že ono pomíchání = znáhodnění pořadí ve výběru, odstínilo vliv individuálních charakteristik jednotlivých pacientů, které právě implikovalo závislost mezi X_i a Y_i).

5.2.2 INTERVALOVÉ ODHADY ZALOŽENÉ NA CLV

A co můžeme dělat v případě, že \mathbf{X} je náhodný výběr z jiného, než normálního modelu? Můžeme využít speciálních vlastností daného modelu, abychom našli funkci $H(\mathbf{X}; \theta)$ použitelnou pro obecný postup odvození intervalového odhadu (jako jsme to udělali v příkladu na str. 82). Ne vždy je to ale možné. Pokud trochu slevíme z přesnosti, můžeme použít limitní věty z kapitoly 4 k tomu, abychom odvodili *intervalový odhad s asymptotickou spolehlivostí* $(1 - \alpha)$, neboli $(\eta_L(\mathbf{X}), \eta_U(\mathbf{X}))$ pro který platí

$$\lim_{n \rightarrow \infty} P_\theta(\eta_L(\mathbf{X}) < g(\theta) < \eta_U(\mathbf{X})) = 1 - \alpha, \quad \forall \theta \in \Theta.$$

Ukažme si to na intervalovém odhadu pro střední hodnotu $E X_1 = \mu$. Předpokládejme, že (X_1, \dots, X_n) je náhodný výběr z rozdělení s neznámou střední hodnotou μ

a konečným rozptylem $\text{var } X_1 = \sigma^2 \in \mathbb{R}^+$. V tom případě jsou splněny předpoklady CLV a platí $\frac{\bar{X}_n - \mu}{\sigma} \sqrt{n} \xrightarrow[n \rightarrow \infty]{\text{sj}} N(0, 1)$. Tedy

$$\begin{aligned} \lim_{n \rightarrow \infty} P_\mu \left(\bar{X}_n - u_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} < \mu < \bar{X}_n + u_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \right) \\ = \lim_{n \rightarrow \infty} P_\mu \left(\left| \frac{\bar{X}_n - \mu}{\sigma} \sqrt{n} \right| < u_{1-\frac{\alpha}{2}} \right) = \Phi(u_{1-\frac{\alpha}{2}}) - \Phi(u_{\frac{\alpha}{2}}) = 1 - \alpha, \end{aligned}$$

a $\left(\bar{X}_n - u_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \bar{X}_n + u_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \right)$ je intervalový odhad parametru μ s asymptotickou spolehlivostí $(1-\alpha)$. Bohužel ale předpis tohoto intervalu spolehlivosti obsahuje neznámé σ . Musíme tedy ještě nahradit neznámé σ jeho konzistentním odhadem. Univerzálně použitelný je silně konzistentní odhad S_n^2 (za předpokladu $\text{var } X_1 = \sigma^2 \in \mathbb{R}^+$ – viz věta 5.1). S_n^2 je i slabě konzistentní odhad, čili $S_n^2 \xrightarrow[n \rightarrow \infty]{\text{P}} \sigma^2$. Věta 4.6 o spojitě transformaci pro konvergenci v pravděpodobnosti nám pak pro funkci $g(x) = \frac{\sigma}{\sqrt{x}}$, spojitou na okolí bodu σ^2 , dává $\frac{\sigma}{\sqrt{S_n^2}} \xrightarrow[n \rightarrow \infty]{\text{P}} 1$. Použitím Cramérovo-Sluckého věty následně dostaneme

$$\sqrt{n} \frac{\bar{X}_n - \mu}{\sigma} \frac{\sigma}{\sqrt{S_n^2}} = \frac{\bar{X}_n - \mu}{\sqrt{S_n^2}} \sqrt{n} \xrightarrow[n \rightarrow \infty]{\text{d}} N(0, 1).$$

Takže můžeme zformulovat větu:

Věta 5.11 Bud' (X_1, \dots, X_n) náhodný výběr z rozdělení, jež je prvkem modelu $\mathcal{F} = \{P_\theta, \theta \in \Theta\}$, ve kterém platí $\text{var}_\theta X_1 \in \mathbb{R}^+$. Potom intervalovým odhadem střední hodnoty, tj. odhadem parametrické funkce $\mu = E_\theta X_1$, o asymptotické spolehlivosti $(1-\alpha)$, je interval

$$\left(\bar{X}_n - u_{1-\frac{\alpha}{2}} \frac{S_n}{\sqrt{n}}, \bar{X}_n + u_{1-\frac{\alpha}{2}} \frac{S_n}{\sqrt{n}} \right). \quad (5.11)$$

Důkaz: Zřejmý. □

Analogickým postupem bychom odvodili i horní a dolní asymptotické intervalové odhady pro μ .

O distribučních funkcích z modelu \mathcal{F} není třeba předpokládat, že jsou spojitě nebo ryze monotónní. Neboť odvozený interval je založený na asymptotických výsledcích a má také jen asymptotickou spolehlivost $(1-\alpha)$. Stačí tedy, že spojitost a ryzí monotonie platí pro limitní distribuční funkci Φ . Samozřejmě ale, čím víc se bude lišit skutečná distribuční funkce F_θ náhodného výběru od Φ , tím „přibližnější“ bude spolehlivost odvozeného intervalového odhadu. Tedy ve skutečnosti může být o něco menší, než deklarovaná $(1-\alpha)$.

Pokud mám k dispozici lepší (méně variabilní) odhad σ^2 než je S_n^2 , je dobré ho použít.

Příklad. Bud' (X_1, \dots, X_n) náhodný výběr s modelu $\mathcal{F} = \{Pois(\lambda), \lambda \in \mathbb{R}^+\}$. Chceme odvodit intervalový odhad λ . Pro Poissonovo rozdělení platí $E X_1 = \text{var } X_1 = \lambda \in \mathbb{R}^+$. Tedy jsou splněny předpoklady věty 5.11 a (5.11) je intervalový odhad parametru λ o asymptotické spolehlivosti $(1 - \alpha)$.

Ale protože $\text{var } X_1 = \lambda = E X_1$, mohu ho odhadnout konzistentně také výběrovým průměrem \bar{X}_n . Ten má menší variabilitu než S_n^2 a alternativní intervalový odhad

$$\left(\bar{X}_n - u_{1-\frac{\alpha}{2}} \frac{\sqrt{\bar{X}_n}}{\sqrt{n}}, \bar{X}_n + u_{1-\frac{\alpha}{2}} \frac{\sqrt{\bar{X}_n}}{\sqrt{n}} \right) \quad (5.12)$$

o asymptotické spolehlivosti $(1 - \alpha)$ bude stabilnější. (Ověřte, že (5.12) opravdu je intervalový odhad o asymptotické spolehlivosti $(1 - \alpha)$).

Poznámka. Může se stát, že máme k dispozici intervalový odhad $(\eta_L(\mathbf{X}), \eta_U(\mathbf{X}))$ o (ať už asymptotické nebo přesné) spolehlivosti $(1 - \alpha)$ pro parametr $\theta \in \Theta$. Ale potřebovali bychom intervalový odhad pro parametrickou funkci $g(\theta)$. Pokud je funkce g spojitá a ryze monotónní na Θ , pak lze problém snadno vyřešit. Pro g ryze rostoucí máme

$$P_\theta (\eta_L(\mathbf{X}) < \theta < \eta_U(\mathbf{X})) = P_\theta (g(\eta_L(\mathbf{X})) < g(\theta) < g(\eta_U(\mathbf{X}))) .$$

Takže $(g(\eta_L(\mathbf{X})), g(\eta_U(\mathbf{X})))$ je intervalový odhad parametrické funkce $g(\theta)$ o spolehlivosti $(1 - \alpha)$. Pro g ryze klesající se jen otočí nerovnosti a dostaneme tvar intervalového odhadu $(g(\eta_U(\mathbf{X})), g(\eta_L(\mathbf{X})))$.

*Zde končí
předn. 23
(21.5.)*

5.3 TESTOVÁNÍ HYPOTÉZ

V teorii odhadu používáme pozorování, abychom určili rozumný odhad neznámé hodnoty parametru, nebo parametrické funkce. Naproti tomu v testování hypotéz je úkolem vyvinout pravidla pro racionální rozhodování v náhodných situacích, ve kterých potřebujeme jednoznačné rozhodnutí (s potenciálně dalekosáhlými důsledky). Rozhodovací problém je popsán jako hypotéza o skutečném náhodném mechanismu, který řídí (produkuje) pozorování. Pozorování jsou pak použita k rozhodnutí o tom, jestli hypotézu zamítnout nebo nikoli. Protože pozorování jsou náhodná, rozhodnutí může zřejmě být i chybné. Cíl je tedy najít rozhodovací pravidla, která udrží pravděpodobnost chyby tak malou, jak je to jen možné – a to nezávisle na tom, jaká je neznámá skutečnost.

Celý proces testování hypotéz si postupně ukážeme na příkladu z oblasti kontroly kvality.

Příklad. Mějme importéra ovoce, který dostane dodávku, řekněme, $N = 10000$, řekněme, pomerančů. A chce vědět, kolik z nich je špatných/zkažených. Aby to zjistil, odebere vzorek 50 pomerančů. A z nich je náhodný počet x shnilých. Importér musí zaplatit dohodnutou cenu, jen když je maximálně 5% pomerančů v dodávce

zkažených. Takže potřebuje rozhodnout, jestli je kvalita dodávky dostatečná. Chce tedy zvolit číslo c = počet shnilých pomerančů ve vzorku, které bude ještě tolerovat. Pak použije rozhodovací pravidlo:

- nanejvýš c shnilých pomerančů ve vzorku \Rightarrow přijmout dodávku

- více než c shnilých pomerančů ve vzorku \Rightarrow požadovat slevu.

Samozřejmě jde o to, zvolit rozumnou hodnotu c . Ale jak?

Obecný postup testování hypotézy probíhá v několika krocích:

1. Formulace statistického modelu: Tím se musí vždy začít. Tedy $\mathbf{X} = (X_1, \dots, X_n)$ je náhodný výběr z $P_{\theta_0} \in \mathcal{F} = \{P_{\theta} : \theta \in \Theta\}$.

2. Formulace nulové hypotézy a alternativní hypotézy (alternativy): Θ se rozdělí na disjunktní Θ_0 a Θ_1 podle principu:

$\theta \in \Theta_0 \Leftrightarrow \theta$ je uspokojivé, tj. θ je považováno za normální případ/situaci a nevyžaduje žádnou další akci,

$\theta \in \Theta_1 \Leftrightarrow \theta$ je problematické, tj. θ představuje deviaci od normální situace, kterou je potřeba detekovat, kdykoli se objeví.

Říkáme pak, že testujeme *nulovou hypotézu* $H_0 : \theta \in \Theta_0$ proti *alternativní hypotéze (alternativě)* $H_1 : \theta \in \Theta_1$.

Poznámka. Je dobré mluvit o nulové hypotéze, než jen o hypotéze, jinak je nebezpečí zmatení – je to totiž alternativa, která popisuje ten „podezřelý“ případ, který v běžné řeči odpovídá významu slova „hypotéza“.

3. Volba hladiny testu: Při rozhodování mohou nastat chyby dvou druhů:

- chybné zamítnutí platné nulové hypotézy (*chyba I. druhu*)

- chybné nezamítnutí neplatné nulové hypotézy (*chyba II. druhu*)

Chyba I. druhu má závažnější důsledky (a např. u našeho importéra je zahanbující), takže je třeba ji udržet tak malou, jak je to jen možné. Proto se stanoví *hladina testu* α , $0 < \alpha < 1$ (např. 5%), a požaduje se po rozhodovacím pravidlu (testu) (zatím stále ještě neurčeném), aby pravděpodobnost chybného zamítnutí nulové hypotézy (= pravděpodobnost chyby I. druhu) nebyla větší než α .

Uvědomme si, že zde vzniká asymetrie mezi nulovou a alternativní hypotézou.

4. Volba rozhodovacího pravidla / kritického oboru testu $W \subset \mathbb{R}^n$:

- náhodný výběr $\mathbf{X} \in W \Rightarrow$ zamítneme H_0 ,

- náhodný výběr $\mathbf{X} \notin W \Rightarrow$ nezamítneme H_0 .

Pozor, množina W je nenáhodná! Pomocí W a rozhodovací procedury je test určen.

5. Provedu experiment: až teď! V praxi velmi důležité, protože jinak je nezanedbatelná šance, že dojde ke klamání nebo sebeklamu a „statistice“ místo statistiky.

Ukažme si teď aplikaci tohoto postupu na příkladu importéra pomerančů:

Příklad. 1.) Z popisu problému je zřejmé, že vhodný model pro situaci je hypergeometrické rozdělení, které jsme viděli v sekci 2.3. $\mathcal{F} = \{\text{Hypergeom}(N = 10000, M = \theta, n = 50), \theta \in \Theta = \{0, \dots, N\}\}$. Připomeňme si, že hypergeometrické rozdělení popisuje situaci, kdy máme celkem N předmětů – M z nich typu I a $(N - M)$ typu II. Z nich n náhodně vybereme. Hypergeometrické rozdělení popisuje pravděpodobnosti toho, že přesně m z n vybraných předmětů je typu I

$$P(X = m) = \frac{\binom{M}{m} \binom{N-M}{n-m}}{\binom{N}{n}}, \quad \text{pro } 0 \leq m \leq M \text{ a } 0 \leq n - m \leq N - M,$$

$$= 0, \quad \text{jinak.}$$

Předměty typu I jsou shnilé pomeranče. Z $n = 50$ vybraných pozoruje importér $m = x$ shnilých, což je náhodný výběr z hypergeometrického rozdělení o rozsahu 1.

2.) Z hlediska importéra je uspokojivá situace, když je kvalita dostatečná, čili kazivost je maximálně 5%, neboli M je maximálně $N * 0.05 = 500$, tj.

$\theta \in \Theta_0 = \{0, \dots, 500\}$,

Problematická situace je, když je kvalita nedostatečná, neboli $M > 500$, tj.

$\theta \in \Theta_1 = \{501, \dots, 10000\}$.

3.) Importér zvolí hladinu testu α .

4.) Vhodný kritický obor bude mít tvar $W = \{c + 1, c + 2, \dots, n = 50\}$, tedy více než c zkažených pomerančů ve vzorku. A c musí být takové, aby byla zachována hladina testu α z bodu 3. Ovšem na druhou stranu chcí W největší možné (a tedy c nejmenší možné), abych maximalizovala pravděpodobnost zamítnutí H_0 , když H_0 neplatí.

Označme si

$$\sum(M) = \sum_{m=c+1}^{50} \frac{\binom{M}{m} \binom{10000-M}{50-m}}{\binom{10000}{50}},$$

to je pravděpodobnost, že $x \in W$ když $\theta = M$, čili že zamítnu hypotézu H_0 . Když chcí splnit hladinu testu α , tak požadují

$$\sum(M) \leq \alpha \quad \forall M \in \Theta_0 = \{0, \dots, 500\}. \quad (5.13)$$

Stačí mi ovšem najít nejmenší možné c tak, aby platilo $\sum(500) \leq \alpha$. Pro ostatní $M \in \Theta_0$ bude nerovnost (5.13) splněna automaticky, protože $\sum(M)$ je pro pevné c rostoucí funkcí M na Θ .

5.) Nakonec importér provede test.

Uvědomme si ještě, že pro $M \in \Theta_1$, ale blízké hraniční hodnotě 500, bude pravděpodobnost $\sum(M)$ (správného) zamítnutí hypotézy H_0 blízká hodnotě $\sum(500) \approx \alpha$. Na příklad pro $\alpha = 0.05$ odvodíme v bodě 4. $c = 5$, neboli při šesti zkažených pomerančích ve vzorku zamítám hypotézu. Ovšem pro $M = \theta = 600$, kazivost 6%, bude $\sum(600) = 0.077$, tedy pravděpodobnost správného zamítnutí hypotézy bude jen necelých osm procent. Pokud ale $M = \theta = 1000$, kazivost 10%, je $\sum(1000) = 0.38$ a pro kazivost 20% je $\sum(2000) = 0.95$.

Kritický obor W je nenáhodná množina. Náhodný je jev $\{\mathbf{X} \in W\}$. Typicky se tento jev dá ekvivalentně popsat jako $\{T_n(\mathbf{X}) \in C\}$, pro nějakou statistiku $T_n(\mathbf{X})$ a $C \subset \mathbb{R}$. $T_n(\mathbf{X})$ říkáme *testová statistika* a C je typicky interval. Podle bodu 3. z obecného postupu testování hypotéz požadujeme

$$\sup_{\theta \in \Theta_0} P_{\theta}(\mathbf{X} \in W) = \sup_{\theta \in \Theta_0} P_{\theta}(T_n(\mathbf{X}) \in C) \leq \alpha,$$

a zároveň chceme

$$\forall \theta \in \Theta_1 \quad P_{\theta}(\mathbf{X} \in W) = P_{\theta}(T_n(\mathbf{X}) \in C) = \text{maximální}.$$

Zde maximální přes všechny možné volby množiny W zachovávající hladinu testu α . Neboli chceme minimalizovat chybu II druhu, resp. chceme maximalizovat pravděpodobnost odhalení, že H_0 neplatí, když opravdu neplatí.

Funkce

$$\beta(\theta) = P_{\theta}(\mathbf{X} \in W), \quad \theta \in \Theta_1,$$

se nazývá *síla testu proti alternativě* θ . Test s kritickým oborem W^* , který by splňoval hladinu testu α a zároveň

$$P_{\theta}(\mathbf{X} \in W^*) \geq P_{\theta}(\mathbf{X} \in V) \quad \forall \theta \in \Theta_1,$$

pro každý jiný test na hladině α s kritickým oborem V , by byl stejnoměrně nejsilnější test H_0 proti H_1 na hladině α . Takový test bychom chtěli. Bohužel, existuje jen někdy, za dost omezujících předpokladů (např. pro Θ_0 a Θ_1 jednobodové) – viz přednáška Matematická statistika I.

Ale pozor! Ani ten stejnoměrně nejsilnější test nemusí být dost dobrý pro danou situaci. Je třeba najít rovnováhu mezi hladinou testu a silou testu. Čím menší hladina, tím menší je obecně i síla testu. Čili čím víc se snažím vyhnout chybě I druhu, tím menší šance odhalit alternativu, když platí, a tím pravděpodobnější je chyba II druhu. Když hladina a síla testu neumožní udělat dobře podložené rozhodnutí, pak může být jedinou (možná nepohodlnou) možností zvýšit dostupnou informaci a pořídit více nebo lepší pozorování.

A jak tedy ty statistické testy, respektive jejich kritické obory konstruovat? Metod je více. Obecně podobně jako v odvozování intervalových odhadů potřebujeme testovou statistiku $T_n(\mathbf{X})$, jejíž rozdělení za platnosti hypotézy H_0 nezávisí na neznámých charakteristikách rozdělení P_{θ} (a je známo alespoň asymptoticky). A také je potřeba, aby to rozdělení statistiky $T_n(\mathbf{X})$ bylo citlivé na skutečnou hodnotu testovaného parametru θ .

Potom kritický obor C volíme tak, aby byla dodržena hladina testu α , a byly v něm zahrnuty ty hodnoty $T_n(\mathbf{X})$, které jsou za platnosti hypotézy H_0 méně pravděpodobné, než za platnosti alternativy.

A protože máme z předchozí kapitoly dobře rozmyšlené intervalové odhady, tak si zde ukážeme, jak konstruovat kritické obory pomocí intervalových odhadů pro vhodně volenou funkci $g(\theta)$.

*Zde končí
předn. 24
(28.5.)*

Příklad. Chceme porovnat průměrnou výšku chlapců a děvčat. Naměříme proto výšky n náhodně vybraných chlapců (X_1, \dots, X_n) , a m dívek (Y_1, \dots, Y_m) , $n, m \in \mathbb{N}$. Budeme postupovat podle obecného postupu testování hypotéz.

1.) Model: pro změřené biologické charakteristiky je přijatelný normální model, takže budeme předpokládat $X_i \sim N(\mu_X, \sigma^2)$, $Y_j \sim N(\mu_Y, \sigma^2)$. Náhodné výběry \mathbf{X} a \mathbf{Y} předpokládáme nezávislé, a parametry jsou $\mu_X, \mu_Y \in \mathbb{R}$, $\sigma^2 \in \mathbb{R}^+$. Předpoklad stejného rozptylu je přijatelný a (jak si pamatujeme z intervalových odhadů) umožní nám snadněji najít vhodné statistiky se známým rozdělením.

2.) $H_0 : \mu_X = \mu_Y$, neboli průměrná výška chlapců a dívek je stejná.

$H_1 : \mu_X \neq \mu_Y$, neboli průměrná výška chlapců a dívek není stejná.

3.) Např. $\alpha = 0.05$.

4.) Nyní je potřeba určit test = vhodnou statistiku $T(\mathbf{X}, \mathbf{Y})$ a kritický obor C . Je přirozené použít $(\bar{X}_n - \bar{Y}_m)$, když se hodně liší, tak svědčí proti H_0 . Ale rozdělení $(\bar{X}_n - \bar{Y}_m)$ závisí na neznámém σ^2 (i za hypotézy!). Je tedy potřeba $(\bar{X}_n - \bar{Y}_m)$ upravit tak, aby na neznámém σ^2 nezáviselo. Umíme to? – Ano, víme totiž z odvození před větou 5.8, že

$$\frac{\bar{X}_n - \bar{Y}_m - (\mu_X - \mu_Y)}{\sqrt{\frac{(n-1)S_X^2 + (m-1)S_Y^2}{n+m-2}}} \sqrt{\frac{mn}{m+n}} \sim t_{m+n-2},$$

ale za platnosti hypotézy H_0 je $\mu_X - \mu_Y = 0$. Tedy za platnosti H_0 platí

$$T(\mathbf{X}, \mathbf{Y}) = \frac{\bar{X}_n - \bar{Y}_m}{\sqrt{\frac{(n-1)S_X^2 + (m-1)S_Y^2}{n+m-2}}} \sqrt{\frac{mn}{m+n}} \sim t_{m+n-2}.$$

Jako vhodný kritický obor lze volit

$$C = \left(-\infty, -t_{m+n-2} \left(1 - \frac{\alpha}{2}\right)\right] \cup \left[t_{m+n-2} \left(1 - \frac{\alpha}{2}\right), +\infty\right), \quad (5.14)$$

neboť ten obsahuje málo pravděpodobné hodnoty statistiky $T(\mathbf{X}, \mathbf{Y})$, a za platnosti nulové hypotézy H_0 máme $P_{\mu_X, \mu_Y}(T(\mathbf{X}, \mathbf{Y}) \in C) = \alpha$.

Takto jsme odvodili kritický obor tak říkajíc „ručně“. Ale je možné tentýž kritický obor odvodit i z vhodného intervalového odhadu. Naše nulová i alternativní hypotéza mluví o rozdílu středních hodnot v normálním modelu. Použijeme tedy intervalový odhad pro rozdíl středních hodnot v normálním modelu. Věta 5.8 říká, že platí

$$P_{\mu_X, \mu_Y} \left((\mu_X - \mu_Y) \in \left(\bar{X}_n - \bar{Y}_m - t_{1-\alpha/2, m+n-2} S^* \sqrt{\frac{m+n}{mn}}, \bar{X}_n - \bar{Y}_m + t_{1-\alpha/2, m+n-2} S^* \sqrt{\frac{m+n}{mn}} \right) \right) = 1 - \alpha.$$

Za platnosti hypotézy H_0 je $\mu_X - \mu_Y = 0$, tedy postupně dostáváme

$$P_{H_0} \left(\bar{X}_n - \bar{Y}_m - t_{1-\alpha/2, m+n-2} S^* \sqrt{\frac{m+n}{mn}} < 0 < \bar{X}_n - \bar{Y}_m + t_{1-\alpha/2, m+n-2} S^* \sqrt{\frac{m+n}{mn}} \right) = 1 - \alpha$$

$$P_{H_0} \left(|\bar{X}_n - \bar{Y}_m| < t_{m+n-2} \left(1 - \frac{\alpha}{2}\right) S^* \sqrt{\frac{m+n}{mn}} \right) = 1 - \alpha$$

$$P_{H_0} \left(T(\mathbf{X}, \mathbf{Y}) \in \left(-t_{m+n-2} \left(1 - \frac{\alpha}{2}\right), t_{m+n-2} \left(1 - \frac{\alpha}{2}\right) \right) \right) = 1 - \alpha$$

a ta množina, kam spadá $T(\mathbf{X}, \mathbf{Y})$ v poslední rovnosti, je doplněk kritického oboru C .

Rozhodovací pravidlo testu tedy vypadá:

$T(\mathbf{X}, \mathbf{Y}) \in C \Rightarrow$ zamítáme H_0

$T(\mathbf{X}, \mathbf{Y}) \notin C \Rightarrow$ nezamítáme H_0 ,

kde C je definováno v (5.14). Ekvivalentně vyjádřeno pomocí kritického oboru přímo pro \mathbf{X} a \mathbf{Y}

$$W = \left\{ (\mathbf{x}, \mathbf{y}) : |\bar{x}_n - \bar{y}_m| \geq t_{m+n-2} \left(1 - \frac{\alpha}{2}\right) S^* \sqrt{\frac{m+n}{mn}} \right\},$$

a rozhodovací pravidlo je

$(\mathbf{X}, \mathbf{Y}) \in W \Rightarrow$ zamítáme H_0

$(\mathbf{X}, \mathbf{Y}) \notin W \Rightarrow$ nezamítáme H_0 .

Test odvozený v příkladu se nazývá *dvouvýběrový t-test*, a v aplikacích se vyskytuje velmi často. Tak často, jak častá je potřeba statistického porovnání středních hodnot dvou různých populací.

„Překlopení“ intervalového odhadu do kritického oboru testu, jak bylo ukázáno v příkladě, lze ovšem provést i obecně:

Mějme intervalový odhad $(\eta_L(\mathbf{X}), \eta_U(\mathbf{X}))$ pro parametrickou funkci $g(\theta)$ s pravděpodobností pokrytí $(1 - \alpha)$ (přesnou nebo asymptotickou). Pak test nulové hypotézy $H_0 : g(\theta) = g(\theta_0)$ proti alternativě $H_1 : g(\theta) \neq g(\theta_0)$, určený rozhodovacím pravidlem:

$g(\theta_0) \notin (\eta_L(\mathbf{X}), \eta_U(\mathbf{X})) \Rightarrow$ zamítáme H_0

$g(\theta_0) \in (\eta_L(\mathbf{X}), \eta_U(\mathbf{X})) \Rightarrow$ nezamítáme H_0

má hladinu α (přesně nebo asymptoticky).

Poznámka. Samozřejmě lze i obráceně z testů odvodit intervalové odhady.

Pro případ jednostranné alternativy – např. $H_0 : g(\theta) = g(\theta_0)$ proti $H_1 : g(\theta) > g(\theta_0)$: pokud máme dolní intervalový odhad $(\eta_D(\mathbf{X}), \infty)$ pro parametr θ s pravděpodobností pokrytí $(1 - \alpha)$, pak test daný rozhodovacím pravidlem:

$g(\theta_0) \leq \eta_D(\mathbf{X}) \Rightarrow$ zamítáme H_0

$g(\theta_0) \in (\eta_D(\mathbf{X}), \infty) \Rightarrow$ nezamítáme H_0

je testem výše popsané H_0 proti H_1 na hladině α .

Pro alternativu $H_1 : g(\theta) < g(\theta_0)$ bychom postupovali analogicky.

Vraťme se ještě k našemu příkladu.

Příklad. Nyní provedeme krok 5. – provedeme test na datech.

Řekněme, že $n = 25$, $m = 20$ a protože nepotřebujeme znát celá data, ale jen výběrové průměry a výběrové rozptyly obou výběrů, stačí nám že $\bar{X}_n = 180.6$, $\bar{Y}_m = 164.9$, $S_X^2 = 6.5$, $S_Y^2 = 9.3$. Po dosazení zjistíme, že oboustranný intervalový odhad s pravděpodobností pokrytí $(1 - 0.05)$ z věty 5.8 je $(14.0; 17.4)$ a $0 \notin (14.0; 17.4)$, takže na hladině 0.05 zamítáme hypotézu, že chlapci a dívky jsou stejně vysocí.

Představme si, že teď chceme na stejných datech testovat, jestli jsou chlapci o více než 10cm vyšší než dívky (a nebo ne). Tedy naše alternativní hypotéza je, že chlapci

jsou o více než 10cm vyšší než dívky. Neboli

$$H_0 : \mu_X - \mu_Y = 10 \times H_1 : \mu_X - \mu_Y > 10$$

Kritický obor je podle odstavce před příkladem

$$W = \left\{ (\mathbf{x}, \mathbf{y}) : \bar{x}_n - \bar{y}_m - t_{1-\alpha, m+n-2} S^* \sqrt{\frac{m+n}{mn}} \geq 10 \right\},$$

a pro naše data $\bar{X}_n - \bar{Y}_m - t_{1-\alpha, m+n-2} S^* \sqrt{\frac{m+n}{mn}} = 14.3$. Tedy $14.3 \geq 10$ a zamítáme i nulovou hypotézu, že chlapci jsou o maximálně 10cm vyšší než dívky, ve prospěch alternativní hypotézy, že chlapci jsou o více než 10cm vyšší než dívky.

V přednášce ještě proběhla diskuse průběhu silofunkce pro alternativy $H_1 : \mu_X \neq \mu_Y$ a $H_1 : \mu_X > \mu_Y$. Ten se velmi liší, neboť heuristicky řečeno, pro případ jednostranné alternativy $H_1 : \mu_X > \mu_Y$ jsou „podezřelé“ jen odchylky od 0 nahoru.

*Zde končí
předn. 25
(29.5.)*

OPRAVY

Opravy a úpravy, které byly provedeny oproti první verzi textu postupně přidávaných přednášek.

Příklady na lineární transformaci v kapitole 2.4 Definice Y měla přehozené koeficienty. Opraveno na správné $Y = aX + b$.

Věta 3.1 Důkaz doplněn a rozepsán podrobněji.

Věta 3.4 Tvzení o $P_X((-\infty, x])$ bylo silnější, než je potřeba. Nyní odpovídá tomu, co je dokázáno.

Poznámka a příklad před větou 3.11 Byla tisková chyba v textu, opraveno na správné $\text{cor}(X, -X) = -1$.

Věta 3.15 Byly překlapy v oborech integrování ve znění věty a v označení distribuční funkce náhodné veličiny U v důkazu. Rovněž na začátku důkazu má být integrál přes \mathbb{R}^2 , nikoli dvojný integrál.

Věta 3.20 Ve znění věty ze vzorce pro podíl vypadlo x . Správně je

$$f_V(v) = \int_0^\infty f_X(vx)f_Y(x)x \, dx.$$

V posledních dvou vzorcích v důkazu věty byly špatně indexy.

Kapitola 3.5.2 - úvod Ve druhém vysazeném vzorci chybělo 2 v $\frac{1}{\sqrt{2\pi\sigma^2}}$.

Důkaz věty 4.2 Na konec důkazu doplněna poznámka o tom, že obě implikace jsou zároveň i ekvivalence.

poznámka za větou 4.6 Doplněna poznámka o tom, co přesně říká věta 4.6 v situaci, kdy je limitní náhodná veličina Y konstantní.

Důkaz věty 4.8 důkaz konvergence řady na str. 56 – v rovnostech (4.7) byla v druhém členu pro hodnotu $n = 1$ vnitřní suma efektivně prázdná. Aby nedošlo k matení v následujících úpravách, byl posunut rozsah vnější sumy v druhém členu až od $n = 2$.

poznámka na konci str. 56 Doplněna jedna konkrétní volba $\text{var } X_n$.

Důkaz CLV, strana 62 dole Doplněny ω do předposledního vzorce na stránce. $\zeta(\omega)$ je různé pro různá ω , ale vždy $\zeta(\omega) \in [0, 1]$.

Příklad na str. 65 Ve vzorci (4.15) vypadlo 1–, také byly opraveny numerické chyby ve výsledcích.

Důkaz věty 4.17 Vypadl předpoklad $x > 0$ před omezováním množiny $\{\omega : X_n(\omega) \cdot Y_n(\omega) \leq x\}$. Nyní doplněn a doplněno vysvětlení, co je jinak pro $x < 0$.

Odvození intervalového odhadu na str. 89 Asi uprostřed strany se tvrdí $(\bar{X}_n - \bar{Y}_m) \sim N(\mu_1 - \mu_2, \sigma^2/n + \sigma^2/m)$. Byl překlep ve vyjádření rozptylu.

Věta 5.8 a 5.9 V intervalovém odhadu byl převrácený zlomek popisující závislost na m, n . Opraveno na správné $\sqrt{\frac{m+n}{mn}}$.

odvození věty 5.11 První vysazený vzorec na straně – v předposledním výrazu vypadly kvantily $u_{1-\frac{\alpha}{2}}$ a $u_{\frac{\alpha}{2}}$. Konvergence uprostřed strany 92 – byl přepis v druhu konvergence, správně je $\xrightarrow[n \rightarrow \infty]{d}$.

LITERATURA

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

Georgii, H.O. (2013). *Stochastics*. Berlin: de Gruyter.

Hogg, R.V. a Tanis, E.A. a Zimmerman, D.L. (2015). *Probability and Statistical Inference*. Pearson Education.

Lachout, P. (2004). *Teorie pravděpodobnosti*. Praha: Karolinum.