NMSA202 Pravděpodobnost a matematická statistika

Poznámky k přednášce

Naposledy upraveno dne 27. června 2023.

Katedra pravděpodobnosti a matematické statistiky
Matematicko-fyzikální fakulta University Karlovy

Tento učební text obsahuje přehled všech definic, tvrzení, vět a jejich důkazů, i poznámek probíraných v přednášce „NMSA202 Pravděpodobnost a matematická statistika“ v rámci bakalářského studia oboru „Obecná matematika“ na MFF UK. Text je rozšířený i o některé příklady a vysvětlení proč, co, a jak děláme, je tedy poněkud „upovídanější“ než klasická skripta.

Nejedná se o plnohodnotnou učebnici, protože zde chybí některé příklady a není zde obsažená látka probíraná na cvičení. Při přípravě na zkoušku je nutné si tento text doplnit poznámkami z přednášek a cvičení

Obsah

Značení

$\emptyset$	prázdná množina
$\mathbb{R}$	množina reálných čísel
$\mathbb{N}$	množina přirozených čísel
$\mathbb{Z}$	množina celých čísel
$\boldsymbol{a}^\mathsf{T}$	transpozice vektoru $\boldsymbol{a}$
$\lVert\boldsymbol{a}\rVert$	eukleidovská norma vektoru $\boldsymbol{a}$
$\mathbb{I}_n$	jednotková matice $n \times n$
Ω	prostor elementárních jevů, stavový prostor
$\mathcal{P}(\Omega)$	potenční množina množiny Ω
$\mathcal{A}$	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
$\mu \ll \nu$	míra μ je absolutně spojitá vzhledem k míře ν
$\mu \perp \nu$	míry μ a ν jsou vzájemně singulární
$\mu\Psi^{-1}$	obraz míry μ v zobrazení Ψ
P	pravděpodobnost
P_X	rozdělení náhodné veličiny X
S_X	nosič rozdělení náhodné veličiny X
$\mathsf{P}_\mathbf{X}$	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 $(\Omega, \mathcal{A}, \mathsf{P})$ s konečným p-tým absolutním momentem
$\lVert h\rVert_\infty$	supremová norma funkce h
$\mathsf{E}\,X$	střední hodnota náhodné veličiny X
$\mathsf{var}\,X$	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

$\mathrm{cov}\,(X_1, X_2)$	kovariance náhodných veličin X_1 a X_2
$\mathrm{cor}\,(X, Y)$	korelační koeficient náhodných veličin X a Y
$\mathrm{Var}\,(\mathbf{X})$	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
$\overline{X}_n$	výběrový průměr náhodného výběru $X_1, \ldots, X_n$
S_n^2	výběrový rozptyl náhodného výběru $X_1, \ldots, X_n$
$\widehat{m}_r$	r-tý výběrový moment náhodného výběru $X_1, \ldots, X_n$
$\mathcal{F}$	pravděpodobnostní model pro pozorovaná data
$\widehat{\theta}_n$	odhad parametru θ na základě náhodného výběru $X_1, \ldots, X_n$
$L(\theta, \mathbf{x})$	věrohodnost θ pro pozorování $\mathbf{x}$
$l(\theta, \mathbf{x})$	logaritmická věrohodnost θ pro pozorování $\mathbf{x}$
$\xrightarrow[n\to\infty]{\mathsf{P}}$	konvergence v pravděpodobnosti
$\xrightarrow[n\to\infty]{\mathsf{sj}}$	konvergence skoro jistě
$\xrightarrow[n\to\infty]{\mathsf{d}}$	konvergence v distribuci

1 Pravděpodobnostní prostor

1.1 Axiomatická definice pravděpodobnosti

První objekt, který si musíme definovat, pokud chceme matematicky popisovat náhodu, je pravděpodobnostní prostor.

Definice 1.1 (pravděpodobnostní prostor) Buď Ω neprázdná množina a $\mathcal{A}$ σ-algebra na množině Ω. *Pravděpodobnost* P je množinová funkce $\mathsf{P} : \mathcal{A} \to [0, 1]$ splňující

(i) $\mathsf{P}(A) \geq 0, \forall A \in \mathcal{A}$;

(ii) $\mathsf{P}(\Omega) = 1$;

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

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

Množina Ω se nazývá *stavový prostor*. Obsahuje všechny možné „realizace náhody“ – to jest všechny možné výsledky popisovaného náhodného experimentu, respektive všechny možnosti, co se může stát v popisované náhodné situaci. Prvky množiny Ω nazýváme *elementární jevy* a samotná Ω se také někdy nazývá *prostor elementárních jevů*.

σ-algebra $\mathcal{A}$ se nazývá *σ-algebra náhodných jevů* a její prvky se nazývají *(náhodné) jevy*. σ-algebra náhodných jevů je vlastně systém podmnožin stavového prostoru Ω, kterým umíme konzistentně přiřadit pravděpodobnost.

Pokud se snažíme vytvořit pravděpodobnostní model pro nějakou náhodnou situaci – to jest vymyslet pravděpodobnostní prostor, který by mohl jako tento model sloužit – je volba Ω a $\mathcal{A}$ často snadná/zřejmá. S volbou vhodné pravděpodobnosti P už to může být složitější. Pravděpodobnost P obsahuje všechnu potřebnou informaci o dané náhodné situaci. To, jak ji zvolíme, určí, jak dobrý/použitelný bude náš model pro popis dané náhodné situace.

Poznámka. Ty tři podmínky z definice 1.1 – nezápornost, normovanost a σ-aditivita – jsou přesně ty tři axiomy, jejichž splnění požadujeme po objektu, který chceme nazývat pravděpodobností. Zcela intuitivně by pravděpodobnost tyto vlastnosti mít měla.

Poznámka. Při použití znalostí z přednášky *Teorie míry a integrálu (TMI)* zjistíme, že pravděpodobnost definovaná v definici 1.1 je vlastně konečná normovaná míra (normovaná tak, aby míra celého prostoru byla 1). V následujícím budeme znalosti z přednášky *TMI* využívat a nebudeme opakovat důkazy v ní probrané. Některé z výsledků si ale explicitně připomene.

Připomeňme si, že $\mathcal{A}$ je σ-algebra na množině Ω, pokud $\mathcal{A} \subset \mathcal{P}(\Omega)$ je neprázdný systém uzavřený na doplňky a spočetná sjednocení. Důsledkem definice je, že $\mathcal{A}$ obsahuje prázdnou množinu $\emptyset$, celé Ω, množinový rozdíl $B \backslash A$, pro $A, B \in \mathcal{A}$, a konečná i spočetná sjednocení i průniky množin z $\mathcal{A}$. Všem těmto množinám tedy umíme konzistentně přiřadit pravděpodobnost. V teorii pravděpodobnosti ovšem prvky $\mathcal{A}$ představují náhodné jevy a my je tak budeme v dalším nazývat.

Terminologie. Pokud je $\omega \in \Omega$ přesně ten elementární jev, který nastal, a $A \in \mathcal{A}$ taková, že $\omega \in A$, pak říkáme, že *nastal jev* A. Pokud $\omega \notin A$, pak říkáme, že *jev* A *nenastal*.
Buď i $B \in \mathcal{A}$. Pokud $\omega \in A \cap B$, pak *nastaly oba jevy* A i B. Pokud $\omega \in A \cup B$, pak *nastal alespoň jeden z jevů* A, B.
Pokud platí $A \subset B$, pak A je *podjev* jevu B. $\emptyset$ se nazývá *jev nemožný*, Ω *jev jistý*.
$A^c = \Omega \backslash A$ se nazývá *jev opačný* nebo *doplňkový* k jevu A.
Pokud platí $A \cap B = \emptyset$, pak jevy A, B nazveme *vzájemně neslučitelné*.

Příklad (Klasický pravděpodobnostní prostor)**.** Buď Ω konečná, $|\Omega| = n \in \mathbb{N}$, $\mathcal{A} = \mathcal{P}(\Omega)$, a $\mathsf{P}(\omega) = \frac{1}{n}$. Pak nutně

$$\mathsf{P}(A) = \frac{|A|}{|\Omega|} = \frac{|A|}{n}, \quad A \in \mathcal{A}.$$

Tento pravděpodobnostní model se hodí třeba k popisu hodu spravedlivou kostkou, pak $n = 6$.

Příklad (Diskrétní pravděpodobnostní prostor)**.** Buď Ω konečná nebo spočetná, $\mathcal{A} = \mathcal{P}(\Omega)$, a mějme funkci $p : \Omega \to [0,1]$ splňující $p(\omega) \geq 0$ a $\sum_{\omega\in\Omega} p(\omega) = 1$. Pak nutně

$$\mathsf{P}(A) = \sum_{\omega\in A} p(\omega), \quad A \in \mathcal{A}.$$

Pro popis jednoho hodu (ne nutně spravedlivou) mincí se hodí diskrétní pravděpodobnostní prostor s $\Omega = \{0,1\}$. $p(1)$ bude pravděpodobnost, že padla panna, $p(0) = 1 - p(1)$ pravděpodobnost, že padl orel.
Pokud budeme hod nezávisle n-krát opakovat, pak bude vhodný popis pomocí diskrétního pravděpodobnostního prostoru s $\Omega = \{0,1\}^n$. Odpovídající $p(\omega)$ si čtenář odvodí sám jako cvičení.
Kdyby nás ale zajímalo jen to, kolikrát padla v n pokusech panna, pak lze ten samý experiment popsat i pomocí diskrétního pravděpodobnostního prostoru s $\Omega = \{0,1,\ldots,n\}$ a vhodně určenými $p(\omega)$.
Vhodný pravděpodobnostní prostor po popis nějaké náhodné situace nemusí být jednoznačně určený, různé (ovšem smysluplné) volby pravděpodobnostního prostoru mohou mít různé výhody, vždy by ale měly vést na stejné odpovědi.

Příklad (Spojitý pravděpodobnostní prostor)**.** Buď $\Omega = \mathbb{R}$, $\mathcal{A} = \mathcal{B}(\mathbb{R})$ borelovská σ-algebra na $\mathbb{R}$, a mějme měřitelnou funkci $g : \mathbb{R} \to [0, \infty)$ takovou, že $\int_{\mathbb{R}} g(x)\,\mathrm{d}x = 1$. Pak

$$P(A) = \int_A g(x)\,\mathrm{d}x, \qquad A \in \mathcal{A},$$

je dobře definovaná pravděpodobnost.
Pokud volíme speciálně $g(x) = \mathbb{1}_{[0,1]}(x)$, pak je P Lebesgueova míra omezená na interval $[0, 1]$.

Příklad (Geometrická pravděpodobnost)**.** Buď $\Omega \in \mathcal{B}^n$ taková, že $0 < \lambda(\Omega) < \infty$. Položme $\mathcal{A} = \mathcal{B}(\Omega)$ a

$$P(A) = \frac{\lambda(A)}{\lambda(\Omega)}, \qquad A \in \mathcal{A}.$$

Pak P je dobře definovaná pravděpodobnost a takovýto pravděpodobnostní prostor může modelovat náhodnou volbu bodu v množině, ale i náhodnou volbu složitějších geometrických objektů (úseček, přímek, kompaktních množin, ...). Více o geometrické pravděpodobnosti lze najít např. v Dupač a Hušková (2013) v sekci 1.2.

Připomeňme si některé vlastnosti pravděpodobnosti, které se nám budou hodit při počítání.

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

(i) $\mathsf{P}(\emptyset) = 0$;

(ii) P je konečně aditivní;

(iii) $\mathsf{P}(A^c) = 1 - \mathsf{P}(A)$;

(iv) $\mathsf{P}(A \cup B) = \mathsf{P}(A) + \mathsf{P}(B) - \mathsf{P}(A \cap B)$;

(v) P je monotónní, tj. $A \subset B \;\Rightarrow\; \mathsf{P}(A) \leq \mathsf{P}(B)$;

(vi) je-li $A_1 \subseteq A_2 \subseteq \dots$, pak $\lim_{n\to\infty} \mathsf{P}(A_n) = \mathsf{P}\left(\bigcup_{i=1}^{\infty} A_i\right)$;

(vii) je-li $A_1 \supseteq A_2 \supseteq \dots$, pak $\lim_{n\to\infty} \mathsf{P}(A_n) = \mathsf{P}\left(\bigcap_{i=1}^{\infty} A_i\right)$;

(viii) $B \subset A \;\Rightarrow\; \mathsf{P}(A \backslash B) = \mathsf{P}(A) - \mathsf{P}(B)$;

(ix) princip inkluze a exkluze, tj.

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

Důkaz: Důkazy bodů (i)–(v) a (viii) jsou snadné a známé z přednášky *TMI*. Důkazy bodů (vi) a (vii) pomocí triku zdisjunktnění jsou také známé z *TMI*. Důkaz bodu (ix) je z nám z přednášky *Diskrétní matematika*. □

Poznámka. Speciální případ bodu (vii) z předchozí věty je tzv. „spojitost pravděpodobnosti v $\emptyset$" tj. tvrzení

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

Zde končí přednn. 1 (15.2.)

1.2 Podmíněná pravděpodobnost

Když se zabýváme pravděpodobnostmi různých jevů, velmi často narazíme na otázku, jak spolu pravděpodobnosti různých jevů souvisí. Přesněji řečeno, jak výskyt jednoho jevu ovlivňuje pravděpodobnost výskytu jevu druhého. Nebo ještě jinak, když víme, že nastal jev A, jaká je potom pravděpodobnost jevu B, a jestli se liší od pravděpodobnosti jevu B v případě, že bychom o výskytu jevu A nic nevěděli. Ke zkoumání takových otázek je užitečná následující definice:

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

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

Že je podmíněná pravděpodobnost opravdu pravděpodobnost, ukazuje následující věta:

Věta 1.2 Buď $(\Omega, \mathcal{A}, \mathsf{P})$ pravděpodobnostní prostor a $B \in \mathcal{A}$ takové, že $\mathsf{P}(B) > 0$. Pak zobrazení $\mathsf{P}(\cdot|B) : \mathcal{A} \to [0, 1]$ splňuje definici pravděpodobnosti 1.1.

Důkaz: Stačí ověřit splnění podmínek (i) – (iii):
Podle definice je $\mathsf{P}(A|B)$ podíl nezáporného a kladného čísla, a tedy je nutně ≥ 0, a (i) je splněno.
Dosazením do definice a snadnou úpravou máme

$$\mathsf{P}(\Omega|B) = \frac{\mathsf{P}(\Omega \cap B)}{\mathsf{P}(B)} = \frac{\mathsf{P}(B)}{\mathsf{P}(B)} = 1,$$

tedy i (ii) je splněno.

Mějme $A_1, A_2, \dots \in \mathcal{A}$ po dvou disjunktní, pak

$$
\begin{aligned}
\mathsf{P}\left(\bigcup_{i=1}^{\infty} A_i \middle| B\right) &= \frac{\mathsf{P}\left(\left(\bigcup_{i=1}^{\infty} A_i\right) \cap B\right)}{\mathsf{P}(B)} = \frac{\mathsf{P}\left(\bigcup_{i=1}^{\infty} (A_i \cap B)\right)}{\mathsf{P}(B)} \\
&= \frac{\sum_{i=1}^{\infty} \mathsf{P}(A_i \cap B)}{\mathsf{P}(B)} = \sum_{i=1}^{\infty} \frac{\mathsf{P}(A_i \cap B)}{\mathsf{P}(B)} = \sum_{i=1}^{\infty} \mathsf{P}(A_i|B),
\end{aligned}
$$

kde ve třetí rovnosti jsme využili faktu, že i množiny $\{(A_i \cap B)\}_{i=1}^{\infty}$ jsou po dvou disjunktní. Tedy i (iii) je splněno. □

Pro podmíněnou pravděpodobnost platí následující pozorování.

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

(i) $\mathsf{P}(A \cup C|B) = \mathsf{P}(A|B) + \mathsf{P}(C|B) - \mathsf{P}(A \cap C|B)$;

(ii) $B \subset A \;\Rightarrow\; \mathsf{P}(A|B) = 1$;

(iii) $A \cap B = \emptyset \;\Rightarrow\; \mathsf{P}(A|B) = 0$;

(iv) $\mathsf{P}(A|\Omega) = \mathsf{P}(A)$;

(v) pokud $\mathsf{P}(\{\omega\}) > 0$ pro nějaké $\omega \in \Omega$, pak $\forall A \in \mathcal{A}$ platí $\mathsf{P}(A|\{\omega\}) \in \{0, 1\}$.

Důkaz: Snadno dosazením do definice podmíněné pravděpodobnosti. Ponecháváme čtenáři jako cvičení. □

Poznámka. Je velmi důležité si uvědomit, že podmíněná pravděpodobnost není míra v druhé proměnné, tj. obecně neplatí rovnost mezi $\mathsf{P}(A|B \cup C)$ a $\mathsf{P}(A|B) + \mathsf{P}(A|C)$, a to ani v případě, kdy jsou B a C disjunktní! Zkuste si nějaký protipříklad najít.

Podmíněná pravděpodobnost je velmi účinný nástroj pro výpočet pravděpodobností složitějších jevů (jak uvidíte i na cvičení). Často se v takových výpočtech používá některá z následujících tří vět.

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

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

Důkaz: Z předpokladu $\mathsf{P}(A_1 \cap \dots \cap A_{n-1}) > 0$ plyne $\mathsf{P}(A_1 \cap \dots \cap A_{n-k}) > 0$ pro každé $k = 1, 2, \dots, n-1$ z monotonie pravděpodobnosti. Tedy všechny podmíněné pravděpodobnosti na pravé straně rovnosti jsou dobře definované.

Důkaz provedeme indukcí. Pro $n = 2$ platí

$$\mathsf{P}(A_1 \cap A_2) = \mathsf{P}(A_2|A_1)\mathsf{P}(A_1)$$

z definice podmíněné pravděpodobnosti. Pro $n > 2$ platí

$$\begin{aligned}\mathsf{P}(A_1 \cap \dots \cap A_n) &= \mathsf{P}(A_n|A_1 \cap \dots \cap A_{n-1})\mathsf{P}(A_1 \cap \dots \cap A_{n-1}) \\ &= \mathsf{P}(A_n|A_1 \cap \dots \cap A_{n-1})\mathsf{P}(A_{n-1}|A_1 \cap \dots \cap A_{n-2}) \dots \mathsf{P}(A_2|A_1)\mathsf{P}(A_1),\end{aligned}$$

kde jsme využili nejdříve definici podmíněné pravděpodobnosti a potom indukční předpoklad pro $n-1$. □

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

$$\mathsf{P}(A) = \sum_n \mathsf{P}(A|B_n)\mathsf{P}(B_n).$$

Důkaz: Z předpokladu $\mathsf{P}(\bigcup_n B_n) = 1$ plyne, že pravděpodobnost opačného jevu $\mathsf{P}((\bigcup_n B_n)^c) = 0$, takže

$$\begin{aligned}\mathsf{P}(A) &= \mathsf{P}\left(A \cap \left(\bigcup_n B_n\right)\right) + \mathsf{P}\left(A \cap \left(\bigcup_n B_n\right)^c\right) = \mathsf{P}\left(\bigcup_n (A \cap B_n)\right) \\ &= \sum_n \mathsf{P}(A \cap B_n) = \sum_n \mathsf{P}(A|B_n)\mathsf{P}(B_n).\end{aligned}$$

V první rovnosti jsme použili konečnou aditivitu míry, ve třetí rovnosti, že jevy $\{(A \cap B_n)\}$ jsou po dvou disjunktní, ve čtvrté rovnosti jsme použili definici podmíněné pravděpodobnosti. □

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

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

Důkaz: Použijeme postupně definici podmíněné pravděpodobnosti a větu 1.5. □

Příklad (Pólyovo urnové schéma)**.** Podrobnosti viz přednáška.

Zde končí předn. 2 (16.2.)

1.3 Nezávislost náhodných jevů

V předchozí kapitole jsme zkoumali podmíněné pravděpodobnosti. Významná je situace, kdy je podmíněná i nepodmíněná pravděpodobnost stejná $\mathsf{P}(A) = \mathsf{P}(A|B)$. Ta nastává, pokud jsou jevy A a B nezávislé.

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

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 těch dvou jevů.

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

Důkaz: Počítejme a postupně použijme konečnou aditivitu pravděpodobnosti, nezávislost jevů A a B a pravděpodobnost opačného jevu:

$$\mathsf{P}(A \cap B^c) = \mathsf{P}(A) - \mathsf{P}(A \cap B) = \mathsf{P}(A) - \mathsf{P}(A)\mathsf{P}(B) = \mathsf{P}(A)(1 - \mathsf{P}(B)) = \mathsf{P}(A)\mathsf{P}(B^c).$$

Tedy i A a B^c jsou nezávislé. Použitím definice podmíněné pravděpodobnosti a nezávislosti A a B dostaneme

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

□

A co když máme víc jevů než dva? Jak je to s nezávislostí?

Příklad. Mějme klasický pravděpodobnostní prostor s $\Omega = \{1, 2, 3, 4\}$ a jevy $A = \{1, 2\}$, $B = \{2, 3\}$, $C = \{1, 3\}$. Platí

$$\begin{aligned} \mathsf{P}(A) &= \mathsf{P}(B) = \mathsf{P}(C) = \frac{1}{2} \\ \mathsf{P}(A \cap B) &= \frac{1}{4} = \mathsf{P}(A)\mathsf{P}(B) \\ \mathsf{P}(A \cap C) &= \frac{1}{4} = \mathsf{P}(A)\mathsf{P}(C) \\ \mathsf{P}(B \cap C) &= \frac{1}{4} = \mathsf{P}(B)\mathsf{P}(C), \end{aligned}$$

takže jevy A, B, C jsou po dvou nezávislé. Ale

$$\mathsf{P}(A \cap B \cap C) = 0 \quad \neq \quad \frac{1}{8} = \mathsf{P}(A)\mathsf{P}(B)\mathsf{P}(C),$$

tedy pro všechny tři jevy už součinový tvar neplatí. Pokud chceme, aby součinový tvar platil pro všechny kombinace jevů, potřebujeme silnější pojem než nezávislost po dvou.

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

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

I pro vzájemnou nezávislost jevů platí, že se přenáší na doplňky.

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

Důkaz: Buďte $A_1, \dots, A_n, B_1, \dots, B_m \in C$. Chceme dokázat

$$\mathsf{P}(A_1 \cap A_2 \cap \dots \cap A_n \cap B_1^c \cap \dots \cap B_m^c) = \mathsf{P}(A_1) \dots \mathsf{P}(A_n)\mathsf{P}(B_1^c) \dots \mathsf{P}(B_m^c),$$

pro každé (přípustné) $n, m \in \mathbb{N}$.

Budeme dokazovat indukcí podle m (pro libovolné n). Pro $m = 1$ rovnost platí z věty 1.7 s volbou $A = A_1 \cap \dots \cap A_n$.
Pro indukční krok $m \to m + 1$ počítejme

$$\begin{aligned}
\mathsf{P}(A_1 \cap \dots \cap A_n \quad \cap \quad & B_1^c \cap \dots \cap B_{m+1}^c) \\
= \quad & \mathsf{P}(A_1 \cap \dots \cap A_n \cap B_1^c \cap \dots \cap B_m^c) - \mathsf{P}(A_1 \cap \dots \cap A_n \cap B_1^c \cap \dots \cap B_{m+1}) \\
= \quad & \mathsf{P}(A_1) \dots \mathsf{P}(A_n)\mathsf{P}(B_1^c) \dots \mathsf{P}(B_m^c)(1 - \mathsf{P}(B_{m+1})) \\
= \quad & \mathsf{P}(A_1) \dots \mathsf{P}(A_n)\mathsf{P}(B_1^c) \dots \mathsf{P}(B_m^c)\mathsf{P}(B_{m+1}^c).
\end{aligned}$$

V první rovnosti jsme použili konečnou aditivitu pravděpodobnosti a v druhé dvakrát indukční předpoklad pro m (neboť oba průniky obsahují jen m doplňkových množin). □

Věta 1.9 Jsou-li jevy $A_1, \dots A_n, B_1, \dots B_m$, $n, m \in \mathbb{N}$ vzájemně nezávislé a $\mathsf{P}(B_1 \cap \dots \cap B_m) > 0$, pak

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

Důkaz: Použijeme postupně definici podmíněné pravděpodobnosti a vzájemnou nezávislost jevů. □

2 Náhodné veličiny

Ne vždy, když modelujeme nějakou náhodnou situaci nebo náhodný experiment, potřebujeme opravdu znát/přesně určit celý složitý pravděpodobnostní prostor $(\Omega, \mathcal{A}, \mathsf{P})$. Někdy nás zajímá jen nějaký aspekt náhodné situace nebo jen některé charakteristiky výsledku náhodného experimentu. Pak je dobré umět měnit „pozorovací hloubku“ neboli umět přecházet mezi jemným/přesnějším stavovým prostorem Ω a hrubším/méně přesným stavovým prostorem řekněme Ω'. K tomu lze použít pojem náhodného elementu, resp. náhodné veličiny.

2.1 Náhodná veličina a její rozdělení

Definice 2.1 Buď $(\Omega, \mathcal{A}, \mathsf{P})$ pravděpodobnostní prostor a $(\Omega', \mathcal{A}')$ měřitelný prostor. Pak každé měřitelné zobrazení $X : \Omega \to \Omega'$ nazveme *náhodný element* z Ω'.

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

Tedy náhodná veličina je <u>funkce</u>.

Značení. Buď $A \in \mathcal{B}$. 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 $\mathsf{P}(X \in A)$.

Definice 2.3 Buď 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 $\sigma(X)$.

Poznámky.

- 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* $X : (\Omega, \mathcal{A}) \to (\mathbb{R}, \mathcal{B})$ nazveme indukovanou pravděpodobnostní míru P_X na $(\mathbb{R}, \mathcal{B})$ definovanou jako $\mathsf{P}_X(B) = \mathsf{P}(\{\omega \in \Omega : X(\omega) \in B\})$, $B \in \mathcal{B}$.

Poznámka. Takže $\mathsf{P}_X = \mathsf{P}X^{-1}$ je obraz míry P v zobrazení X a pravděpodobnostní prostor $(\Omega, \mathcal{A}, \mathsf{P})$ se zobrazí na pravděpodobnostní prostor $(\mathbb{R}, \mathcal{B}, \mathsf{P}_X)$.

Z teorie míry máme také větu, která nám umožňuje přecházet mezi integrováním vzhledem k pravděpodobnosti P a integrováním vzhledem k jejímu obrazu P_X. Je to věta o přenosu integrace.

Věta 2.1 (o přenosu integrace) Buď $(M, \mathcal{A}, \mu)$ prostor s mírou, $g : (M, \mathcal{A}) \to (N, \mathcal{F})$ měřitelné zobrazení a $\nu = \mu g^{-1}$ obraz míry μ v zobrazení g. Buď $h : (N, \mathcal{F}) \to (\mathbb{R}, \mathcal{B})$ měřitelná funkce. Pak platí

$$\int_N h(y)\,\mathrm{d}\nu(y) = \int_M h(g(x))\,\mathrm{d}\mu(x)$$

kdykoli má jedna strana smysl.

Důkaz: Standardním postupem teorie míry – snadné cvičení. □

Dosadíme-li do věty o přenosu integrace následujícím způsobem:

$$(M, \mathcal{A}, \mu) = (\Omega, \mathcal{A}, \mathsf{P}) \qquad (N, \mathcal{F}) = (\mathbb{R}, \mathcal{B}) \qquad g = X \qquad \nu = \mathsf{P}_X$$

dostaneme speciálně větu o přenosu integrace pro rozdělení P_X náhodné veličiny X.

Věta 2.2 (o přenosu integrace pro P_X) Buď X náhodná veličina a buď $h : (\mathbb{R}, \mathcal{B}) \to (\mathbb{R}, \mathcal{B})$ měřitelná funkce. Pak platí

$$\int_\Omega h(X(\omega))\,\mathrm{d}\mathsf{P}(\omega) = \int_\mathbb{R} h(x)\,\mathrm{d}\mathsf{P}_X(x),$$

pokud má alespoň jedna strana smysl.

Poznámka. Takže pro práci s náhodnou veličinou nepotřebujeme znát celé $(\Omega, \mathcal{A}, \mathsf{P})$ 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 z přednášky *TMI* známe Radonovu-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 $\mathsf{P}_X \ll \mu$. Potom $f_X = \frac{\mathrm{d}\mathsf{P}_X}{\mathrm{d}\mu}$ se nazývá *hustota náhodné veličiny X vzhledem k míře μ.*

Poznámky.

- f_X je určena jednoznačně μ-skoro všude.
- Pokud je měřitelná $g : (\mathbb{R}, \mathcal{B}) \to (\mathbb{R}, \mathcal{B})$ nezáporná, nebo pro ni platí $\int_\mathbb{R} |g(x)|\,\mathrm{d}\mathsf{P}_X(x) < \infty$, pak $\int_\mathbb{R} g(x)\,\mathrm{d}\mathsf{P}_X(x) = \int_\mathbb{R} g(x) f_X(x)\,\mathrm{d}\mu(x)$ (důsledek tvrzení 12.1 ze skript Rataj (2022)).

Zde končí přednáška 3 (22.2.)

Věta 2.3 Buď X náhodná veličina, P_X její rozdělení a $B \in \mathcal{B}$. Pak platí následující rovnosti

$$\begin{aligned}\mathsf{P}(X \in B) &= \mathsf{P}(\{\omega \in \Omega : X(\omega) \in B\}) \\ &= \int_\Omega \mathbb{1}_B(X(\omega))\,\mathrm{d}\mathsf{P}(\omega) = \int_\mathbb{R} \mathbb{1}_B(x)\,\mathrm{d}\mathsf{P}_X(x) = \int_B 1\,\mathrm{d}\mathsf{P}_X(x) = \mathsf{P}_X(B).\end{aligned}$$

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

$$\mathsf{P}(X \in B) = \int_\Omega \mathbb{1}_B(X(\omega))\,\mathrm{d}\mathsf{P}(\omega) = \int_\mathbb{R} \mathbb{1}_B(x) f_X(x)\,\mathrm{d}\mu(x) = \int_B f_X(x)\,\mathrm{d}\mu(x).$$

Důkaz: Věta je důsledkem věty o přenosu integrace 2.2, Radonovy-Nikodýmovy věty a tvrzení 12.1 ze skript Rataj (2022). □

Takže při znalosti P_X nebo f_X jsme schopni určit pravděpodobnosti všech jevů ze $\sigma(X)$. Věta 2.3 nám dává návod, jak počítat pravděpodobnosti typu $\mathsf{P}(X \in B)$ různými způsoby, podle toho, jakou informaci o náhodné veličině X máme k dispozici.

Příklad. Buď $\Omega = \{1, 2, 3, 4, 5, 6\}$, $\mathcal{A} = \mathcal{P}(\Omega)$, $\mathsf{P}(\omega) = \frac{1}{6}$, $\omega \in \Omega$. A buď náhodná veličina X dána předpisem $X(\omega) = \mathbb{1}_{\{2,4,6\}}(\omega)$, $\omega \in \Omega$. Tento pravděpodobnostní prostor může popisovat například jeden hod spravedlivou kostkou a X je indikátor toho, že na kostce padlo sudé číslo.
Určete $\sigma(X)$, P_X a f_X. Spočítejte pravděpodobnosti $\mathsf{P}(X = 0)$, $\mathsf{P}(X > 0)$, $\mathsf{P}(X \geq 0)$, pomocí P a X, pomocí rozdělení P_X a pomocí hustoty f_X.
Řešení – viz přednáška.

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

Poznámka. Zřejmě $F_X(x) = \mathsf{P}_X((-\infty, x])$, $x \in \mathbb{R}$. Množinový systém $\{(-\infty, x]), x \in \mathbb{R}\}$ je uzavřený na průniky, pro $A_n = (-\infty, n]$, $n \in \mathbb{N}$ platí $A_n \nearrow \mathbb{R}$, a P_X je konečná míra, takže z věty o jednoznačnosti míry plyne, že distribuční funkce jednoznačně určuje rozdělení P_X náhodné veličiny X. Ale různé náhodné veličiny (ať už definované na různých pravděpodobnostních prostorech, nebo definované na stejném pravděpodobnostním prostoru, ale s různými hodnotami pro stejná $\omega \in \Omega$) mohou mít stejnou distribuční funkci (a tedy stejné rozdělení).

Příklad. (pokračování příkladu o hodu kostkou) Zjistili jsme, že rozdělení náhodné veličiny X je $\mathsf{P}_X = \frac{1}{2}(\delta_0 + \delta_1)$, tedy odpovídající distribuční funkce je

$$F_X(x) = \frac{1}{2}\Big(\mathbb{1}_{[0,\infty)}(x) + \mathbb{1}_{[1,\infty)}(x)\Big).$$

Uvažujme ještě náhodnou veličinu $Y(\omega) = \mathbb{1}_{1,2,3}(\omega)$, $\omega \in \Omega$, tedy indikátor toho, že na kostce padl menší počet ok než 4. Náhodné veličiny X a Y jsou různé (jako funkce z Ω do $\mathbb{R}$), ale když určíme P_Y, f_Y (hustotu vzhledem k čítací míře na $\mathbb{Z}$) a F_Y, zjistíme, že jsou stejné jako pro náhodnou veličinu X.

Věta 2.4 (Vlastnosti distribuční funkce) Buď F_X distribuční funkce náhodné veličiny X. Pak

(i) F_X je neklesající;

(ii) F_X je zprava spojitá;

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

Důkaz: Snadno plyne z monotonie, spojitosti a normovanosti pravděpodobnostní míry. Podrobný důkaz ponecháváme čtenáři jako cvičení. □

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

$$\mu((-\infty, a]) = F(a), \qquad \forall a \in \mathbb{R}. \tag{2.1}$$

Tato míra se nazývá Lebesgueovou-Stieltjesovou mírou příslušnou 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 2.5 Buď $F : \mathbb{R} \to \mathbb{R}$ funkce splňující body (i) – (iii) z věty 2.4. Pak existuje pravděpodobnostní prostor $(\Omega, \mathcal{A}, \mathsf{P})$ a náhodná veličina $X : (\Omega, \mathcal{A}) \to (\mathbb{R}, \mathcal{B})$ taková, že $F = F_X$.

Důkaz: Existenci míry μ splňující (2.1) máme z předchozí poznámky. Stačí tedy nadefinovat kompatibilní pravděpodobnostní prostor $(\Omega, \mathcal{A}, \mathsf{P})$ a funkci X z Ω do $\mathbb{R}$. Můžeme volit $\Omega = \mathbb{R}$, $\mathcal{A} = \mathcal{B}$, $\mathsf{P} = \mu$, a funkci $X(\omega) = \omega$, $\omega \in \Omega$. Funkce X je identita, a jako taková je měřitelná, a tedy je to náhodná veličina. Navíc pro $a \in \mathbb{R}$ platí

$$F_X(a) = \mathsf{P}(X \le a) = \mathsf{P}(\{\omega \in \Omega : X(\omega) \le a\}) = \mu(\{x \in \mathbb{R} : x \le a\}) = \mu((-\infty, a]) = F(a),$$

takže distribuční funkce F_X náhodné veličiny X je přesně rovna F. □

Poznámka. Mějme distribuční funkci F_X nějaké náhodné veličiny X. Je dobré si uvědomit, že z konečné aditivity pravděpodobnostní míry P lze snadno odvodit rovnost

$$\mathsf{P}(X \in (a, b]) = F_X(b) - F_X(a), \qquad a < b \in \mathbb{R}.$$

Ovšem při vyjádření pravděpodobností $\mathsf{P}(X \in (a, b))$ a $\mathsf{P}(X \in [a, b])$ nesmíme zapomenout na možnou nespojitost distribuční funkce F_X zleva a naopak musíme využít spojitost pravděpodobnostní míry P:

$$\begin{aligned}
\mathsf{P}(X \in (a, b)) &= \mathsf{P}(\{\omega \in \Omega : X(\omega) \in (a, b)\}) = \mathsf{P}\left(\left\{\omega \in \Omega : X(\omega) \in \bigcup_{n\in\mathbb{N}} \left(a, b - \frac{1}{n}\right]\right\}\right) \\
&= \mathsf{P}\left(\bigcup_{n\in\mathbb{N}} \left\{\omega \in \Omega : X(\omega) \in \left(a, b - \frac{1}{n}\right]\right\}\right) = \lim_{n\to\infty} \mathsf{P}(X \in (a, b - 1/n]) \\
&= \lim_{n\to\infty} (F_X(b - 1/n) - F_X(a)) = \lim_{t\to b-} F_X(t) - F_X(a).
\end{aligned}$$

Obdobně odvodíme

$$\mathsf{P}(X \in [a,b]) = F_X(b) - \lim_{t\to a-} F_X(t), \qquad a < b \in \mathbb{R}.$$

Poznámka. Pro náhodnou veličinu X (a jí příslušné rozdělení P_X a distribuční funkci F_X) platí

$$\mathsf{P}(X \in B) = \int_{\mathbb{R}} \mathbb{1}_B(x)\,\mathrm{d}\mathsf{P}_X(x) = \int_{\mathbb{R}} \mathbb{1}_B(x)\,\mathrm{d}F_X(x),$$

kde poslední výraz je Lebesgueův-Stieltjesův integrál podle F_X. Takže i z distribuční funkce F_X umíme spočítat pravděpodobnosti všech jevů ze $\sigma(X)$.

Poznámka. A jak vlastně může vypadat P_X? Z teorie míry víme, že každou Lebesgueovu-Stieltjesovu míru μ můžeme rozložit na tři části:

$$\mu = \mu_a + \mu_d + \mu_{sc}, \tag{2.2}$$

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

$$\mathsf{P}_X = \sum_{i\in I} p_i\,\delta_{x_i},$$

kde δ_u značí Diracovu míru v bodě u. Odpovídající distribuční funkce tvaru

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

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 $\mathsf{P}_X \ll \nu$ 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 (23.2.)

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

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

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

Příklad. Příklad, kdy je k modelování náhodného množství potřeba náhodná veličina, která má jak diskrétní, tak i absolutně spojitou část, je modelování množství srážek (spadlých např. za den). S kladnou pravděpodobností se může stát, že neprší, takže naprší 0 — to bude ta diskrétní část. A množství, které naprší, když prší, je výhodné modelovat spojitě. Podrobnosti viz přednáška.

Příklad. V teorii míry jste viděli Cantorovu funkci F_C, která je distribuční funkcí Cantorovy míry μ_C. Neboť platí $\mu_C(\mathbb{R}) = 1$, je μ_C míra pravděpodobnostní, a tedy je to i rozdělení nějaké náhodné veličiny. Ovšem μ_C má v rozkladu (2.2) diskrétní i absolutně spojitou část nulovou. μ_C je spojitá singulární míra, to jest je neatomická a singulární vzhledem k λ. Distribuční funkce F_C je spojitá, ale $F_C' = 0$ λ-skoro všude na $\mathbb{R}$.

Definice 2.9 Buď F_X distribuční funkce náhodné veličiny X. Funkce

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

se nazývá *kvantilová funkce* náhodné veličiny X.

Poznámka. Kvantilová funkce je neklesající a zleva spojitá (čtenář si snadno dokáže jako cvičení). 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

Rozdělení P_X náhodné veličiny X nám dává kompletní informaci o pravděpodobnostech jevů s touto náhodnou veličinou souvisejících. Často ale potřebujeme i nějaké zjednodušené charakteristiky, ideálně vyjádřitelné jedním číslem. Dvě základní charakteristiky, které bychom chtěli definovat, jsou „typická hodnota“ náhodné veličiny a „variabilita“ náhodné veličiny, resp. „rozptýlenost“ (náhodných) hodnot náhodné veličiny kolem její typické hodnoty.

Nejpřirozenější je definovat typickou hodnotu náhodné veličiny jako její průměrnou hodnotu — tak, jak to dělá následující definice.

Definice 2.10 *Střední hodnota* náhodné veličiny X je číslo $\mathsf{E}\,X$ dané výrazem

$$\mathsf{E}\,X = \int_\Omega X(\omega)\,\mathrm{d}\mathsf{P}(\omega),$$

pokud má tento integrál smysl.

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

Poznámka. Jak počítat $\mathsf{E}\,X$? Ne nutně z definice — nepotřebuji znát celou pravděpodobnost P, stačí mi P_X, resp. F_X, neboť

$$\mathsf{E}\,X = \int_\Omega X(\omega)\,\mathrm{d}\mathsf{P}(\omega) = \int_\mathbb{R} x\,\mathrm{d}\mathsf{P}_X(x) = \int_\mathbb{R} x\,\mathrm{d}F_X(x),$$

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

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

a pro absolutně spojitou náhodnou veličinu dostanu

$$\mathsf{E}\,X = \int_{\mathbb{R}} x\, f_X(x)\,\mathrm{d}x.$$

Tedy $\mathsf{E}\,X$ 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.

Příklad. Modelování výše platů a „typického/průměrného“ platu. Podrobnosti viz přednáška.

Jiná možnost jak definovat „typickou“ hodnotu náhodné veličiny, je pomocí mediánu.

Definice 2.11 *Medián* rozdělení náhodné veličiny X je číslo $q_{\frac{1}{2}}$ splňující $\mathsf{P}\left(X \le q_{\frac{1}{2}}\right) \ge \frac{1}{2}$ a $\mathsf{P}\left(X \ge q_{\frac{1}{2}}\right) \ge \frac{1}{2}$.

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.

A jak spočítáme střední hodnotu funkce g nějaké náhodné veličiny X? Je potřeba znát celé rozdělení $g(X)$? Nikoli.

Věta 2.6 Buď X náhodná veličina a $g : \mathbb{R} \to \mathbb{R}$ měřitelná funkce. Pak $g(X)$ je také náhodná veličina a platí

$$\mathsf{E}\,g(X) = \int_{\mathbb{R}} g(x)\,\mathrm{d}\mathsf{P}_X(x) = \int_{\mathbb{R}} g(x)\,\mathrm{d}F_X(x),$$

pokud jeden z nich existuje. Speciálně pro diskrétní náhodnou veličinu

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

a pro absolutně spojitou náhodnou veličinu

$$\mathsf{E}\,g(X) = \int_{\mathbb{R}} g(x) f_X(x)\,\mathrm{d}x.$$

Důkaz: Složení dvou měřitelných zobrazení je opět měřitelné, takže $g(X) : (\Omega, \mathcal{A}) \to (\mathbb{R}, \mathcal{B})$ je náhodná veličina. Platí

$$\mathsf{E}\,g(X) = \int_{\Omega} g(X(\omega))\,\mathrm{d}\mathsf{P}(\omega) = \int_{\mathbb{R}} g(x)\,\mathrm{d}\mathsf{P}_X(x),$$

kde jsme nejdřív použili definici střední hodnoty a pak větu 2.2 o přenosu integrace. $\square$

Terminologie. Řekneme, že něco platí P–*skoro jistě*, pokud to platí pro všechna $\omega \in \Omega \setminus N$, kde $N \in \mathcal{A}$ je množina pravděpodobnosti 0 (tj. $\mathsf{P}(N) = 0$). Pokud je jasné, jakou pravděpodobnostní míru P používáme, říkáme jen: platí *skoro jistě*. Zkráceně píšeme P-s.j., resp. jen s.j.

Přímým důsledkem vlastností Lebesgueova integrálu jsou následující vlastnosti střední hodnoty.

Věta 2.7 (Vlastnosti střední hodnoty) Buďte X, Y náhodné veličiny a $a, b \in \mathbb{R}$. Pak platí:

(i) $\mathsf{E}\,(a + bX) = a + b\mathsf{E}\,X, \quad$ pro $X \in L^1$,

(ii) $\mathsf{E}\,(X + Y) = \mathsf{E}\,X + \mathsf{E}\,Y, \quad$ pro $X, Y \in L^1$,

(iii) $\mathsf{P}(X \geq 0) = 1 \Rightarrow \mathsf{E}\,X \geq 0$,

(iv) $X \in L^1 \Rightarrow |X| \in L^1$,

(v) $X \leq Y$ P–s.j. $\Rightarrow$ $\mathsf{E}\,X \leq \mathsf{E}\,Y$ (pokud existují).

Mimo střední hodnoty $\mathsf{E}\,X$ se často používají i vyšší momenty náhodné veličiny.

Definice 2.12 Buď $n \in \mathbb{N}$. Pak
n-tý moment náhodné veličiny X definujeme jako $\mathsf{E}\,X^n$;
n-tý absolutní moment náhodné veličiny X definujeme jako $\mathsf{E}\,|X|^n$;
n-tý centrální moment náhodné veličiny X definujeme jako $\mathsf{E}\,(X - \mathsf{E}\,X)^n$, existuje-li $\mathsf{E}\,X$;
n-tý absolutní centrální moment náhodné veličiny X definujeme jako $\mathsf{E}\,|X - \mathsf{E}\,X|^n$, existuje-li $\mathsf{E}\,X$.

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

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

Definice 2.13 *Rozptyl* náhodné veličiny X je definován jako $\mathsf{E}\,(X - \mathsf{E}\,X)^2$. Značíme ho $\operatorname{var} X$.

Poznámka. Rozptyl je střední čtvercová odchylka od střední hodnoty $\mathsf{E}\,X$ — měřeno/váženo pomocí P_X. Rozptyl tedy popisuje, jak moc jsou hodnoty náhodné veličiny X rozptýlené kolem $\mathsf{E}\,X$.

Poznámka. Zřejmě je $\operatorname{var} X = \mathsf{E}\,(X - \mathsf{E}\,X)^2 \geq 0$ a z vlastností Lebesgueova integrálu plyne, že $\operatorname{var} X = \mathsf{E}\,(X - \mathsf{E}\,X)^2 = 0$, právě když $X = \mathsf{E}\,X$ P-s.j.

Zde končí přednášk. 5 (1.3.)

Věta 2.8 Buď X náhodná veličina s konečným rozptylem a $a, b \in \mathbb{R}$. Platí

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

Důkaz: Za použití linearity integrálu postupně počítáme:

$$\begin{aligned}\operatorname{var}(a+bX) &= \mathsf{E}\,(a+bX-\mathsf{E}\,(a+bX))^2 = \mathsf{E}\,(a+bX-a-b\mathsf{E}\,X)^2\\ &= \mathsf{E}\,(b(X-\mathsf{E}\,X))^2 = b^2\,\mathsf{E}\,(X-\mathsf{E}\,X)^2 = b^2\operatorname{var}X.\end{aligned}$$

□

Poznámka. Jak počítat rozptyl? Buď přímo:

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

Nebo se může hodit přepis

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

Rozptyl a vyšší momenty náhodné veličiny je možno použít i pro odvození horní meze na pravděpodobnosti velkých odchylek náhodné veličiny X od její střední hodnoty $\mathsf{E}\,X$, resp. od nuly. Následující dvě věty ukazují jak.

Věta 2.9 (Čebyševova nerovnost) Buď X náhodná veličina z L^1 a $a>0$. Pak platí

$$\mathsf{P}(|X-\mathsf{E}\,X|\ge a)\le\frac{\operatorname{var}X}{a^2}.$$

Důkaz: Označme množinu, na níž je splněna podmínka

$$B=\{x\in\mathbb{R}:|x-\mathsf{E}\,X|\ge a\}=\left\{x\in\mathbb{R}:\frac{|x-\mathsf{E}\,X|^2}{a^2}\ge 1\right\}.$$

Pak můžeme psát

$$\begin{aligned}\mathsf{P}(|X-\mathsf{E}\,X|\ge a) &= \int_B \mathrm{d}P_X(x)=\int_B 1\,\mathrm{d}P_X(x)\le\int_B\frac{|x-\mathsf{E}\,X|^2}{a^2}\,\mathrm{d}P_X(x)\\ &\le \int_{\mathbb{R}}\frac{|x-\mathsf{E}\,X|^2}{a^2}\,\mathrm{d}P_X(x)=\frac{\mathsf{E}\,(X-\mathsf{E}\,X)^2}{a^2}=\frac{\operatorname{var}X}{a^2}.\end{aligned}$$

U druhé nerovnosti jsme využili, že $\frac{|x-\mathsf{E}\,X|^2}{a^2}\ge 0$ na celém $\mathbb{R}$, tedy i $\int_{\mathbb{R}\setminus B}\frac{|x-\mathsf{E}\,X|^2}{a^2}\ge 0$. □

Analogicky je možno dokázat i Markovovu nerovnost:

Věta 2.10 (Markovova nerovnost) Buď X náhodná veličina z L^n, $n\in\mathbb{N}$ a $a>0$. Pak platí

$$\mathsf{P}(|X|\ge a)\le\frac{\mathsf{E}\,|X|^n}{a^n}.$$

Poznámka. Oba výše uvedené odhady jsou velmi hrubé.

V dalším výkladu se nám bude hodit i následující nerovnost mezi momenty různých řádů.

Věta 2.11 (Nerovnost mezi L^p normami na pravděpodobnostních prostorech) Buď X náhodná veličina, $0 < \alpha < \beta$ a $\mathsf{E}\,|X|^\beta < \infty$. Pak

$$\left(\mathsf{E}\,|X|^\alpha\right)^{\frac{1}{\alpha}} \leq \left(\mathsf{E}\,|X|^\beta\right)^{\frac{1}{\beta}}, \tag{2.3}$$

a speciálně $\mathsf{E}\,|X| \leq \sqrt{\mathsf{E}\,X^2}$.

Důkaz: Důkaz provedeme dosazením do Hölderovy nerovnosti. Potřebujeme tedy měřitelné funkce f, g z $\mathbb{R}$, $f \in L^p(\mathsf{P}_X)$, $g \in L^q(\mathsf{P}_X)$, kde $1 \leq p, q \leq \infty$ a $\frac{1}{p} + \frac{1}{q} = 1$. Volme $f = |x|^\alpha$, $g = 1$, $p = \frac{\beta}{\alpha}$, $q = \frac{\beta}{\beta-\alpha}$.

Funkce $g = 1$ zřejmě patří do každého $L^c(\mathsf{P}_X)$, $c > 0$, neboť $\mathsf{P}_X(\mathbb{R}) = 1$. Dále $\int_{\mathbb{R}} (f(x))^{\frac{\beta}{\alpha}}\, \mathrm{d}\mathsf{P}_X(x) = \int_{\mathbb{R}} |x|^\beta\, \mathrm{d}\mathsf{P}_X(x) = \mathsf{E}\,|X|^\beta < \infty$ z předpokladu věty. Předpoklady Hölderovy nerovnosti jsou tedy splněny. Můžeme psát

$$\begin{aligned}
\mathsf{E}\,|X|^\alpha &= \int_{\mathbb{R}} |x|^\alpha\, \mathrm{d}\mathsf{P}_X(x) = \int_{\mathbb{R}} |x|^\alpha \cdot 1\, \mathrm{d}\mathsf{P}_X(x) \\
&\leq \left(\int_{\mathbb{R}} |x|^\beta\, \mathrm{d}\mathsf{P}_X(x)\right)^{\frac{\alpha}{\beta}} \cdot \left(\int_{\mathbb{R}} 1^q\, \mathrm{d}\mathsf{P}_X(x)\right)^{\frac{1}{q}} = \left(\mathsf{E}\,|X|^\beta\right)^{\frac{\alpha}{\beta}} \cdot 1 = \left(\mathsf{E}\,|X|^\beta\right)^{\frac{\alpha}{\beta}},
\end{aligned}$$

což je nerovnost ekvivalentní s nerovností (2.3). □

Poznámka. Takže jsme dokázali, že pro pravděpodobnostní míru P platí $L^\beta(\Omega, \mathcal{A}, \mathsf{P}) \subseteq L^\alpha(\Omega, \mathcal{A}, \mathsf{P})$, $\forall\ 0 < \alpha < \beta$. Neboli, je-li absolutní moment řádu n konečný, tak je konečný i pro každé m splňující $0 < m < n$.

2.3 Některá rozdělení náhodné veličiny

V této kapitole si ukážeme několik základních a často používaných rozdělení náhodné veličiny.

2.3.1 Alternativní (Bernoulliho) rozdělení

Náhodná veličina X s *alternativním rozdělením* nabývá pouze hodnot 1 a 0 s pravděpodobnostmi p a $1-p$, kde $p \in (0,1)$ se nazývá *parametr alternativního rozdělení*. Takže

$$\mathsf{P}_X = p\delta_1 + (1-p)\delta_0.$$

Pro momenty alternativního rozdělení platí

$$\mathsf{E}\,X^k = p,\ \ k \in \mathbb{N}, \qquad\qquad \operatorname{var} X = p\,(1-p).$$

Alternativní rozdělení se používá k modelování výsledku jednoho náhodného pokusu, u kterého mohou nastat jen dvě varianty — často označované jako úspěch resp. neúspěch. Zkráceně značíme alternativní rozdělení jako *Alt(p)*.

2.3.2 Binomické rozdělení

Náhodná veličina X má *binomické rozdělení s parametry* p, n, kde $p \in (0,1)$, $n \in \mathbb{N}$, pokud platí

$$\mathsf{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \qquad k \in \{0, 1, \dots n\}.$$

Použijeme-li binomickou větu, vidíme, že $\sum_{k=0}^{n} \mathsf{P}(X = k) = 1$, tedy výše uvedené pravděpodobnosti opravdu odpovídají pravděpodobnostnímu rozdělení. Zkráceně ho budeme označovat *Binom(n,p)*. Pro první dva momenty binomického rozdělení platí

$$\mathsf{E}\,X = np, \qquad \mathsf{var}\,X = np(1-p).$$

Binomické rozdělení je odvozené od alternativního — pokud opakujeme n-krát nezávisle pokus s pravděpodobností úspěchu p, má celkový počet pozorovaných úspěchů binomické rozdělení s parametry n, p.

2.3.3 Geometrické rozdělení

Náhodná veličina X má *geometrické rozdělení s parametrem* p, kde $p \in (0,1)$, pokud platí

$$\mathsf{P}(X = k) = p(1-p)^k, \quad k \in \mathbb{N}_0.$$

Snadno ověříme, že součet výše uvedených pravděpodobností je roven 1, a tedy opravdu odpovídají pravděpodobnostnímu rozdělení. Zkráceně ho budeme označovat *Geom(p)*. Pro první dva momenty geometrického rozdělení platí

$$\mathsf{E}\,X = \frac{(1-p)}{p}, \qquad \mathsf{var}\,X = \frac{(1-p)}{p^2}.$$

I geometrické rozdělení je odvozené od alternativního — pokud sledujeme posloupnost nezávislých náhodných pokusů s pravděpodobností úspěchu p, pak počet neúspěšných pokusů předcházejících prvnímu úspěšnému je náhodná veličina s rozdělením *Geom(p)*.

2.3.4 Hypergeometrické rozdělení

Náhodná veličina X má *hypergeometrické rozdělení s parametry* N, M, n, kde N, M, $n \in \mathbb{N}$, $M < N$ a $n < N$, pokud platí

$$\mathsf{P}(X = m) = \frac{\binom{M}{m}\binom{N-M}{n-m}}{\binom{N}{n}} \qquad \text{pro} \quad 0 \le m \le M \quad \text{a} \quad 0 \le n - m \le N - M.$$

Toto rozdělení je vhodné pro popis situace, kdy máme (např. v urně) N předmětů, z nichž M je prvního a $N - M$ je druhého druhu. Z urny náhodně (a najednou) vybereme n předmětů. Počet vybraných předmětů, které budou prvního druhu, má hypergeometrické rozdělení *Hg(N,M,n)*.

2.3.5 Poissonovo rozdělení

Náhodná veličina X má *Poissonovo rozdělení s parametrem* λ, kde $\lambda \in (0, \infty)$, pokud platí

$$\mathsf{P}(X = k) = \mathrm{e}^{-\lambda} \frac{\lambda^k}{k!}, \qquad k \in \mathbb{N}_0.$$

Snadno ověříme, že součet výše uvedených pravděpodobností je roven 1, a tedy skutečně odpovídají pravděpodobnostnímu rozdělení. Zkráceně ho budeme označovat *Pois(λ)*. Je známou vlastností Poissonova rozdělení, že se jeho střední hodnota a rozptyl rovnají, neboť platí

$$\mathsf{E}\, X = \mathsf{var}\, X = \lambda.$$

Poissonovo rozdělení se často používá k modelování náhodného počtu událostí, které nastaly v nějakém časovém intervalu. Parametr λ se pak nazývá *intenzita* Poissonova rozdělení a má význam středního počtu pozorovaných událostí.
Poissonovo rozdělení je také možno odvodit jako limitní případ binomického rozdělení *Binom (n,p)*, když snižujeme pravděpodobnost úspěchu $p = p_n$ a zvyšujeme počet pokusů n tak, že $\lim\limits_{n\to\infty} np_n = \lambda$. Pro podrobnosti viz proseminář nebo Dupač a Hušková (2013), kapitola 2.4.

2.3.6 Rovnoměrné rozdělení

Náhodná veličina X má *rovnoměrné rozdělení na intervalu* $[a, b]$, kde $a < b \in \mathbb{R}$ (zkráceně značíme *R(a,b)*, pokud má hustotu

$$f_X(x) = \frac{1}{b-a} \mathbb{1}_{(a,b)}(x), \qquad x \in \mathbb{R}.$$

Snadno zjistíme, že distribuční funkce je

$$F_X(x) = \int_{-\infty}^{x} f_X(t)\,\mathrm{d}t = \begin{cases} 0 & x \le a \\ \frac{x-a}{b-a} & a \le x \le b \\ 1 & x \ge b, \end{cases}$$

a P_X je rovno $\frac{1}{b-a}$ násobku Lebesgueovy míry omezené na interval $[a, b]$. Pro první dva momenty rovnoměrného rozdělení platí

$$\mathsf{E}\, X = \frac{b+a}{2}, \qquad \mathsf{var}\, X = \frac{(b-a)^2}{12}.$$

Rovnoměrné rozdělení je absolutně spojité rozdělení, které odpovídá generování náhodného čísla z intervalu $[a, b]$ a to „rovnoměrně“ náhodně. Tím je míněno, že každé dva podúseky intervalu $[a, b]$ stejné délky mají stejnou pravděpodobnost, že se v nich ono náhodné číslo bude nacházet — tento požadavek na rozdělení P_X přesně odpovídá tomu, že P_X je násobek Lebesgueovy míry (na intervalu [a,b]). Rovnoměrné rozdělení, např. $R(-0.5; 0.5)$ nebo $R(-0.005; 0.005)$ se používá jako model pro náhodnou chybu vzniklou zaokrouhlováním.

Zde končí předn. 6 (2.3.)

2.3.7 Exponenciální rozdělení

Náhodná veličina X má *exponenciální rozdělení s parametrem* λ, kde $\lambda > 0$ (zkráceně značíme *Exp(*λ*)*), pokud má hustotu

$$f_X(x) = \lambda e^{-\lambda x}\, \mathbb{1}_{(0,\infty)}(x), \qquad x \in \mathbb{R}.$$

Distribuční funkce je

$$F_X(x) = \begin{cases} 1 - e^{-\lambda x} & x > 0 \\ 0 & x \le 0. \end{cases}$$

Pro první dva momenty exponenciálního rozdělení platí

$$\mathsf{E}\, X = \frac{1}{\lambda}, \qquad \mathsf{var}\, X = \frac{1}{\lambda^2}.$$

Exponenciální rozdělení se často používá k modelování náhodné doby čekání na nějakou událost (doby trvání nějakého děje). Speciální vlastností exponenciálního rozdělení je tzv. vlastnost ztráty paměti — fakt, že rozdělení zbývající doby čekání nezávisí na tom, jak dlouho už čekáme.

Tvrzení 2.12 Nezáporná absolutně spojitá náhodná veličina X má exponenciální rozdělení právě když

$$\mathsf{P}(X > x + y \mid X > y) = \mathsf{P}(X > x), \qquad \forall x > 0, y > 0.$$

Důkaz: Viz např. Billingsley (1995) str.190. □

2.3.8 Normované normální rozdělení

Náhodná veličina X má *normované normální rozdělení* (zkráceně značíme *N(0,1)*), pokud má hustotu

$$f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \qquad x \in \mathbb{R}.$$

Hustota normovaného normálního rozdělení se tradičně značí symbolem $\varphi(x)$ a distribuční funkce normovaného normálního rozdělení se tradičně značí symbolem Φ:

$$F_X(x) = \Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}\, dt, \qquad x \in \mathbb{R}.$$

Hodnoty Φ jsou tabelovány, resp. jsou obsaženy ve standardních statistických softwarech. Pro liché momenty normovaného normálního rozdělení platí

$$\mathsf{E}\, X^{2k+1} = 0, \qquad k \in \mathbb{N}_0,$$

a pro rozptyl $\mathsf{var}\, X = 1$.

2.3.9 Obecné normální rozdělení

Náhodná veličina X má *normální rozdělení s parametry* μ *a* σ^2, kde $\mu \in \mathbb{R}$ a $\sigma^2 > 0$ (zkráceně značíme $N(\mu, \sigma^2)$), pokud má hustotu

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}}, \qquad x \in \mathbb{R}.$$

Lineární transformací normálního rozdělení získáme opět normální rozdělení, ovšem s jinými parametry.

Tvrzení 2.13 Buď Z náhodná veličina s normovaným normálním rozdělením $Z \sim N(0,1)$. Buď $Y = aZ + b$, kde $a \neq 0$, $b \in \mathbb{R}$. Pak Y má normální rozdělení $N(b, a^2)$.

Důkaz: Viz příklad v následující sekci. □

Snadným důsledkem tvrzení 2.13 je i vzorec pro distribuční funkci F_X rozdělení $N(\mu, \sigma^2)$ — platí:

$$F_X(t) = \Phi\left(\frac{t-\mu}{\sigma}\right), \qquad t \in \mathbb{R},$$

kde $\sigma = \sqrt{\sigma^2}$.
Za použití tvrzení 2.13 nebo přímo odvodíme pro první dva momenty normálního rozdělení $N(\mu, \sigma^2)$

$$\mathsf{E}\, X = \mu, \qquad \mathsf{var}\, X = \sigma^2.$$

V kapitole o limitních větách uvidíme, že součet velkého počtu nezávislých náhodných veličin má přibližně normální rozdělení. Proto se normální rozdělení dobře hodí a často používá jako model pro náhodnou chybu — např. (náhodnou nikoli systematickou) chybu fyzikálních měření, a také jako model rozdělení nějaké variabilní hodnoty, která je výsledkem velkého množství drobných náhodných vlivů — např. rozdělení nějakého rozměru nebo charakteristiky nebo fyziologické hodnoty v homogenní populaci jedinců.

Poznámka. Momenty většiny představených rozdělení budou spočítány na cvičení (případně srovnej s Dupač a Hušková (2013), kapitola 2.4.)

Značení. To, že má nějaká náhodná veličina konkrétní rozdělení, značíme zkráceně pomocí symbolu $\sim$, např. $X \sim Pois(3)$ značí, že náhodná veličina X má Poissonovo rozdělení s intenzitou 3.

2.4 Rozdělení funkce náhodné veličiny

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

Věta 2.14 (o rozdělení funkce náhodné veličiny) Buď X náhodná veličina s distribuční funkcí F_X a $g : \mathbb{R} \to \mathbb{R}$ měřitelná funkce. Pak $Y = g(X)$ je náhodná veličina s distribuční funkcí

$$F_Y(y) = \int_{\{x:g(x)\le y\}} \mathrm{d}F_X(x).$$

Důkaz: Platí

$$F_Y(y) = \mathsf{P}(Y \le y) = \mathsf{P}(g(X) \le y) = \mathsf{P}_X(\{x : g(x) \le y\}) = \int_{\{x:g(x)\le y\}} \mathrm{d}\mathsf{P}_X = \int_{\{x:g(x)\le y\}} \mathrm{d}F_X,$$

kde ve třetí rovnosti jsme použili větu 2.2 o přenosu integrace pro P_X. □

Poznámka. Speciálně pro absolutně spojitou náhodnou veličinu X dostaneme

$$F_Y(y) = \int_{\{x:g(x)\le y\}} f_X(x)\,\mathrm{d}x.$$

Pro diskrétní náhodnou veličinu X dostaneme

$$F_Y(y) = \sum_{x_i:g(x_i)\le y} p_i, \tag{2.4}$$

a zřejmě $\mathsf{P}_Y = \sum_{i\in I} p_i \delta_{g(x_i)}$, 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

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

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

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

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 dále Y lineární transformace X, tj. $Y = aX + b$, kde $a \neq 0$, $b \in \mathbb{R}$. Pak podle předchozího příkladu má Y hustotu

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

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 2.15 (o monotónní transformaci) Buď X náhodná veličina s distribuční funkcí F_X a buď S_X nosič rozdělení P_X (tj. $\mathsf{P}(X \in S_X) = 1$). Buď $g : S_X \to \mathbb{R}$ 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 2.14 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 $g : S_X \to \mathbb{R}$ ryze monotónní funkce diferencovatelná λ-s.v. na S_X. Pak náhodná veličina $Y = g(X)$ má hustotu

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

vzhledem k λ na $\mathbb{R}$.

I pro diskrétní náhodnou veličinu dochází pro monotónní transformaci ke zjednodušení vzorce (2.4):

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

Vyzkoušíme důsledek věty 2.15 na lineární transformaci:

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

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

(jak už jsme zjistili v příkladu za větou 2.14).

Zde končí předn. 7 (8.3.)

3 Náhodné vektory

3.1 Náhodné vektory

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

Definice 3.1 Buď $(\Omega, \mathcal{A}, \mathsf{P})$ pravděpodobnostní prostor. Měřitelné zobrazení $\mathbf{X} : (\Omega, \mathcal{A}) \to (\mathbb{R}^n, \mathcal{B}^n)$ nazveme *(n-rozměrným) náhodným vektorem.*

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

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

Poznámky.

- Takže opět $\mathsf{P}_\mathbf{X}$ je obraz míry P v zobrazení $\mathbf{X}$. Vše je obdobné jako u náhodných veličin.
- Platí, že $\mathbf{X} = (X_1, \dots X_n)^\mathsf{T}$ 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 $\mathsf{P}_\mathbf{X}$ náhodného vektoru $\mathbf{X}$. Jak uvidíme dále, všechna informace definující míru $\mathsf{P}_\mathbf{X}$ je obsažena i v distribuční funkci $F_\mathbf{X}$ náhodného vektoru $\mathbf{X}$, tedy funkci z $\mathbb{R}^n$ do $\mathbb{R}$.

Definice 3.3 *(Sdružená) distribuční funkce* náhodného vektoru $\mathbf{X}$ je definována jako

$$F_\mathbf{X}(\mathbf{x}) = \mathsf{P}\left(\bigcap_{i=1}^{n} \{X_i \le x_i\}\right), \qquad \mathbf{x} = (x_1, \dots x_n)^\mathsf{T} \in \mathbb{R}^n.$$

Zopakujme si:

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

(i) měřitelnými obdélníky, tj. systémem $\mathcal{S} = \{B_1 \times \cdots \times B_n : B_i \in \mathcal{B}, i = 1, \dots n\}$,

(ii) otevřenými intervaly $(\mathbf{a}, \mathbf{b}) = (a_1, b_1) \times \cdots \times (a_n, b_n)$, $\mathbf{a}, \mathbf{b} \in \mathbb{R}^n$

(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 $\mathsf{P}_{\mathbf{X}}$. Tedy $F_{\mathbf{X}}$ opravdu je ekvivalentní charakterizace rozdělení $\mathbf{X}$.

A jak z $\mathsf{P}_{\mathbf{X}}$ 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

$$\lim_{x_n \to \infty} F_{\mathbf{X}}(x_1, \ldots, x_n) = F_{(X_1, \ldots X_{n-1})^\top}(x_1, \ldots, x_{n-1}),$$

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

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

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

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 $\lim\limits_{k\to\infty} \mathsf{P}(D_k) = \mathsf{P}(B)$. Z monotonie míry P máme $\mathsf{P}(D_k) \le \mathsf{P}(B_k) \le \mathsf{P}(B)$, tedy i $\lim\limits_{k\to\infty} \mathsf{P}(B_k) = \mathsf{P}(B)$, což jsme měli dokázat. □

Poznámka. Z definice distribuční funkce zřejmě

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

pro každou permutaci $\pi : \{1, \ldots, n\} \to \{1, \ldots, n\}$. Takže z předchozí věty získáme limitními přechody z F_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 $\sigma((X_1, \ldots, X_{n-1})^\top) \subseteq \sigma(\mathbf{X}) \subseteq \mathcal{A}$.

A jak získat z $\mathsf{P}_{\mathbf{X}}$ marginální rozdělení podvektoru?

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

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

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

Buď $\mathbf{X} = (X_1, X_2)^\mathsf{T}$ náhodný vektor. Z předchozí poznámky vidíme, že $\mathsf{P}_\mathbf{X}$ 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 $\mathsf{P}_{(X,Y)^\mathsf{T}}$ je diskrétní míra s nosičem $\{0, 1\} \times \{0, 1, 2\}$ a můžeme jí popsat tabulkou pravděpodobností pro dané hodnoty:

$X \ \backslash \ Y$	0	1	2	$\sum$
0	$\frac{1}{8}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{2}$
1	$\frac{1}{8}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{2}$
$\sum$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$	1

Zřejmě marginální rozdělení X je $\mathsf{P}_X = \frac{1}{2}\delta_0 + \frac{1}{2}\delta_1$ a pro Y je $\mathsf{P}_Y = \frac{1}{4}\delta_0 + \frac{1}{2}\delta_1 + \frac{1}{4}\delta_2$.

Uvažujme teď X stejné, ale Y buď náhodná veličina popisující počet sudých čísel dohromady na kostkách číslo 1 a 2. Opět je $\mathsf{P}_{(X,Y)^\mathsf{T}}$ diskrétní míra s nosičem $\{0, 1\} \times \{0, 1, 2\}$ a můžeme jí popsat tabulkou pravděpodobností pro dané hodnoty:

$X \ \backslash \ Y$	0	1	2	$\sum$
0	$\frac{1}{4}$	$\frac{1}{4}$	0	$\frac{1}{2}$
1	0	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{2}$
$\sum$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$	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 $\mathsf{P}_\mathbf{X}$ do jednotlivých souřadnic. Víme z teorie míry, že tyto projekce určují celou $\mathsf{P}_\mathbf{X}$ 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 $\mathbf{c}$ z $\mathbb{R}^n$ takových, že $c_i \in \{a_i, b_i\}$ $\forall i \in \{1, \ldots, n\}$ 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) $\lim\limits_{x_i\to\infty, 1\le i\le n} F_{\mathbf{X}}(x_1,\dots,x_n)=1$;

(ii) $\lim\limits_{x_j\to-\infty} F_{\mathbf{X}}(x_1,\dots,x_n)=0,\ \forall j\in\{1,\dots,n\},\ \forall x_1,\dots,x_{j-1},x_{j+1},\dots,x_n\in\mathbb{R}$;

(iii) $F_{\mathbf{X}}$ je zprava spojitá v každé proměnné;

(iv) pro $(\mathbf{a},\mathbf{b}]$ neprázdný interval v $\mathbb{R}^n$ platí

$$\sum_{k=0}^{n}(-1)^k\sum_{\mathbf{c}\in\Delta_{n,k}}F_{\mathbf{X}}(\mathbf{c})\ge 0. \tag{3.1}$$

Zde končí přednáška 8 (8.3.)

Důkaz: (i) Uvědomme si, že $x_i\to\infty,\ \forall\, 1\le i\le n$ právě tehdy, když $\min_{1\le i\le n}x_i\to\infty$, a že z monotonie pravděpodobnosti máme

$$1\ge F_{\mathbf{X}}(x_1,\dots,x_n)\ge F_{\mathbf{X}}(\min_{1\le i\le n}x_i\ (1,\dots,1)).$$

Stačí tedy dokázat, že pro funkci $H(x)=F_{\mathbf{X}}(x(1,\dots,1))$, $x\in\mathbb{R}$, platí $\lim_{x\to\infty}H(x)=1$. Ovšem z monotonie pravděpodobnosti plyne, že H je funkce neklesající, a z normovanosti pravděpodobnosti plyne, že $H(x)\le 1,\ \forall x\in\mathbb{R}$. Tedy nutně musí existovat $\lim_{x\to\infty}H(x)$, která je ≤ 1, a která se rovná $\lim_{k\in\mathbb{N},k\to\infty}H(k)$. Označme $B_k=(-\infty,k(1,\dots,1)]$. Platí $B_k\nearrow\mathbb{R}^n$ a tedy ze spojitosti míry $\mathsf{P}_{\mathbf{X}}$ máme

$$\lim_{k\in\mathbb{N},k\to\infty}H(k)=\lim_{k\to\infty}\mathsf{P}_{\mathbf{X}}(B_k)=\mathsf{P}_{\mathbf{X}}\left(\bigcup_{k=1}^{\infty}B_k\right)=\mathsf{P}_{\mathbf{X}}(\mathbb{R}^n)=1,$$

čímž je (i) dokázáno.

(ii) Buďte $x_1,\dots,x_{j-1},x_{j+1},\dots,x_n\in\mathbb{R}$ pevné. Označme

$$H(x)=F_{\mathbf{X}}(x_1,\dots,x_{j-1},x,x_{j+1},\dots,x_n),\qquad x\in\mathbb{R}.$$

Z monotonie pravděpodobnosti je H funkce neklesající a nezáporná. Musí tedy existovat nezáporná $\lim_{x\to-\infty}H(x)$ a musí být rovna $\lim_{k\in\mathbb{N},k\to\infty}H(-k)$. Označme $B_k=(-\infty,(x_1,\dots,x_{j-1},-k,x_{j+1},\dots,x_n)]$. Platí $B_k\searrow\emptyset$ a tedy ze spojitosti míry $\mathsf{P}_{\mathbf{X}}$ máme

$$\lim_{x\to-\infty}H(x)=\lim_{k\in\mathbb{N},k\to\infty}H(-k)=\lim_{k\to\infty}\mathsf{P}_{\mathbf{X}}(B_k)=\mathsf{P}_{\mathbf{X}}\left(\bigcap_{k=1}^{\infty}B_k\right)=\mathsf{P}_{\mathbf{X}}(\emptyset)=0,$$

čímž je (ii) dokázáno.

(iii) Buďte $x_1,\dots,x_{j-1},x_{j+1},\dots,x_n\in\mathbb{R}$ pevné. Uvažujme opět funkci H z bodu (ii). Protože je funkce H neklesající, musí existovat $\lim_{y\to x_+}H(y)\ge H(x)$. Potřebujeme dokázat, že nastává rovnost. Víme, že $\lim_{y\to x_+} H(x) = \lim_{k\to\infty} H(x + \frac{1}{k})$, stačí tedy určit limitu na pravé straně. Označme $B_k = (-\infty, (x_1, \dots, x_{j-1}, x + \frac{1}{k}, x_{j+1}, \dots, x_n)]$. Platí $B_k \searrow B = (-\infty, (x_1, \dots, x_{j-1}, x, x_{j+1}, \dots, x_n])$ a tedy ze spojitosti míry $\mathsf{P}_{\mathbf{X}}$ máme

$$\lim_{y\to x_+} H(y) = \lim_{k\in\mathbb{N}, k\to\infty} H\left(x + \frac{1}{k}\right) = \lim_{k\to\infty} \mathsf{P}_{\mathbf{X}}(B_k) = \mathsf{P}_{\mathbf{X}}\left(\bigcap_{k=1}^{\infty} B_k\right) = H(x),$$

čímž je (iii) dokázáno.

(iv) Abychom dokázali (3.1), je třeba dokázat, že levá strana nerovnosti je rovna $\mathsf{P}(\mathbf{X} \in (\mathbf{a}, \mathbf{b}])$, což je nutně nezáporné. Podrobný důkaz lze najít např. v Dupač a 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

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

□

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

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

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

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

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

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

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

a tedy $\mu_F = \mathsf{P}_{\mathbf{X}}$. □

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 $(\Omega, \mathcal{A}, \mathsf{P})$ a nějaký náhodný vektor $\mathbf{X} : \Omega \to \mathbb{R}^n$, který má přesně toto rozdělení.

I pro náhodné vektory rozlišíme dva speciální případy $\mathsf{P}_\mathbf{X}$:

Definice 3.4 Náhodný vektor **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

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

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

Poznámka. Každé marginální rozdělení podvektoru vektoru **X** je také diskrétní (ověřte).

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

$$F_\mathbf{X}(\mathbf{x}) = \int_{-\infty}^{x_1} \ldots \int_{-\infty}^{x_n} f_\mathbf{X}(t_1, \ldots, t_n)\, \mathrm{d}t_n \ldots \mathrm{d}t_1, \qquad \forall\, \mathbf{x} \in \mathbb{R}^n.$$

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

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

Poznámka. Každé marginální rozdělení podvektoru absolutně spojitého náhodného vektoru **X** je také absolutně spojité (ověřte).

Vlastnost diskrétního rozdělení, resp. absolutně spojitého rozdělení, pro marginální rozdělení nemusí vynucovat 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 $\mathsf{P}_\mathbf{X}$ 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 2.5).

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

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

$$\mathsf{P}_{\mathbf{X}}((-\infty, \mathbf{x}]) = F_{\mathbf{X}}(\mathbf{x}) = \int_{(-\infty, \mathbf{x}]} f_{\mathbf{X}}(t_1, \ldots, t_n)\, \mathrm{d}(\nu_1 \otimes \cdots \otimes \nu_n)(\mathbf{t}), \qquad \forall \mathbf{x} \in \mathbb{R}^n,$$

a

$$\mathsf{P}_{\mathbf{X}}(B) = \int_B f_{\mathbf{X}}(\mathbf{t})\, \mathrm{d}(\nu_1 \otimes \cdots \otimes \nu_n)(\mathbf{t}), \qquad \forall B \in \mathcal{B}^n.$$

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

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

kde

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

pro ν_i-s.v. $y_i \in \mathbb{R}$.

Důkaz: Existence a tvrzení o $f_{\mathbf{X}}$ plyne z Radonovy-Nikodýmovy věty.
Existence a tvrzení o f_i plyne z Fubiniovy věty a věty 3.1 o marginální distribuční funkci. □

Poznámka. Zřejmě je funkce f_i rovna $\frac{\mathrm{d}\mathsf{P}_{X_i}}{\mathrm{d}\nu_i}$ 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 **X**, aby nebylo možné $\mathsf{P}_{\mathbf{X}} \ll \nu_1 \otimes \cdots \otimes \nu_n$?

3.2 Nezávislé náhodné veličiny

V této kapitole podrobněji prozkoumáme situaci, kdy je rozdělení $\mathsf{P}_{\mathbf{X}}$ náhodného vektoru $\mathbf{X} = (X_1, \ldots, X_n)^{\mathsf{T}}$ ve speciálním součinovém tvaru – tj. $\mathsf{P}_{\mathbf{X}} = \mathsf{P}_{X_1} \otimes \cdots \otimes \mathsf{P}_{X_n}$. Z teorie míry víme, že v takovém případě $\mathsf{P}_{\mathbf{X}}$ přesně odpovídá součinu svých projekcí. A v pravděpodobnosti tento případ odpovídá situaci, kdy rozdělení jednotlivých marginálních veličin X_i z **X** nenese v sobě žádnou informaci o rozdělení ostatních marginálů X_j, $j \neq i$, z **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 Buď $\{X_i, i \in I\}$ systém náhodných veličin na $(\Omega, \mathcal{A}, \mathsf{P})$, kde $I \neq \emptyset$ je libovolná indexová množina. Náhodné veličiny $\{X_i, i \in I\}$ nazveme *(vzájemně) nezávislé*, pokud pro každou konečnou množinu $J \subset I$ platí

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

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 $\mathsf{P}\left(\bigcap_{i \in I}\{X_i \in B_i\}\right) = \prod_{i \in I} \mathsf{P}(X_i \in B_i)$, $\forall B_i \in \mathcal{B}$, $i \in I$, neboť platnost ostatních podmínek z definice 3.6 pro $J \subset I$ dostanu dosazením $\mathbb{R}$ za B_i pro $i \in I \setminus J$.

Odpovídá naše definice nezávislých náhodných veličin tomu, co jsme chtěli? Tedy součinové míře pro $\mathsf{P}_{\mathbf{X}}$? Ano:

Věta 3.5 (o rozdělení vektoru s nezávislými složkami) Buď $\mathbf{X} = (X_1, \ldots, X_n)^\top$ náhodný vektor. Pak $\{X_i\}_{i=1}^n$ jsou nezávislé náhodné veličiny právě tehdy, když

$$\mathsf{P}_{\mathbf{X}} = \mathsf{P}_{X_1} \otimes \cdots \otimes \mathsf{P}_{X_n}.$$

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

$$\mathsf{P}_{\mathbf{X}}(B_1 \times \cdots \times B_n) = \prod_{i=1}^n \mathsf{P}_{X_i}(B_i)$$

z definice nezávislosti. Tedy $\mathsf{P}_{\mathbf{X}}$ se rovná součinové míře na měřitelných obdélnících, což je systém množin uzavřený na průniky a generující $\mathcal{B}^n$. Protože je $\mathsf{P}_{\mathbf{X}}(\mathbb{R}^n) = 1$ 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) Buď $\mathbf{X} = (X_1, \ldots, X_n)^\top$ náhodný vektor. Pak $\{X_i\}_{i=1}^n$ jsou nezávislé náhodné veličiny právě tehdy, když

$$F_{\mathbf{X}}(\mathbf{x}) = \prod_{i=1}^n F_{X_i}(x_i), \;\; \forall \mathbf{x} \in \mathbb{R}^n. \tag{3.2}$$

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

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

kde druhá rovnost plyne z definice 3.6.

Obráceně, pokud platí rovnost (3.2), tak $\mathsf{P}_{\mathbf{X}}$ se rovná součinové míře $\mathsf{P}_{X_1} \otimes \cdots \otimes \mathsf{P}_{X_n}$ na množinách ze systému $\mathcal{S} = \{(-\infty, \mathbf{x}] : \mathbf{x} \in \mathbb{R}^n\}$. Systém $\mathcal{S}$ je uzavřený na průniky a generuje $\mathcal{B}^n$. Protože je $\mathsf{P}_{\mathbf{X}}(\mathbb{R}^n) = 1$ 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 $\mathsf{P}_{\mathbf{X}}$, resp. sdruženou distribuční funkci $F_{\mathbf{X}}$.

Zde končí předn. 9 (9.3.)

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) Buď $\mathsf{P}_{\mathbf{X}}$ rozdělení n-rozměrného náhodného vektoru $\mathbf{X}$ splňující $\mathsf{P}_{\mathbf{X}} \ll \nu_1 \otimes \cdots \otimes \nu_n$ (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

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

kde $f_{X_i} = \frac{\mathrm{d}\mathsf{P}_{X_i}}{\mathrm{d}\nu_i}$, $i \in \{1, \ldots, n\}$.

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

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

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

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

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

"$\Leftarrow$" Dokážeme obdobně obráceným postupem. □

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

A co se stane, když máme náhodný vektor $\mathbf{X}$ s nezávislými marginálními rozděleními a transformujeme 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 $g_i : \mathbb{R} \to \mathbb{R}$, $i \in I$, 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

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

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. □

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

Příklad. Buď $(X, Y)^\mathsf{T}$ náhodný vektor s nezávislými složkami. Pak je i $(X^2, Y^2)^\mathsf{T}$ náhodný vektor s nezávislými složkami. Ale např. $(X + Y, X - Y)^\mathsf{T}$ už nemusí mít nezávislá marginální rozdělení – 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 $\mathsf{E}\,\mathbf{X} = (\mathsf{E}\,X_1, \dots, \mathsf{E}\,X_n)^\mathsf{T}$.

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

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

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

Důkaz: Složení měřitelných funkcí je zřejmě měřitelné, první a druhá rovnost plynou z věty o přenosu integrace 2.1, třetí z Radonovy-Nikodýmovy věty. □

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

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

Definice 3.7 Buďte X, Y náhodné veličiny definované na tomtéž pravděpodobnostním prostoru. Pak *kovariance* X a Y je definována jako

$$\operatorname{cov}(X, Y) = \mathsf{E}\,[(X - \mathsf{E}\,X)(Y - \mathsf{E}\,Y)]\,.$$

Koeficient korelace je definován jako

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

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

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

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

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

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

$$\mathsf{E}\,|XY| \leq (\mathsf{E}\,|X|^p)^{\frac{1}{p}}\,(\mathsf{E}\,|Y|^q)^{\frac{1}{q}}\,,$$

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

Důkaz: Viz *TMI*. □

A co nám Hölderova nerovnost říká o kovarianci?

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

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

takže

$$|\operatorname{cov}(X, Y)| \leq \mathsf{E}\,|(X - \mathsf{E}\,X)(Y - \mathsf{E}\,Y)| \leq \sqrt{\mathsf{E}\,(X - \mathsf{E}\,X)^2\ \mathsf{E}\,(Y - \mathsf{E}\,Y)^2} = \sqrt{\operatorname{var} X\ \operatorname{var} Y},$$

a $\operatorname{cor}(X, Y) \in [-1, 1]$.
Navíc tedy $|\operatorname{cor}(X, Y)| = 1$ 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 $\mathsf{E}\,X^2 < \infty$ a $\mathsf{E}\,Y^2 < \infty$, pak $\operatorname{cov}(X, Y) \in \mathbb{R}$ (ověřte).

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

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

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

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

A teď si rozmyslíme, jak vypají smíšené momenty náhodného vektoru, když má ten vektor nezávislé složky. Uvidíme, že mnoho věcí se zjednoduší. Začněme $\mathsf{E}\, X_1 X_2$:

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

Důkaz: Pokud je $\mathsf{E}\, X_1 X_2$ definována, pak platí

$$\begin{aligned} \mathsf{E}\, X_1 X_2 &= \int_{\mathbb{R}^2} x_1 x_2 \, \mathrm{d}\mathsf{P}_{\mathbf{X}}(\mathbf{x}) = \int_{\mathbb{R}^2} x_1 x_2 \, \mathrm{d}(\mathsf{P}_{X_1} \otimes \mathsf{P}_{X_2})(\mathbf{x}) = \int_{\mathbb{R}} \int_{\mathbb{R}} x_1 x_2 \, \mathrm{d}\mathsf{P}_{X_1}(x_1) \, \mathrm{d}\mathsf{P}_{X_2}(x_2) \\ &= \int_{\mathbb{R}} x_1 \, \mathrm{d}\mathsf{P}_{X_1}(x_1) \int_{\mathbb{R}} x_2 \, \mathrm{d}\mathsf{P}_{X_2}(x_2), \end{aligned}$$

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

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

$$\begin{aligned} \mathsf{E}\, g_n(X_1, X_2) &= \int_{\mathbb{R}^2} |x_1| \mathbb{1}_{[|x_1| \le n]} |x_2| \mathbb{1}_{[|x_2| \le n]} \, \mathrm{d}(\mathsf{P}_{X_1} \otimes \mathsf{P}_{X_2})(\mathbf{x}) \\ &= \mathsf{E}\, |X_1| \mathbb{1}_{[|X_1| \le n]} \mathsf{E}\, |X_2| \mathbb{1}_{[|X_2| \le n]} \le \mathsf{E}\, |X_1| \mathsf{E}\, |X_2|, \end{aligned} \tag{3.3}$$

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 $\mathsf{E}\, g_n(X_1, X_2) \nearrow \mathsf{E}\, |X_1 X_2|$ a $\mathsf{E}\, |X_1 X_2| \le \mathsf{E}\, |X_1| \mathsf{E}\, |X_2| < \infty$ z nerovnosti (3.3). $\square$

Poznámka. Uvědomte si, že k důkazu konečnosti momentu $\mathsf{E}\, X_1X_2$ z předpokladu konečnosti $\mathsf{E}\,|X_1|$ a $\mathsf{E}\,|X_2|$ nelze přímo využít Hölderovu nerovnost. Dokonce bez předpokladu nezávislosti X_1 a X_2 nemusí být $\mathsf{E}\, X_1X_2$ konečné (najděte (proti)příklad). Nezávislost je tedy důležitá nejen pro tvar $\mathsf{E}\, X_1X_2$, ale také pro to, aby pro konečnost $\mathsf{E}\, X_1X_2$ 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. 10 (15.3.)

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

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

Důkaz: Počítejme:

$$\operatorname{cov}(X_i, X_j) = \mathsf{E}\,[(X_i - \mathsf{E}\, X_i)(X_j - \mathsf{E}\, X_j)] = \\ \mathsf{E}\, X_iX_j - \mathsf{E}\,((\mathsf{E}\, X_i)X_j) - \mathsf{E}\,(X_i\mathsf{E}\, X_j) + \mathsf{E}\, X_i\mathsf{E}\, X_j = \mathsf{E}\, X_iX_j - \mathsf{E}\, X_i\mathsf{E}\, X_j.$$

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

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

Dále počítejme

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

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 $\operatorname{cov}(X_i, X_j) = 0$ pro $i \neq j$. Tedy dvojná suma je rovna 0 a

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

□

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

$$\operatorname{var}\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n} a_i^2 \operatorname{var} X_i + 2\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} a_i a_j \operatorname{cov}(X_i, X_j). \tag{3.4}$$

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

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

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

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

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

$X - Y \ \backslash\ X + Y$	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.:

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

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

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

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 $a_{i,j} = \operatorname{cov}(X_i, X_j)$, $i, j = 1, \ldots, n$, tj.

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

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

A jaké má varianční matice vlastnosti?

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

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

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

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

(iii) $|\operatorname{cov}(X_i, X_j)| \le \sqrt{\operatorname{var} X_i \operatorname{var} X_j}$ 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 $X_1, \ldots, X_n$ vzájemně nezávislé, pak je $\operatorname{Var} \mathbf{X}$ diagonální.

(v) $\operatorname{Var} \mathbf{X}$ je singulární právě tehdy, když existují $a_1, \ldots a_n \in \mathbb{R}$, 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 positivní semidefinitnost potřebujeme ověřit, že $\mathbf{a} \operatorname{Var} \mathbf{X} \, \mathbf{a}^{\mathsf{T}} \ge 0$. Počítejme:

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

druhá rovnost plyne z rozpisu (3.4).

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

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

(iii) Už jsme ukázali jako důsledek Hölderovy nerovnosti 3.10.
(iv) Je zřejmé z věty 3.12.
(v) Positivně semidefinitní matice $\mathsf{Var}\,\mathbf{X}$ je singulární právě tehdy, když není positivně definitní, čili když existuje $\mathbf{a} \in \mathbb{R}^n$, $\mathbf{a} \neq \mathbf{0}$ tak, že $\mathbf{a}\, \mathsf{Var}\,\mathbf{X}\, \mathbf{a}^\mathsf{T} = 0$. Ale

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

právě když $\mathbf{aX} - \mathsf{E}\,\mathbf{aX} = 0$ skoro jistě. □

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 $\mathsf{E}\, X_i = \mu$, $\mathsf{var}\, X_i = \sigma^2$, $\forall i = 1, \dots, n$, $\mu \in \mathbb{R}$, $\sigma^2 > 0$. Pak pro $\overline{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$ platí $\mathsf{E}\,\overline{X}_n = \mu$ a $\mathsf{var}\,\overline{X}_n = \frac{\sigma^2}{n}$.

Důkaz: Dostáváme snadno rozepsáním z linearity střední hodnoty a věty 3.13 o vlastnostech varianční matice, bod (ii). □

3.4 Rozdělení transformovaného náhodného vektoru

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

Uvažujme náhodný vektor $(X, Y)^\mathsf{T}$ s nezávislými složkami a měřitelné zobrazení $\Psi : \mathbb{R}^2 \to \mathbb{R}$. Buď $U = \Psi(X, Y)$. Rozdělení P_U je zřejmě $(\mathsf{P}_X \otimes \mathsf{P}_Y)\Psi^{-1}$, obraz součinové míry $\mathsf{P}_X \otimes \mathsf{P}_Y$. Ale jak tahle míra vypadá? Bude jednodušší odvodit distribuční funkci F_U, ta také charakterizuje rozdělení náhodné veličiny U.

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

$$\begin{aligned}
G_U(u) &= \int_\mathbb{R}\int_{\{y:\, \Psi(x,y)\le u\}} \mathrm{d}F_Y(y)\, \mathrm{d}F_X(x) = \int_\mathbb{R}\int_{\{y:\, \Psi(x,y)\le u\}} \mathrm{d}\mathsf{P}_Y(y)\, \mathrm{d}\mathsf{P}_X(x) \\
&= \int_\mathbb{R}\int_{\{x:\, \Psi(x,y)\le u\}} \mathrm{d}F_X(x)\, \mathrm{d}F_Y(y) = \int_\mathbb{R}\int_{\{x:\, \Psi(x,y)\le u\}} \mathrm{d}\mathsf{P}_X(x)\, \mathrm{d}\mathsf{P}_Y(y), \qquad u \in \mathbb{R},
\end{aligned}$$

a

$$\mathsf{E}\, U = \int_{\mathbb{R}} \int_{\mathbb{R}} \Psi(x, y) \, \mathrm{d}\mathsf{P}_Y(y) \, \mathrm{d}\mathsf{P}_X(x),$$

pokud má $\mathsf{E}\, U$ smysl.

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

$$\mathsf{E}\, U = \int_{\mathbb{R}^2} \Psi(x, y) \, \mathrm{d}(\mathsf{P}_X \otimes \mathsf{P}_Y)(x, y) = \int_{\mathbb{R}} \int_{\mathbb{R}} \Psi(x, y) \, \mathrm{d}\mathsf{P}_Y(y) \, \mathrm{d}\mathsf{P}_X(x).$$

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

$$\begin{aligned} G_U(u) &= \mathsf{P}(U \le u) = \mathsf{E}\, \mathbb{1}_{(-\infty, u]}(U) = \int_{\mathbb{R}} \int_{\mathbb{R}} \mathbb{1}(\Psi(x, y) \le u) \, \mathrm{d}\mathsf{P}_Y(y) \, \mathrm{d}\mathsf{P}_X(x) \\ &= \int_{\mathbb{R}} \int_{\{y : \Psi(x, y) \le u\}} \mathrm{d}F_Y(y) \, \mathrm{d}F_X(x), \quad u \in \mathbb{R}. \end{aligned}$$

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

Zde končí předn. 11 (16.3.)

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

$$F_U(u) = \int_{\mathbb{R}} F_X(u - y) \, \mathrm{d}F_Y(y) = \int_{\mathbb{R}} F_Y(u - x) \, \mathrm{d}F_X(x), \quad u \in \mathbb{R}. \tag{3.5}$$

Důkaz: Dosazením do předchozí věty máme

$$F_U(u) = \int_{\mathbb{R}} \int_{\{x + y \le u\}} \mathrm{d}F_Y(y) \, \mathrm{d}F_X(x) = \int_{\mathbb{R}} F_Y(u - x) \, \mathrm{d}F_X(x) = \int_{\mathbb{R}} F_X(u - y) \, \mathrm{d}F_Y(y), \quad u \in \mathbb{R}.$$

□

Rozdělení součtu nezávislých náhodných veličin má i svůj speciální název.

Definice 3.10 Buď $\Psi : (x, y) \mapsto (x + y)$. Pak pravděpodobnostní rozdělení $(\mathsf{P}_X \otimes \mathsf{P}_Y)\Psi^{-1}$ se nazývá *konvoluce pravděpodobnostních rozdělení* P_X a P_Y a značíme ji $\mathsf{P}_U = \mathsf{P}_X * \mathsf{P}_Y$.

Buďte F_X a F_Y distribuční funkce. Pak distribuční funkce F_U definovaná pomocí vztahu (3.5) se nazývá *konvoluce distribučních funkcí* a značíme $F_U = F_X * F_Y$.

Důsledek (věty o rozdělení součtu)**.** Buďte X a Y nezávislé absolutně spojité náhodné veličiny. Pak náhodná veličina $U = X + Y$ je také absolutně spojitá a její hustota je dána

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

Důkaz: Dosazením do (3.5) máme pro každé $u \in \mathbb{R}$

$$F_U(u) = \int_{\mathbb{R}} \int_{-\infty}^{u-x} f_Y(y) f_X(x)\,\mathrm{d}y\,\mathrm{d}x = \int_{\mathbb{R}} \int_{-\infty}^{u} f_Y(z-x) f_X(x)\,\mathrm{d}z\,\mathrm{d}x$$
$$= \int_{-\infty}^{u} \left[\int_{\mathbb{R}} f_Y(z-x) f_X(x)\,\mathrm{d}x \right] \mathrm{d}z,$$

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

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

□

Poznámka. Předpoklad nezávislosti X a Y není pro odvození hustoty součtu podstatný. Ověřte, že pro absolutně spojitý náhodný vektor $(X, Y)^\mathsf{T}$ se sdruženou hustotou $f_{(X,Y)^\mathsf{T}}(x, y)$ platí, že náhodná veličina $U = X + Y$ je také absolutně spojitá a její hustota je dána vzorcem

$$f_U(u) = \int_{\mathbb{R}} f_{(X,Y)^\mathsf{T}}(u-y, y)\,\mathrm{d}y = \int_{\mathbb{R}} f_{(X,Y)^\mathsf{T}}(x, u-x)\,\mathrm{d}x, \quad u \in \mathbb{R}.$$

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

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

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

Důkaz: Snadno z věty o úplné pravděpodobnosti. □

Pro příklady použití všech výše uvedených vět odkazujeme na cvičení.

A co když nejsou složky náhodného vektoru $\mathbf{X}$ nezávislé? Jak určíme $\mathsf{P}_\mathbf{X}\Psi^{-1}$?

Pokud má $\mathbf{X}$ diskrétní rozdělení $\mathsf{P}_\mathbf{X} = \sum_{j \in I} p_j\ \delta_{\mathbf{x}_j}$, kde I je spočetná indexová množina, pak prostě přepočítáme $\mathsf{P}_{\Psi(\mathbf{X})}$ jako obraz míry $\mathsf{P}_\mathbf{X}$ při zobrazení Ψ

$$\mathsf{P}_{\Psi(\mathbf{X})} = \mathsf{P}_\mathbf{X}\Psi^{-1} = \sum_{j \in I} p_j\ \delta_{\Psi(\mathbf{x}_j)}.$$

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

Pokud má $\mathbf{X}$ absolutně spojité 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 $\phi : M \to \mathbb{R}^n$ difeomorfismus. Buď $h : \phi(M) \to \mathbb{R}$ lebesgueovsky měřitelná funkce a $N \subset \phi(M)$ lebesgueovsky měřitelná množina. Pak

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

pokud má alespoň jedna strana smysl. Zde $\mathrm{J}\,\phi(\mathbf{t})$ je Jakobián zobrazení ϕ.

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 $\mathsf{P}(\mathbf{X} \in S_{\mathbf{X}}) = 1$ a $g : S_{\mathbf{X}} \to \mathbb{R}^n$ difeomorfismus. Pak rozdělení náhodného vektoru $\mathbf{Y} = g(\mathbf{X})$ má vzhledem k λ^n hustotu

$$f_{\mathbf{Y}}(\mathbf{y}) = f_{\mathbf{X}}(g^{-1}(\mathbf{y}))\left|\mathrm{J}g^{-1}(\mathbf{y})\right|\mathbb{1}_{g(S_{\mathbf{X}})}(\mathbf{y}),\ \ \mathbf{y} \in \mathbb{R}^n. \tag{3.6}$$

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

$$\mathcal{J} g^{-1}(\mathbf{y}) = \left(\mathcal{J} g(\mathbf{x})\Big|_{\mathbf{x}=g^{-1}(\mathbf{y})}\right)^{-1} \qquad \text{a} \qquad \left|\mathrm{J}g^{-1}(\mathbf{y})\right| = \left|\mathrm{J}g(\mathbf{x})\Big|_{\mathbf{x}=g^{-1}(\mathbf{y})}\right|^{-1},$$

kde $\mathcal{J} g(\mathbf{x})$ značí Jacobiho matici zobrazení g v bodě $\mathbf{x}$.

Důkaz: Z věty o obrazu míry víme $\mathsf{P}_{\mathbf{X}}(A) = \mathsf{P}_{\mathbf{Y}}(g(A))$ pro každé $A \in \mathcal{B}^n$, resp. každé $g(A) \in \mathcal{B}^n$. Pokud existuje hustota $f_{\mathbf{Y}}$, pak také $\mathsf{P}_{\mathbf{Y}}(g(A)) = \int_{g(A)} f_{\mathbf{Y}}(\mathbf{y})\,\mathrm{d}\mathbf{y}$. 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

$$\mathsf{P}_{\mathbf{X}}(A) = \int_A f_{\mathbf{X}}(\mathbf{x})\,\mathrm{d}\mathbf{x} = \int_{g(A)} f_{\mathbf{X}}\left(g^{-1}(\mathbf{y})\right)\,|\mathrm{J}g^{-1}(\mathbf{y})|\,\mathrm{d}\mathbf{y},$$

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

$$\int_{S^c_{\mathbf{X}}} f_{\mathbf{X}}(\mathbf{x})\,\mathrm{d}\mathbf{x} = 0,$$

a

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

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

Takže pro každou $B \in \mathcal{B}^n$ a funkci (3.6) dostáváme

$$\begin{aligned}
\mathsf{P}_{\mathbf{Y}}(B) &= \mathsf{P}_{\mathbf{Y}}(B \cap g(S_{\mathbf{X}})) + \mathsf{P}_{\mathbf{Y}}(B \backslash g(S_{\mathbf{X}})) = \mathsf{P}_{\mathbf{X}}\left(g^{-1}(B \cap g(S_{\mathbf{X}}))\right) + 0 \\
&= \int_{B \cap g(S_{\mathbf{X}})} f_{\mathbf{Y}}(\mathbf{y})\,\mathrm{d}\mathbf{y} + 0 = \int_{B \cap g(S_{\mathbf{X}})} f_{\mathbf{Y}}(\mathbf{y})\,\mathrm{d}\mathbf{y} + \int_{B \backslash g(S_{\mathbf{X}})} f_{\mathbf{Y}}(\mathbf{y})\,\mathrm{d}\mathbf{y} = \int_B f_{\mathbf{Y}}(\mathbf{y})\,\mathrm{d}\mathbf{y}.
\end{aligned}$$

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

□

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 $\mathsf{P}_{\mathbf{X}}$ 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) Buďte 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

$$f_U(u) = \int_{\mathbb{R}} f_X\left(\frac{u}{x}\right) f_Y(x) \frac{1}{|x|} \mathbb{1}(x \neq 0)\, \mathrm{d}x. \tag{3.7}$$

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

$$f_V(v) = \int_0^\infty f_X(vx) f_Y(x) x\, \mathrm{d}x.$$

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

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

Buď zobrazení g definované jako $g(x, y) = (xy, y)$, a označme $(U, V)^\mathsf{T} = g((X, Y)^\mathsf{T})$. Pak g je difeomorfismus na otevřené množině $S_{(X,Y)^\mathsf{T}} = \mathbb{R}^2 \backslash \mathbb{R} \times \{0\}$. Platí $\mathsf{P}_{(X,Y)^\mathsf{T}}\left((S_{(X,Y)^\mathsf{T}})^c\right) = \mathsf{P}(Y = 0) = 0$, neboť Y je z předpokladů absolutně spojitá náhodná veličina. Tedy $S_{(X,Y)^\mathsf{T}}$ je opravdu nosič rozdělení $\mathsf{P}_{(X,Y)^\mathsf{T}}$.

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

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

Předpoklady věty 3.19 jsou splněny a dostáváme

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

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

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

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

□

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 $(X,Y)^\top$ by byl absolutně spojitý náhodný vektor s hustotou $f_{(X,Y)^\top}$, pak by $U = XY$ byla stále absolutně spojitá náhodná veličina ovšem s hustotou

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

3.5 Příklady rozdělení náhodného vektoru

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

3.5.1 Multinomické rozdělení

nebylo v LS 22/23 přednášeno na přednášce

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

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

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

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 $\sum_{\mathbf{x}\in S_{\mathbf{X}}} \mathsf{P}(\mathbf{X} = \mathbf{x}) = 1$, takže $\mathsf{P}_{\mathbf{X}}$ 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 $\mathbf{X} = (X_1,\dots X_k)^\top$ 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

$$\mathsf{E}\,X_i = np_i \qquad \mathsf{var}\,X_i = np_i(1-p_i) \qquad \mathsf{cov}\,(X_i,X_j) = -np_ip_j,\quad i \neq j.$$

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

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

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 tento součet má hodnotu 1. Tedy

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

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 definovali normované normální rozdělení $N(0,1)$ jako pravděpodobnostní rozdělení s hustotou

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

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

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

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

Nyní definujeme mnohorozměrnou verzi normálního rozdělení.

Definice 3.11 Buď $\mathbf{Z} = (Z_1, \dots, Z_r)^\mathsf{T}$, $r \in \mathbb{N}$, kde Z_i jsou vzájemně nezávislé a mají $N(0,1)$ rozdělení. Buď $A_{n\times r}$, $n \in \mathbb{N}$, 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 $\Sigma = AA^\mathsf{T}$. Značíme $N_n(\mu, \Sigma)$.

Poznámka. Náhodný vektor $\mathbf{Z}$ z definice má zřejmě $\mathsf{E}\,\mathbf{Z} = \mathbf{0}$ a $\operatorname{Var}\mathbf{Z} = \mathbb{I}_r$ 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. $N(\mu, \mathbb{O}) = \delta_\mu$ kde $\mu \in \mathbb{R}^n$ je pevný vektor).

Důsledky (definice mnohorozměrného normálního rozdělení)**.** Buďte $\mathbf{X}$ 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 $\mathsf{E}\,\mathbf{X} = \mu$ a $\mathsf{Var}\,\mathbf{X} = \Sigma$.

(iii) Pro $k \in \mathbb{N}$ a matici $B_{k\times n}$ platí, že náhodný vektor $\mathbf{Y} = B\mathbf{X}$ má normální rozdělení $N_k(B\mu, B\Sigma B^\top)$.

(iv) Speciálně pro vektor $\mathbf{c} = (c_1, \dots, c_n) \in \mathbb{R}^n$ má náhodný vektor $\mathbf{cX}$ jednorozměrné normální rozdělení $N(\mathbf{c}\mu, \mathbf{c}\Sigma\mathbf{c}^\top)$.

(v) Marginální rozdělení X_i je $N(\mu_i, \sigma_i^2)$, kde $\sigma_i^2 = \Sigma_{i,i} = \mathsf{var}\,X_i$, $i = 1, \dots, n$.

(vi) Pokud je $\rho_{ij} = \mathsf{cor}\,(X_i, X_j)$ definována, pak má podvektor $(X_i, X_j)^\top$ rozdělení

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

Zde končí předn. 12 (22.3.)

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

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

a

$$\mathsf{Var}\,\mathbf{X} = \mathsf{Var}\,(A\mathbf{Z} + \mu) = A\mathsf{Var}\,\mathbf{Z}A^\top = A\mathbb{I}_r A^\top = AA^\top.$$

(iii) Podle definice platí, že $\mathbf{X} = A\mathbf{Z} + \mu$, tedy $Y = BA\mathbf{Z} + B\mu$ a podle definice 3.11 má k-rozměrné normální rozdělení s parametry $B\mu$ a $BA(BA)^\top = B\Sigma B^\top$.
(iv) Speciální případ bodu (iii) pro matici $1 \times n$.
(v) Použijeme bod (iv) pro speciální volbu $\mathbf{c} = (0, \dots, 1, \dots, 0) = \mathrm{e}_i$. Pak $X_i = \mathbf{cX}$ má rozdělení $N(\mathrm{e}_i\mu, \mathrm{e}_i\Sigma\mathrm{e}_i^\top)$.
(vi) Použijeme bod (iii) s maticí $B_{2\times n}$ složenou z vektorů e_i a e_j. □

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

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

má opět normální rozdělení $N(\sum_{i=1}^n c_i\mu_i, \mathbf{c}\Sigma\mathbf{c}^\top)$. 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 sdružené 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 spojité? A jakou má hustotu? – Odpověď závisí na tom, jaká je Σ.

Víme z lineární algebry, že každou symetrickou positivně semidefinitní matici $\Sigma \equiv \Sigma_{n\times n}$ lze rozepsat jako $\Sigma = AA^\mathsf{T}$ pro nějakou $A \equiv A_{n\times r}$, $r \leq n$ s plnou sloupcovou hodností. Tedy 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 $r < n$. Proto nelze očekávat, že by $\mathsf{P}_\mathbf{X} \ll \lambda^n$. 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 $\mathsf{Var}\,\mathbf{X} = \Sigma$. Z věty 3.13 o vlastnostech varianční matice, bod (v) plyne, že existuje $\mathbf{a} = (a_1, \dots, a_n) \in \mathbb{R}^n$ takové, že $\mathbf{aX} = k$ skoro jistě, kde $k \in \mathbb{R}$ je nějaká konstanta. Označme

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

Pak $\mathsf{P}_\mathbf{X}(E) = 1$ ale $\lambda^n(E) = 0$. Takže opravdu neplatí $\mathsf{P}_\mathbf{X} \ll \lambda^n$.

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 $\mathsf{P}_\mathbf{X} = \delta_\mu$.

Pro Σ regulární je ale rozdělení $N_n(\mu, \Sigma)$ absolutně spojité, 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 $\mathsf{P}_\mathbf{X} \ll \lambda^n$ a

$$f_\mathbf{X}(\mathbf{x}) = \frac{1}{(2\pi)^{\frac{n}{2}}\sqrt{\det \Sigma}}\, e^{-\frac{1}{2}(\mathbf{x}-\mu)^\mathsf{T}\Sigma^{-1}(\mathbf{x}-\mu)}, \qquad \mathbf{x} \in \mathbb{R}^n. \tag{3.8}$$

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

$$f_\mathbf{Z}(\mathbf{z}) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} e^{-\frac{z_i^2}{2}} = \frac{1}{(2\pi)^{\frac{n}{2}}\sqrt{\det \mathbb{I}_n}}\, e^{-\frac{1}{2}(\mathbf{z}-\mathbf{0})^\mathsf{T}\mathbb{I}_n^{-1}(\mathbf{z}-\mathbf{0})}, \qquad \mathbf{z} \in \mathbb{R}^n,$$

tedy (3.8) platí.

Buď Σ positivně definitní a $\mu \in \mathbb{R}^n$. Pak existuje matice $A \equiv A_{n\times n}$ tak, že $\Sigma = AA^\mathsf{T}$. 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 $g : \mathbb{R}^n \to \mathbb{R}^n$, $g(\mathbf{x}) = A\mathbf{x} + \mu$ 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

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

což je přesně (3.8), neboť $(A^{-1})^\mathsf{T}A^{-1} = (AA^\mathsf{T})^{-1} = \Sigma^{-1}$. □

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

Příklad. Uvažujme náhodný vektor $\mathbf{Z} = (Z_1, Z_2)^\top$ s dvourozměrným normálním rozdělením $N_2(\mu, \Sigma)$. Označme

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

pak

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

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

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

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

b) Pro náhodný vektor $\mathbf{Z}$ definovaný jako

$$\mathbf{Z} = \mu + \begin{pmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{pmatrix} \begin{pmatrix} X \\ Y \end{pmatrix},$$

kde X a Y jsou z bodu a), platí $\mathbf{Z} \sim N_2(\mu, \Sigma)$ pro původní obecnou Σ. A obráceně pro původní obecné $\mathbf{Z}$ platí

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

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 součinovém tvaru právě tehdy, když $\rho = \operatorname{cor}(X, Y) = 0$. Tedy hustota podvektoru $(X_i, X_j)^\top$ z důsledku (iv) ze str. 54 je v součinovém tvaru právě tehdy, když $\rho_{ij} = 0$. V tom případě je hustota rovna

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

Ovšem z věty 3.7 je hustota $f_{(X_i,X_j)^\top}$ v součinovém 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 $\operatorname{cor}(X, Y) = 0$, 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 $(X, Y)^\top$ 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 $(X, Y)^\top$ má sdružené normální rozdělení (a korelace $\operatorname{cor}(X, Y) = 0$).

V části přednášky o statistice, budeme používat ještě dvě rozdělení odvozená od normálního rozdělení. Představíme si je už zde.

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

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

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

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

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

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

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

Definice 3.13 Buďte $X \sim N(0, 1)$ a $Y \sim \chi^2_n$, $n \in \mathbb{N}$, nezávislé náhodné veličiny. Pak náhodná veličina $T = \frac{X}{\sqrt{\frac{Y}{n}}}$ má tzv. *Studentovo t_n-rozdělení* (neboli *t-rozdělení o n stupních volnosti*) s hustotou

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

Hustota t_n-rozdělení 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

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

a distribuční funkci

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

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

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

není definován a tedy neexistuje střední hodnota $\mathsf{E}\,T$ Cauchyho (resp. t_1) rozdělení. A ani žádně 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 $t \in \mathbb{R}$

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

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^2_n a t_n rozdělení, ale je třeba je umět odvodit jako příklad.

Zde končí přední. 13 (23.3.)

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 a nezávisle. Co nám naše matematická teorie říká o takové situaci, a jak to odpovídá realitě?

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

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

Co nám říká naše zkušenost o chování $\overline{X}_n$? Třeba ve fyzice, když provádíme n nezávislých identických experimentů s naměřeným výsledkem X_i (pozor, je náhodný), pak jejich průměr $\overline{X}_n$ (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

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

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

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow[n\to\infty]{} \mathsf{E}\, X_1,$$

budeme si ovšem muset vyjasnit, co znamená ta $\longrightarrow$.

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

$$\sqrt{n}\left(\overline{X}_n - \mathsf{E}\, X_1\right) \xrightarrow[n\to\infty]{} X,$$

kde $X \sim N(0, \sigma^2)$, neboli náhodná veličina na pravé straně šipky má normální rozdělení se střední hodnotou 0 a nějakým (nenulovým) rozptylem. Centrální limitní věta tedy mluví o tom, jak moc se empirický průměr $\overline{X}_n$ liší od střední hodnoty $\mathsf{E}\, X_1$. I zde si budeme muset vyjasnit, co znamená $\longrightarrow$.

4.1 Cantelliho a Borelova věta

Začneme ovšem jednodušeji:

Příklad. Buď $\{X_i\}_{i=1}^{\infty}$ posloupnost nezávislých stejně rozdělených náhodných veličin na tomtéž pravděpodobnostním prostoru. A nechť X_i mají alternativní rozdělení s pravděpodobností úspěchu $p = \frac{1}{2}$, tj. pro všechna $i \in \mathbb{N}$ platí

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

Takové posloupnosti $\{X_i\}_{i=1}^{\infty}$ se říká *posloupnost bernoulliovských pokusů* nebo jen *bernoulliovská 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 zvídavé čtenáře tu mohlo napadnout — a jak vůbec víme, že bernoulliovská posloupnost existuje? Nebo obecněji, jak víme, že existuje pravděpodobnostní prostor $(\Omega, \mathcal{A}, \mathsf{P})$ 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 zvídavé čtenáře musíme odkázat na kurz Teorie pravděpodobnosti I, který se učí ve třetím ročníku zaměření stochastika. A nebo na Proseminář z pravděpodobnosti a matematické statistiky. Nadále tedy budeme brát existenci nekonečné posloupnosti nezávislých náhodných veličin jako fakt bez důkazu.

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

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

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

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

Podobně, pokud je $\omega \in \liminf_{n\to\infty} A_n$, 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 $\liminf_{n\to\infty} A_n$ 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 lim sup a lim inf odpovídají lim sup a lim inf pro funkce v tom smyslu, že

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

Poznámka. Ověřte, že platí

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

Tato rovnost může být užitečná při výpočtech.

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

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

Důkaz: Počítejme:

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

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

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

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

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

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

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 $A_1, A_2, \ldots$ nezávislé, může být pravděpodobnost $\limsup_{n\to\infty} A_n$ jen 0 nebo 1, nic jiného nemůže nastat. Bez předpokladu nezávislosti ovšem může být $\mathsf{P}\left(\limsup_{n\to\infty} A_n\right)$ rovna čemukoli z $[0,1]$ (ověřte).

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

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

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

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

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

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

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

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

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

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

A jak je to tedy s naší bernoulliovskou posloupností?

Příklad. Mějme posloupnost bernoulliovských 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

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

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

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

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

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

$$\limsup_{n\to\infty} A_n = \emptyset \qquad \text{a} \qquad \mathsf{P}\left(\limsup_{n\to\infty} A_n\right) = \mathsf{P}(\emptyset) = 0.$$

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 $(\Omega, \mathcal{A}, \mathsf{P})$ a neznáme náhodné veličiny X_n jako funkce z Ω. Známe jen něco o rozdělení X_n, jako třeba v následujícím příkladě:

Příklad. Buďte $X_1, X_2, \ldots$ nezávislé náhodné veličiny definované na tomtéž pravděpodobnostním prostoru, o kterých víme, že $\mathsf{E}\, X_n = 0$ a $\operatorname{var} X_n = 1$ 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

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

Z Čebyševovy nerovnosti 2.9 víme, že $\mathsf{P}(|X_N| \geq n) \leq \frac{\operatorname{var} X_n}{n^2}$, takže

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

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

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

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

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

4.2 Konvergence posloupnosti náhodných veličin

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

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

Připomeňme si Stirlingovu formuli, tj. tvrzení

$$n! \sim \sqrt{2\pi n}\, n^n \mathrm{e}^{-n} \qquad \text{pro} \quad n \to \infty.$$

Trochu více formálně – platí:

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

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

$$\mathsf{P}\left(\frac{1}{2n}\sum_{i=1}^{2n} X_i = \frac{1}{2}\right) = \binom{2n}{n} 2^{-2n} \quad \sim \quad \frac{(2n)^{2n}\sqrt{2\pi 2n}}{(n^n\sqrt{2\pi n})^2}\, 2^{-2n} = \frac{1}{\sqrt{\pi n}} \xrightarrow[n\to\infty]{} 0.$$

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

$$\forall \epsilon > 0 \quad \mathsf{P}\left(\left|\overline{X}_n - \mathsf{E}\, X_1\right| \le \epsilon\right) \xrightarrow[n\to\infty]{} 1, \tag{4.1}$$

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

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

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

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

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

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

Zde končí předn. 14 (29.3.)

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

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 $Y_n \xrightarrow[n\to\infty]{\mathsf{sj}} Y \Rightarrow Y_n \xrightarrow[n\to\infty]{\mathsf{P}} Y$.

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 $Y_n \xrightarrow[n\to\infty]{L^p} Y \Rightarrow Y_n \xrightarrow[n\to\infty]{\mathsf{P}} Y$.

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

Příklad. Buď $\Omega = [0, 1]$, $\mathcal{A} = \mathcal{B}(\Omega)$, $\mathsf{P} = \lambda$ a náhodné veličiny X_n, $n \in \mathbb{N}$ definujme jako

$$X_n(\omega) = \mathbb{1}\left(\omega \in \left[\frac{n - 2^{\lfloor \log_2 n \rfloor}}{2^{\lfloor \log_2 n \rfloor}}, \frac{n - 2^{\lfloor \log_2 n \rfloor} + 1}{2^{\lfloor \log_2 n \rfloor}}\right]\right), \qquad \omega \in \Omega.$$

To jest $X_1 = 1$, $X_2 = \mathbb{1}_{[0, \frac{1}{2}]}$, $X_3 = \mathbb{1}_{[\frac{1}{2}, 1]}$, $X_4 = \mathbb{1}_{[0, \frac{1}{4}]}$, $X_5 = \mathbb{1}_{[\frac{1}{4}, \frac{1}{2}]}$, ... Pak platí:

- $\forall \omega$ existuje nekonečně mnoho indexů takových, že $X_n(\omega) = 1$, takže X_n nekonvergují k 0 P-skoro jistě.
- Pro $\epsilon \in (0, 1)$ je $\mathsf{P}(|X_n| > \epsilon) = \mathsf{P}(X_n = 1) \xrightarrow[n\to\infty]{} 0$, takže $X_n \xrightarrow[n\to\infty]{\mathsf{P}} 0$.
- Také $\mathsf{E}\,|X_n - 0|^p \xrightarrow[n\to\infty]{} 0$, takže $X_n \xrightarrow[n\to\infty]{L^p} 0$ 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 $(\Omega, \mathcal{A}, \mathsf{P})$. Pak platí

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

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

$$Y_n \xrightarrow[n\to\infty]{\mathsf{P}} 0 \;\Rightarrow\; a_n Y_n \xrightarrow[n\to\infty]{\mathsf{P}} 0.$$

Důkaz: Buď $\epsilon > 0$. Pak

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

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

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

□

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

Věta 4.6 (o spojité transformaci) Buďte $Y_1, Y_2, \ldots$ náhodné veličiny na stejném pravděpodobnostním prostoru $(\Omega, \mathcal{A}, \mathsf{P})$ a $g : \mathbb{R} \to \mathbb{R}$ funkce spojitá na otevřené množině S_Y, nosiči rozdělení náhodné veličiny Y (tj. platí $\mathsf{P}(Y \in S_Y) = 1$). Pak

(i) $Y_n \xrightarrow[n\to\infty]{\mathsf{P}} Y \;\Rightarrow\; g(Y_n) \xrightarrow[n\to\infty]{\mathsf{P}} g(Y)$

(ii) $Y_n \xrightarrow[n\to\infty]{\text{sj}} Y \;\Rightarrow\; g(Y_n) \xrightarrow[n\to\infty]{\text{sj}} g(Y)$.

Důkaz: (i) Důkaz pouze z definice je technický, vynecháme z časových důvodů. Krátký důkaz využívající dalších vlastností konvergence v pravděpodobnosti lze najít např. v Kallenberg (2002), Lemma 4.3.

(ii) Buď A množina splňující $\mathsf{P}(A) = 1$ taková, že $Y_n(\omega) \xrightarrow[n\to\infty]{} Y(\omega)$ pro každé $\omega \in A$. Pak pro každé $\omega \in M = Y^{-1}(S_Y) \cap A$ uplatníme větu o limitě spojitě transformované posloupnosti reálných čísel, abychom získali, že $g(Y_n)(\omega) \to g(Y)(\omega)$. Z předpokladů věty ovšem máme, že $\mathsf{P}(M) = 1$, tedy $g(Y_n) \xrightarrow[n\to\infty]{\text{sj}} g(Y)$. □

Poznámka. V předpokladech věty se požaduje, aby nosič S_Y byl otevřená množina. V případě, že $\mathsf{P}(Y = c) = 1$, se tedy požaduje, aby g byla spojitá na nějakém otevřeném okolí S_Y bodu c.

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

4.3 Slabý zákon velkých čísel

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

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

$$\lim_{n\to\infty} \left(\frac{1}{n^2} \sum_{i=1}^{n} \operatorname{var} X_i \right) = 0. \tag{4.2}$$

Potom platí $|\overline{X}_n - \mathsf{E}\,\overline{X}_n| \xrightarrow[n\to\infty]{\mathsf{P}} 0$.

Důkaz: Použijeme postupně Čebyševovu nerovnost z věty 2.9 a vzájemnou nezávislost $\{X_i\}$:

$$\mathsf{P}(|\overline{X}_n - \mathsf{E}\,\overline{X}_n| > \epsilon) \le \frac{\operatorname{var}\overline{X}_n}{\epsilon^2} = \frac{1}{\epsilon^2}\frac{1}{n^2}\sum_{i=1}^{n}\operatorname{var}X_i,$$

což konverguje k 0 při $n \to \infty$ pro každé $\epsilon > 0$ z předpokladu (4.2). □

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 $(X_n - \mathsf{E}\,X_n)$ konvergují v pravděpodobnosti k 0. Posun všech X_n o střední hodnotu je potřeba, abychom průměrovali jen náhodné variace, nikoli systematické (nenáhodné) vychýlení $\mathsf{E}\,X_n$.

A co nám ZVČ říká o bernoulliovské posloupnosti?

Příklad. Buď $\{X_n\}_{n=1}^{\infty}$ bernoulliovská posloupnost s pravděpodobností úspěchu p. Tedy náhodné veličiny X_n jsou vzájemně nezávislé, alternativně rozdělené, splňující $\mathsf{E}\,X_n^2 < \infty\ \forall n \in \mathbb{N}$ a $\mathsf{E}\,X_n = p$, $\operatorname{var}X_n = p(1-p)$. Podmínka (4.2) je splněna, neboť $\frac{1}{n^2}(np(1-p)) \xrightarrow[n\to\infty]{} 0$. $\{X_n\}_{n=1}^{\infty}$ tedy splňuje Čebyševův ZVČ a platí $|\overline{X}_n - p| \xrightarrow[n\to\infty]{\mathsf{P}} 0$, neboli

$$\mathsf{P}(|\overline{X}_n - p| \le \epsilon) \xrightarrow[n\to\infty]{} 1, \qquad \forall \epsilon > 0.$$

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

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

Pokud je ale $\mathsf{E}\,|X_n| = \infty$, pak tvrzení ZVČ platit nemusí. Uvažujme například $\{X_n\}_{n=1}^{\infty}$ nezávislé, stejně rozdělené s Cauchyho rozdělením (neboli t_1-rozdělení, viz str.53). Pak $\mathsf{E}\,X_n$ neexistuje a platí $\frac{1}{n}\sum_{i=1}^{n} X_i \sim$ Cauchy (ověřte, resp. viz cvičení). Tedy $\overline{X}_n$ určitě nekonverguje ke konstantě.

4.4 Silný zákon velkých čísel

V předchozí sekci jsme si dokázali, že pro bernoulliovskou posloupnost s $p = \frac{1}{2}$ platí, že četnost úspěchů $\overline{X}_n$ bude pro velké n blízko $p = \frac{1}{2} = \mathsf{E}\,X_1$ s velkou pravděpodobností. Ale my bychom chtěli ukázat ještě víc: když hodíme např. 100× mincí, tak se může stát, že s malou pravděpodobností se $\overline{X}_n$ 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 $|\overline{X}_n - \mathsf{E}\,\overline{X}_n| \xrightarrow[n\to\infty]{\text{sj}} 0$.

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

$$\sum_{n=1}^{\infty} \frac{\operatorname{var} X_n}{n^2} < \infty. \tag{4.3}$$

Potom platí $|\overline{X}_n - \mathsf{E}\,\overline{X}_n| \xrightarrow[n\to\infty]{\text{sj}} 0$.

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

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

$$\mathsf{P}\left(\max_{1\le k\le n} |S_k - \mathsf{E}\,S_k| \ge \epsilon\right) \le \frac{\operatorname{var} S_n}{\epsilon^2} = \frac{\sum_{k=1}^{n} \operatorname{var} X_k}{\epsilon^2}, \qquad \forall \epsilon > 0,\ \forall n \in \mathbb{N}. \tag{4.4}$$

Zde končí předn. 15 (30.3)

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

$$A_k = \{\max_{1\le j\le k} |S_j| < \epsilon\} \qquad B_k = \bigcap_{j=1}^{k-1} \{|S_j| < \epsilon\} \cap \{|S_k| \ge \epsilon\}.$$

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

$$\begin{aligned}\int_{B_k} S_n^2 \,\mathrm{d}\mathsf{P} &= \mathsf{E}\,(S_n \mathbb{1}_{B_k})^2 = \mathsf{E}\,((S_n - S_k)\mathbb{1}_{B_k} + S_k \mathbb{1}_{B_k})^2 \\ &= \mathsf{E}\,(S_n - S_k)^2 \mathbb{1}_{B_k} + 2\mathsf{E}\,(S_n - S_k) S_k \mathbb{1}_{B_k} + \mathsf{E}\,S_k^2 \mathbb{1}_{B_k} \ge \mathsf{E}\,S_k^2 \mathbb{1}_{B_k},\end{aligned}$$

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

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

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 $B_k \subseteq \{|S_k| \ge \epsilon\}$ a tedy

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

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

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

Zřejmě také platí

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

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

Poznámka. Kolmogorovova nerovnost není užitečná jen pro důkaz SZVČ. Mějme posloupnost nezávislých stejně rozdělených náhodných veličin $\{X_j\}_{j=1}^{\infty}$ s $\mathrm{E}\, X_1 = 0$. 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

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

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

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

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

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

To jest $\max_{1 \leq k \leq 2^n} |S_k - \mathrm{E}\, S_k| \leq \epsilon 2^n$ platí s pravděpodobností 1 až na konečně mnoho vyjímek, a $\limsup\limits_{n \to \infty} \frac{|S_{2^n} - \mathrm{E}\, S_{2^n}|}{2^n} \leq \epsilon$ 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 $|S_m - \mathrm{E}\, S_m| \leq \max_{1 \leq k \leq 2^{n+1}} |S_k - \mathrm{E}\, S_k|$, tedy s pravděpodobností 1

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

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

$$\limsup_{m\to\infty} \frac{|S_m(\omega) - \mathsf{E}\, S_m|}{m} \leq 2\epsilon \qquad \forall \omega \notin N(\epsilon).$$

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

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

a $\lim\limits_{m\to\infty} \frac{|S_m(\omega) - \mathsf{E}\, S_m|}{m} = 0$ 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).

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

První člen je konečný, druhý budeme dále upravovat

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

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. □

Poznámka. Z rozpisu (4.7) lze nahlédnout, že kdyby byly všechny $\operatorname{var} X_k$ 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 $\sum_{n=1}^{\infty} \frac{\operatorname{var} X_n}{n^2} < \infty$ je postačující pro SZVČ, nikoli nutná. Existuje posloupnost $\{X_n\}_{n=1}^{\infty}$ nezávislých náhodných veličin splňující $\sum_{n=1}^{\infty} \frac{\operatorname{var} X_n}{n^2} = \infty$, $\mathsf{E}\, X_n = 0$, a $\mathsf{P}(\lim\limits_{n\to\infty} \overline{X}_n = 0) = 1$, tj. $\overline{X}_n \xrightarrow[n\to\infty]{\text{sj}} 0$.

Abychom ji zkonstruovali volme X_n nezávislé s

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

kde

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

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

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

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

A co nám SZVČ říká o bernoulliovské posloupnosti?

Příklad. Buď $\{X_n\}_{n=1}^{\infty}$ bernoulliovská posloupnost s pravděpodobností úspěchu p. Neboli X_n jsou vzájemně nezávislé, alternativně rozdělené, $\mathsf{E}\, X_n^2 < \infty\ \forall n \in \mathbb{N}$ a $\mathsf{E}\, X_n = p$, $\operatorname{var} X_n = p(1-p)$. 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 $|\overline{X}_n - p| \xrightarrow[n\to\infty]{\text{sj}} 0$, neboli

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

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

Poznámka. ! Pozor ! SZVČ neříká, že

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

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!

Zde končí předn. 16 (5.4.)

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

Nicméně v případě nezávislých, stejně rozdělených náhodných veličin 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

$$\overline{X}_n \xrightarrow[n\to\infty]{\text{sj}} \mu \ \text{ pro nějaké } \ \mu \in \mathbb{R} \quad \Leftrightarrow \quad \mathsf{E}\,|X_1| < \infty.$$

V takovém případě $\mathsf{E}\,X_1 = \mu$.

K důkazu budeme potřebovat následující lemma.

Lemma 4.11 Pro libovolnou náhodnou veličinu X platí

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

a

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

Důkaz: $\mathsf{E}\,|X|$ si můžeme přepsat jako integrál podle pravděpodobnosti P a ten rozdělit na integrály přes po dvou disjunktní množiny $M_n = \{\omega : n \leq |X(\omega)| < n+1\}$, $n \in \mathbb{N}_0$, pro které platí

$$\bigcup_{n\in\mathbb{N}_0} M_n = \Omega.$$

Na každé množině M_n zvlášť odhadneme integrand $|X(\omega)|$ shora, takže dostaneme

$$\mathsf{E}\,|X| = \int_\Omega |X(\omega)|\,\mathrm{d}\mathsf{P}(\omega) = \sum_{n=0}^{\infty} \int\limits_{\{\omega: n\leq |X(\omega)|<n+1\}} |X(\omega)|\,\mathrm{d}\mathsf{P}(\omega) \leq \sum_{n=0}^{\infty}(n+1)\mathsf{P}(n \leq |X| < n+1).$$

A obdobně dostaneme i odhad zespoda:

$$\sum_{n=0}^{\infty} \int\limits_{\{\omega: n\leq |X(\omega)|<n+1\}} |X(\omega)|\,\mathrm{d}\mathsf{P}(\omega) \geq \sum_{n=0}^{\infty} n\ \mathsf{P}(n \leq |X| < n+1) = \sum_{n=0}^{\infty}(n+1)\mathsf{P}(n \leq |X| < n+1) - 1.$$

Zbývá dokázat první část lemmatu, což uděláme pomocí prohození pořadí sčítání (můžeme, neboť všechny členy jsou nezáporné) ve dvojné sumě:

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

□

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

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

Ty budou také vzájemně nezávislé a zřejmě

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

Z konečnosti $\mathsf{E}\,|X_1|$ a lemmatu 4.11 máme $\sum_{n=1}^{\infty} \mathsf{P}(|X_1| \geq n) < \infty$, takže z Cantelliho věty plyne $\mathsf{P}(\limsup_{n\to\infty}\{X_n \neq Y_n\}) = 0$. Tedy $X_n = Y_n$ s pravděpodobností 1 až na konečně mnoho n. Proto $(\overline{X}_n - \overline{Y}_n) \xrightarrow[n\to\infty]{\text{sj}} 0$. Stačí tedy dokázat, že SZVČ splňují $\{Y_n\}_{n=1}^{\infty}$, a že $\lim_{n\to\infty} \mathsf{E}\,\overline{Y}_n = \mu$. Potom už bude platit

$$\overline{X}_n - \mu = (\overline{X}_n - \overline{Y}_n) + (\overline{Y}_n - \mathsf{E}\,\overline{Y}_n) + (\mathsf{E}\,\overline{Y}_n - \mu) \quad \xrightarrow[n\to\infty]{\text{sj}} \quad 0 + 0 + 0 = 0.$$

Ovšem

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

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

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).
- $\mathsf{E}\,|Y_n|^2 < \infty$, neboť Y_n jsou omezené. Navíc platí $\operatorname{var} Y_n \leq \mathsf{E}\,|Y_n|^2$.
- Stačí tedy ukázat $\sum_{n=1}^{\infty} \frac{\mathsf{E}\,Y_n^2}{n^2} < \infty$, neboli $\int \sum_{n=1}^{\infty} \frac{X_1^2 \mathbb{1}_{\{|X_1|<n\}}}{n^2}\,\mathrm{d}\mathsf{P} < \infty$.

Poslední integrál omezíme seshora následujícím způsobem:

$$\begin{aligned}
\int_\Omega \sum_{n=1}^{\infty} \frac{X_1^2(\omega)\mathbb{1}_{\{|X_1(\omega)|<n\}}}{n^2}\,\mathrm{d}\mathsf{P}(\omega) &= \sum_{j=0}^{\infty} \int_\Omega \sum_{n=1}^{\infty} \frac{1}{n^2} X_1^2(\omega)\, \mathbb{1}_{\{|X_1(\omega)|<n\}} \mathbb{1}_{\{j \leq |X_1(\omega)| < j+1\}}\,\mathrm{d}\mathsf{P}(\omega) \\
&= \sum_{j=0}^{\infty} \int_\Omega \sum_{n=j+1}^{\infty} \frac{1}{n^2} X_1^2(\omega)\, \mathbb{1}_{\{|X_1(\omega)|<n\}} \mathbb{1}_{\{j \leq |X_1(\omega)| < j+1\}}\,\mathrm{d}\mathsf{P}(\omega) \\
&\leq \sum_{j=0}^{\infty} (j+1)^2 \int_\Omega \sum_{n=j+1}^{\infty} \frac{1}{n^2} \mathbb{1}_{\{|X_1(\omega)|<n\}} \mathbb{1}_{\{j \leq |X_1(\omega)| < j+1\}}\,\mathrm{d}\mathsf{P}(\omega) \\
&\leq 2\sum_{j=0}^{\infty} (j+1)\mathsf{P}(j \leq |X_1| < j+1) = 2\sum_{j=0}^{\infty} \mathsf{P}(|X_1| \geq j) \\
&\leq 2(\mathsf{E}\,|X_1| + 1) < \infty.
\end{aligned}$$

Použili jsme nerovnosti

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

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

Dokažme nyní " $\Rightarrow$ ".
Můžeme psát $X_n = n\overline{X}_n - (n-1)\overline{X}_{n-1}$, takže

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

z předpokladu věty. To ovšem implikuje

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

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

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

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

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

- $\mathsf{E}\,|X_1| < \infty \quad \overset{\text{SZVČ}}{\Longrightarrow} \quad \overline{X}_n \xrightarrow[n\to\infty]{\text{sj}} \mu$
- $\operatorname{var} X_1 = \sigma^2 < \infty \quad \overset{\text{SZVČ}}{\Longrightarrow} \quad \overline{X_n^2} = \frac{1}{n}\sum_{i=1}^{n} X_i^2 \xrightarrow[n\to\infty]{\text{sj}} (\sigma^2 + \mu^2) = \mathsf{E}\,X_1^2$

A tedy

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

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

A také víme:

(i) $(\overline{X}_n - \mu) \xrightarrow[n\to\infty]{\text{sj}} 0$ ze SZVČ.

(ii) Z Čebyševovy nerovnosti $\mathsf{P}(|\overline{X}_n - \mu| > \epsilon) \leq \frac{\operatorname{var} \overline{X}_n}{\epsilon^2} = \frac{\sigma^2}{n\epsilon^2}$, neboli „pravděpodobnost velké odchylky je nejvýše řádu $\frac{1}{n}$."

(iii) Prostým výpočtem $\mathsf{E}\left(\sqrt{n}\,\frac{\overline{X}_n-\mu}{\sigma}\right) = 0$, a $\operatorname{var}\left(\sqrt{n}\,\frac{\overline{X}_n-\mu}{\sigma}\right) = 1$.

(iv) Ze cvičení, že nejen $(\overline{X}_n - \mu) \xrightarrow[n\to\infty]{\text{sj}} 0$, ale i $n^q(\overline{X}_n - \mu) \xrightarrow[n\to\infty]{\mathsf{P}} 0$, pro $q \in (0, \frac{1}{2})$.

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

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

Zde končí předn. 17 (6.4.)

4.5 Centrální limitní věta

Centrální limitní věta (CLV) mluví o tom, že správně znormované odchylky výběrového průměru od střední hodnoty $\sqrt{n}\left(\frac{\overline{X}_n-\mu}{\sigma}\right)$ 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 a buď Y také náhodná veličina. Řekneme, že náhodné veličiny Y_n *konvergují v distribuci* k náhodné veličině Y, pokud $\lim_{n\to\infty} F_{Y_n}(x) = F_Y(x)$ v každém bodě $x \in \mathbb{R}$, ve kterém je F_Y spojitá. Značíme $Y_n \xrightarrow[n\to\infty]{\mathsf{d}} Y$.

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 spojité, pak konvergence v distribuci požaduje, aby $\lim_{n\to\infty} F_{Y_n}(x) = F_Y(x)$ pro všechna $x \in \mathbb{R}$. 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 $\mathsf{P}_Y(\{x\}) > 0$), tak v těchto bodech nemusí platit $\lim_{n\to\infty} F_{Y_n}(x) = F_Y(x)$.

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 spojité rozdělení. A platí $Y_n \xrightarrow[n\to\infty]{\mathsf{d}} Y$, pro Y s $\mathsf{P}_Y = \delta_0$. Tedy $\mathsf{P}(Y = 0) = 1$, 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 $\mathsf{P}_{Y_n} = \delta_{a_n}$. Pak $Y_n \xrightarrow[n\to\infty]{\mathrm{d}} Y$, kde $\mathsf{P}_Y = \delta_a$, a v bodě a nemusí platit $\lim\limits_{n\to\infty} F_{Y_n}(a) = F_Y(a)$. Ověřte.

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

Poznámka. Konvergence $Y_n \xrightarrow[n\to\infty]{\mathrm{d}} Y$ odpovídá takzvané slabé konvergenci pravděpodobnostních měr $\mathsf{P}_{Y_n} \xrightarrow[n\to\infty]{\mathrm{w}} \mathsf{P}_Y$, resp. slabé* konvergenci se kterou se setkáte ve funkcionální analýze.

Značení. Budeme psát například i $Y_n \xrightarrow[n\to\infty]{\mathrm{d}} N(0,1)$, a budeme tím mínit $Y_n \xrightarrow[n\to\infty]{\mathrm{d}} Y$ 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 $Y_n \xrightarrow[n\to\infty]{\mathrm{d}} Y$ právě tehdy, když $\mathsf{E}\, h(Y_n) \xrightarrow[n\to\infty]{} \mathsf{E}\, h(Y)$ pro každou spojitou a omezenou funkci $h : \mathbb{R} \to \mathbb{R}$.

Důkaz: Důkaz je technický a vynecháme jej z časových důvodů. Lze najít v Georgii (2013), kapitola 5.3. nebo Lachout (2004). □

Poznámka. Na ověření $Y_n \xrightarrow[n\to\infty]{\mathrm{d}} Y$ stačí ověřit $\mathsf{E}\, h(Y_n) \xrightarrow[n\to\infty]{} \mathsf{E}\, h(Y)$ i pro menší třídu funkcí než jsou spojité 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 $X_1, X_2, \ldots$ vzájemně nezávislé, stejně rozdělené náhodné veličiny, $\mathsf{E}\, X_i = \mu \in \mathbb{R}$, $\operatorname{var} X_i = \sigma^2 \in (0, \infty)$, $\forall i \in \mathbb{N}$. Potom pro náhodné veličiny

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

platí $Z_n \xrightarrow[n\to\infty]{\mathrm{d}} Z$, 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 $Z_n \xrightarrow[n\to\infty]{\mathsf{d}} Z$, 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 $X_i \sim Cauchy$, 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
$$f_{Z_n}(z) = \frac{\sqrt{n}}{\pi(n+z^2)}, \qquad z \in \mathbb{R},$$
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 $\mathsf{E}\, X_i = 0$ a $\operatorname{var} X_i = 1$. 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 $\mathsf{E}\, h(Z_n) \xrightarrow[n\to\infty]{} \mathsf{E}\, h(Z)$ 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 jsou všechny nezávislé na $\{X_i\}_{i=1}^\infty$ a které mají všechny rozdělení $N(0,1)$. Pak $\mathsf{E}\, Y_i = 0$, $\operatorname{var} Y_i = 1$, a $T_n = \frac{\sum_{i=1}^n Y_i}{\sqrt{n}} \sim N(0,1)$ (viz sekce 3.5.2). Takže chceme dokázat
$$|\mathsf{E}\,(h(Z_n) - h(T_n))| \xrightarrow[n\to\infty]{} 0. \tag{4.11}$$
K tomu použijeme vyjádření $h(Z_n) - h(T_n)$ jako tzv. teleskopickou sumu. Nejdříve si zaveďme označení
$$X_{i,n} = \frac{X_i}{\sqrt{n}}, \qquad Y_{i,n} = \frac{Y_i}{\sqrt{n}}, \qquad W_{i,n} = \sum_{j=1}^{i-1} Y_{j,n} + \sum_{j=i+1}^{n} X_{j,n}.$$

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

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

neboť

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

Náhodné veličiny $X_{i,n}$ a $Y_{i,n}$ jsou malé a funkce 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

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

kde

$$R_{X_{i,n}}(\omega) = \frac{1}{2}\, X_{i,n}^2(\omega)\, \Big[h''(W_{i,n}(\omega) + \zeta(\omega) X_{i,n}(\omega)) - h''(W_{i,n}(\omega))\Big],$$
$$\text{pro nějaké } \ \zeta(\omega) \in [0,1], \ \ \omega \in \Omega.$$

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

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

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

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

Dohromady tedy můžeme odhadnout shora

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

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

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

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

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

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 první rovnost. Druhá rovnost plyne z linearity střední hodnoty a třetí rovnost plyne z rovnosti momentů X_i a Y_i.

Nyní použijeme (4.14), takže

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

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

$$\mathsf{E}\,\left(X_1^2 \mathbb{1}_{(|X_1|>\delta\sqrt{n})}\right) = \mathsf{E}\,X_1^2 - \mathsf{E}\,\left(X_1^2 \mathbb{1}_{(|X_1|\leq\delta\sqrt{n})}\right) \xrightarrow[n\to\infty]{} \mathsf{E}\,X_1^2 - \mathsf{E}\,X_1^2 = 1 - 1 = 0,$$

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

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

což implikuje (4.11) a tím je věta dokázána. □

Zde končí předn. 18 (13.4.)

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 bernoulliovských pokusů. V tomto případě si odpověď neformulujeme jako příklad, ale jako speciální centrální limitní větu.

Věta 4.14 (de Moivreova–Laplaceova centrální limitní věta) Buďte $\{Y_n\}_{n=1}^\infty$ náhodné veličiny splňující $Y_n \sim Binom(n, p)$, $0 < p < 1$. Pak

$$\frac{Y_n - np}{\sqrt{np(1-p)}} \xrightarrow[n\to\infty]{\mathsf{d}} N(0, 1). \tag{4.15}$$

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

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

Ovšem $V_n = \sum_{i=1}^n X_i \sim Binom(n, p)$, a tedy i $\frac{V_n - np}{\sqrt{np(1-p)}} \xrightarrow[n\to\infty]{\mathsf{d}} N(0, 1)$. A protože konvergence v distribuci požaduje jen konvergenci distribučních funkcí odpovídajících náhodných veličin, musí konvergence (4.15) platit pro jakékoli náhodné veličiny $\{Y_n\}_{n=1}^\infty$ splňující $Y_n \sim Binom(n, p)$, nejen pro $\{V_n\}_{n=1}^\infty$. □

Poznámka. CLV pro binomické rozdělení, resp. pro bernoulliovskou posloupnost byla dokázána o mnoho dříve (A. de Moivre 1733, P.S. Laplace 1812) než obecná CLV (A.M. Ljapunov 1901, J.W. Lindenberg 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-Esseénovu nerovnost, která tvrdí (ve značení CLV 4.13)

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

Odhaduje tedy shora rychlost konvergence rozdílu distribučních funkcí zúčastněných v CLV a ukazuje, že řádově odpovídá $\frac{1}{\sqrt{n}}$ a závisí na normovaném třetím momentu $\frac{\mathsf{E}\,(|X_1 - \mathsf{E}\,(X_1)|^3)}{\sigma^3}$, 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 $\mathsf{P}(|\frac{1}{100}\sum_{i=1}^{100} X_i - \frac{1}{2}| > \frac{1}{10})$. Ovšem $Y_n = \sum_{i=1}^n X_i$ má binomické rozdělení s parametry $p = \frac{1}{2} \in (0, 1)$ a n, tedy splňuje předpoklady CVL pro binomické rozdělení 4.14 a podle této věty platí

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

Neboli $\lim\limits_{n\to\infty} F_{Z_n}(x) = \Phi(x)$. Přepišme

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

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

$$\mathsf{P}\left(\left|\frac{1}{100}\sum_{i=1}^{100} X_i - \frac{1}{2}\right| > \frac{1}{10}\right) \doteq 1 - \Phi(2) + \Phi(-2) = 2\Phi(-2) \doteq 0.046. \tag{4.16}$$

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

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

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

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

tedy na tuto přesnost bychom potřebovali mnohem více pokusů než 1000. Pro $n = 10000$ pokusů bychom opět dostali $\mathsf{P}(|\overline{X}_{10000} - \frac{1}{2}| > \frac{1}{100}) \doteq 0.046$, jako v (4.16). Neboli tento příklad také ilustruje, že odchylka empirického průměru od p, resp. $\mathsf{E}\, X_1$ se opravdu zmenšuje s řádem $\frac{1}{\sqrt{n}}$.

Samozřejmě lze odpovídat i na otázky „v jakém rozmezí se nachází rozdíl mezi $\overline{X}_n$ a p po $n = 100$ pokusech s pravděpodobností 0.99“ nebo „jak velké musí být n, aby odchylka mezi $\overline{X}_n$ 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 Moivreovu-Laplaceovu CLV říká, že aproximace pro binomické rozdělení je prakticky použitelná pokud $np(1-p) \geq 9$ (viz např. Dupač a 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 spojité transformaci?

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

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

Poznámka. Lemma využijeme ve statistice pro konzistentní odhady.

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

$$\frac{\sum_{i=1}^{n}(X_i - \mu)}{\sqrt{n\,\sigma^2}} \xrightarrow[n\to\infty]{\mathsf{d}} N(0,1). \tag{4.17}$$

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

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

A co víme o $(\overline{X_n - \mu})^2$? Aplikací lemmatu 4.15 o spojité transformaci pro konvergenci v distribuci na posloupnost z (4.17) a funkci $g(x) = x^2$ dostaneme

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

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 Buď $\{X_n\}_{n=1}^{\infty}$ posloupnost náhodných veličin definovaných na stejném $(\Omega, \mathcal{A}, \mathsf{P})$. Pak $(X_n - X) \xrightarrow[n\to\infty]{\mathsf{P}} 0 \Rightarrow X_n \xrightarrow[n\to\infty]{\mathsf{d}} X$.

Důkaz: Neuvádíme, viz např. Lachout (2004). □

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

Poznámka. Obrácená implikace obecně neplatí (přesněji – platí jen když je X konstantní s.j. – viz např. Lachout (2004)). Uvažujme $(\Omega, \mathcal{A}, \mathsf{P}) = ([0,1], \mathcal{B}([0,1]), \lambda)$ a $X_n = \mathbb{1}_{[0,\frac{1}{2}]}$ pro n sudé, $X_n = \mathbb{1}_{[\frac{1}{2},1]}$ pro n liché. Platí $\mathsf{P}(X_n = 0) = \mathsf{P}(X_n = 1) = \frac{1}{2}$, $\forall n \in \mathbb{N}$, takže $X_n \xrightarrow[n\to\infty]{\mathsf{d}} X$, kde $X \sim Alt(\frac{1}{2})$. 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) Buďte $\{X_n\}_{n=1}^{\infty}$, $\{Y_n\}_{n=1}^{\infty}$, X náhodné veličiny a nechť $X_n \xrightarrow[n\to\infty]{\mathsf{d}} X$ a $Y_n \xrightarrow[n\to\infty]{\mathsf{P}} a \in \mathbb{R}$. Potom $(X_n + Y_n) \xrightarrow[n\to\infty]{\mathsf{d}} (X + a)$ a $X_n \cdot Y_n \xrightarrow[n\to\infty]{\mathsf{d}} a\,X$.

důkaz nebyl v LS 22/23 přednášen

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 $X_n \cdot Y_n \xrightarrow[n\to\infty]{\mathsf{d}} a\,X$, pro případ $a > 0$ (případy $a < 0$ a $a = 0$ se dokazují podobně).

Buď tedy $a > 0$ a nechť $Y_n \xrightarrow[n\to\infty]{\mathsf{P}} a \in \mathbb{R}$. Chceme ukázat, že platí $\lim_{n\to\infty} \mathsf{P}(X_n \cdot Y_n \le x) = \mathsf{P}(Xa \le x) = \Phi(\frac{x}{a})$. Buď $x > 0$. Odhadneme množinu $\{\omega : X_n(\omega)\cdot Y_n(\omega) \le x\}$ zespoda i seshora. Buď $\epsilon > 0$, ale malé tak, aby $(a-\epsilon) > 0$. Pak platí:

$$\begin{aligned}
&\{X_n(a-\epsilon) \le x\} \cup \{|Y_n - a| > \epsilon\} \\
&\supseteq \{\omega : X_n(\omega)\cdot Y_n(\omega) \le x\} = \{X_n \cdot Y_n \le x, |Y_n - a| \le \epsilon\} \cup \{X_n \cdot Y_n \le x, |Y_n - a| > \epsilon\} \\
&\supseteq \{X_n \cdot Y_n \le x, |Y_n - a| \le \epsilon\} \supseteq \{X_n(a+\epsilon) \le x, |Y_n - a| \le \epsilon\} \\
&\supseteq \{X_n(a+\epsilon) \le x\} \setminus \{|Y_n - a| > \epsilon\}.
\end{aligned}$$

Z toho

$$\mathsf{P}\left(X_n \le \frac{x}{a+\epsilon}\right) - \mathsf{P}(|Y_n - a| > \epsilon) \le \mathsf{P}(X_n \cdot Y_n \le x) \le \mathsf{P}\left(X_n \le \frac{x}{a-\epsilon}\right) + \mathsf{P}(|Y_n - a| > \epsilon).$$

Spočítejme nyní limity pro $n \to \infty$. Dostaneme:

$$\Phi\left(\frac{x}{a+\epsilon}\right) - 0 \le \liminf_{n\to\infty} \mathsf{P}(X_n \cdot Y_n \le x) \le \limsup_{n\to\infty} \mathsf{P}(X_n \cdot Y_n \le x) \le \Phi\left(\frac{x}{a-\epsilon}\right) + 0,$$

z předpokladů věty. Pokud teď spočítáme limitu pro $\epsilon \to 0$, dostaneme ze spojitosti Φ (je spojitá všude)

$$\lim_{n\to\infty} \mathsf{P}(X_n \cdot Y_n \le x) = \Phi\left(\frac{x}{a}\right) = \mathsf{P}(aX \le x). \tag{4.18}$$

Pro $x < 0$ postupujeme obdobně a dostaneme se k inkluzím

$$\{X_n(a+\epsilon) \le x\} \cup \{|Y_n - a| > \epsilon\} \supseteq \{\omega : X_n(\omega)\cdot Y_n(\omega) \le x\} \supseteq \{X_n(a-\epsilon) \le x\} \setminus \{|Y_n - a| > \epsilon\},$$

a tedy nerovnostem

$$\mathsf{P}\left(X_n \le \frac{x}{a-\epsilon}\right) - \mathsf{P}(|Y_n - a| > \epsilon) \le \mathsf{P}(X_n \cdot Y_n \le x) \le \mathsf{P}\left(X_n \le \frac{x}{a+\epsilon}\right) + \mathsf{P}(|Y_n - a| > \epsilon).$$

Limitními předchody obdržíme opět (4.18). 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 Cramérovy–Sluckého věty? Pro posloupnost nezávislých stejně rozdělených náhodných veličin $\{X_i\}_{i=1}^{\infty}$ s $\mathsf{E}\, X_i = \mu \in \mathbb{R}$ a $\operatorname{var} X_i = \sigma^2 \in \mathbb{R}^+$ máme z CLV

$$Z_n = \frac{\sum_{i=1}^{n} X_i - n\mu}{\sqrt{n\sigma^2}} \xrightarrow[n\to\infty]{\mathsf{d}} N(0,1).$$

Představme si, že neznámou střední hodnotu μ chceme odhadnout právě z pozorování posloupnosti $\{X_i\}_{i=1}^{\infty}$. Pak máme problém, že na levé straně konvergence se vyskytuje také hodnota σ^2, kterou rovněž neznáme. Pokud ale najdeme nějakou posloupnost náhodných veličin $\{Y_i\}_{i=1}^{\infty}$ takovou, že $Y_n \xrightarrow[n\to\infty]{\mathsf{P}} \sigma^2$, pak z lemmatu o spojité transformaci pro konvergenci v pravděpodobnosti dostaneme $\frac{Y_n}{\sigma^2} \xrightarrow[n\to\infty]{\mathsf{P}} 1$ a také $U_n = \sqrt{\frac{\sigma^2}{Y_n}} \xrightarrow[n\to\infty]{\mathsf{P}} 1$.

Takže $Z_n \xrightarrow[n\to\infty]{\mathsf{d}} N(0,1)$ a $U_n \xrightarrow[n\to\infty]{\mathsf{d}} 1$, a z Cramérovy–Sluckého věty dostanu

$$Z_n \cdot U_n = \frac{\sum_{i=1}^{n} X_i - n\mu}{\sqrt{n}\sqrt{Y_n}} \xrightarrow[n\to\infty]{\mathsf{d}} N(0,1).$$

Nyní už je na levé straně neznámé jen to, co chci statisticky odhadnout. *Zde končí předn. 19 (19.4.)*

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 těchto parametrech a nakonec i posoudit shodu modelu se skutečností (nebo přesněji, ověřit, že model není v rozporu se skutečností).

Definice 5.1 Nechť $n \in \mathbb{N}$ a $X_1, \dots, X_n$ 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 pravděpodobnostních 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 $\mathcal{F} = \{\mathsf{P}_X \text{ taková, že } \mathsf{E}\, X \in \mathbb{R}\}$. Pro pozorovaná data můžeme spočítat empirickou distribuční funkci $\widehat{F}_n$ (viz následující definice), a použít ji jako náš odhad neznámé distribuční funkce F. Glivenkova-Cantelliho věta (důkaz viz např. Kallenberg (2002) nebo proseminář) tvrdí, že

$$\lim_{n\to\infty} \sup_{x\in\mathbb{R}} |\widehat{F}_n(x) - F(x)| = 0, \qquad \text{s.j.},$$

tedy se zvětšujícím se rozsahem výběru bude náš odhad $\widehat{F}_n$ konvergovat k neznámé F skoro jistě. To je chování, které by rozumný odhad měl splňovat.

Definice 5.3 Buď $X_1, \ldots, X_n$ náhodný výběr. *Empirická distribuční funkce* je definována předpisem

$$\widehat{F}_n(x, \omega) = \frac{1}{n} \sum_{k=1}^{n} \mathbb{1}_{(X_k \leq x)}(\omega) = \frac{1}{n} \sum_{k=1}^{n} \mathbb{1}_{(X_k(\omega) \leq x)}, \qquad x \in \mathbb{R}, \omega \in \Omega.$$

Druhý možný přístup statistického modelování je parametrický. O neznámém rozdělení P_X, resp. distribuční funkci F, máme poměrně přesnou představu a neznámých je jen několik (málo) parametrů. Neboli neznámá $F = F_{\theta_0}$ patří do třídy distribučních funkcí $\mathcal{F} = \{\mathcal{F}_\theta, \theta \in \Theta\}$, kde $\Theta \subseteq \mathbb{R}^d$, $d \in \mathbb{N}$, je borelovská podmnožina $\mathbb{R}^d$. Θ se nazývá *parametrický prostor* a θ je neznámý (pro $d > 1$ vektorový) parametr. Tímto přístupem se budeme podrobněji zabývat v následujících kapitolách.

Příklad. Parametrický model může být např. $\mathcal{F} = \{N(\mu, \sigma^2) : \mu \in \mathbb{R}, \sigma^2 > 0\} = \{\mathsf{P}_\theta : \theta \in \Theta\}$, s $\theta = (\mu, \sigma^2)$ a $\Theta = \mathbb{R} \times \mathbb{R}^+$ – třída všech normálních rozdělení. Zde $d = 2$ a hledáme neznámé prametry: střední hodnotu μ a rozptyl σ^2. Třeba když opakovaně měříme nějakou fyzikální veličinu s náhodnou chybou. Zajímá nás μ = skutečná hodnota fyzikální veličiny, kterou chceme změřit, a σ^2 = přesnost přístroje, kterým ji měříme.

Nebo může být $\mathcal{F} = \{Alt(p) : p \in (0, 1)\} = \{\mathsf{P}_\theta : \theta \in \Theta\}$ s $\theta = p$ a $\Theta = (0, 1)$ – třída všech alternativních rozdělení s pravděpodobností úspěchu p. Parametr je jednorozměrná pravděpodobnost úspěchu p. Například pozorujeme posloupnost bernoulliovských 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.

5.1 Bodový odhad

Mějme tedy model – třídu parametrických rozdělení $\{\mathsf{P}_\theta : \theta \in \Theta\}$ – a $X_1, \ldots, X_n$ náhodný výběr z P_θ, kdy ovšem konkrétní hodnotu parametru θ neznáme a 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 $g : \Theta \to \mathbb{R}$ nazveme *parametrickou funkcí.*

Na základě náhodného výběru $X_1, \ldots, X_n$ z P_θ tedy chceme odhadnout $g(\theta)$.

Definice 5.5 *Bodový odhad φ_n parametrické funkce $g(\theta)$*, je borelovsky měřitelné zobrazení $\varphi_n : \mathbb{R}^n \to \mathbb{R}$, jehož předpis nezávisí na θ (tedy ani na P_θ či F_θ) a jehož definiční obor obsahuje obor hodnot $(X_1, \ldots, X_n)$.

Příklad. Uvažujme posloupnost hodů mincí. Takový experiment lze tedy modelovat jako bernoulliovskou posloupnost $X_1, X_2, \ldots$. Pozorujeme počáteční úsek $X_1, \ldots, X_n$ a zajímá nás, jestli je mince spravedlivá. Pak $X_1, \ldots, X_n$ je náhodný výběr z alternativního rozdělení s neznámou pravděpodobností úspěchu p. Modelem tedy je $\mathcal{F} = \{Alt(p) : p \in [0,1]\} = \{\mathsf{P}_\theta : \theta \in \Theta\}$ 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 $\varphi_n : \mathbb{R}^n \to \mathbb{R}$, definovanou jako

$$\varphi_n(x_1, \ldots x_n) = \frac{1}{n} \sum_{k=1}^{n} x_k = \overline{x}_n.$$

Uvědomme si, že pro $(x_1, \ldots, x_n) \in \{0,1\}^n$ je $\varphi_n(x_1, \ldots x_n) \in [0,1]$, což chceme.

V příkladu je $\varphi_n : \mathbb{R}^n \to \mathbb{R}$ funkce, pro konkrétní pozorovaná $x_1, \ldots, x_n$ je tedy $\varphi_n(x_1, \ldots, x_n) = \overline{x}_n$ číslo. Ovšem $\varphi_n(X_1, \ldots, X_n) = \overline{X}_n$ 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 $\varphi_n(X_1, \ldots, X_n)$ se nazývají estimator, zatímco číslo $\varphi_n(x_1, \ldots, x_n)$ se nazývá estimate.

Uvědomme si také, že funkce $\varphi_n(x_1, \ldots, x_n) = \overline{x}_n$ (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 $\varphi_n(X_1, \ldots, X_n) = \overline{X}_n$ na hodnotě parametru $\theta = p$ závisí. Konkrétně v příkladu výše má $\varphi_n(X_1, \ldots, X_n)$ rozdělení jako $\frac{1}{n}Z$, kde $Z \sim Binom(n, p)$. Ideálně by rozdělení odhadu $\varphi_n(X_1, \ldots, X_n)$ 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 jen do určité míry.

Značení. V dalším budeme někdy zkráceně značit odhad $\varphi_n(X_1, \ldots, X_n)$ parametru θ na základě náhodného výběru $X_1, \ldots, X_n$ jen symbolem $\widehat{\theta}_n$.

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 definujeme některé žádoucí vlastnosti, které by „dobrý“ odhad měl mít.

Definice 5.6 Bodový odhad φ_n parametrické funkce $g(\theta)$ se nazývá *nestranný*, pokud pro každé $\theta \in \Theta$ platí

$$\mathsf{E}_\theta\, \varphi_n(X_1, \ldots, X_n) = g(\theta).$$

Použili jsme značení E_θ střední hodnoty počítané vzhledem k rozdělení $\mathsf{P}_\theta \otimes \cdots \otimes \mathsf{P}_\theta$, tedy když je $X_1, \ldots, X_n$ 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

$$\mathsf{P}(\lim_{n\to\infty} \varphi_n(X_1, \dots, X_n) = g(\theta)) = 1,$$

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 $\varphi_n(X_1, \dots, X_n) = \overline{X}_n$ je nestranný odhad pravděpodobnosti úspěchu p, neboť

$$\mathsf{E}_p\overline{X}_n = \frac{1}{n}\sum_{k=1}^{n} \mathsf{E}\, X_k = \frac{1}{n}\sum_{k=1}^{n} p = p.$$

A posloupnost odhadů φ_n je silně konzistentní, neboť $\overline{X}_n \xrightarrow[n\to\infty]{\text{sj}} \mathsf{E}_p X_1 = p$ ze SZVČ pro stejně rozdělené náhodné veličiny.

A co vlastně ty vlastnosti nestrannosti a konzistence znamenají? Nestrannost odpovídá tomu, že odhad $\widehat{\theta}_n$ 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

$$\mathsf{E}_\theta\, \varphi_n(X_1, \dots, X_n) \xrightarrow[n\to\infty]{} g(\theta), \qquad \forall\theta \in \Theta.$$

Neboli systematická chyba odhadu φ_n s $n \to \infty$ 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

$$\mathsf{E}_\theta(\varphi_n(X_1, \dots, X_n) - g(\theta)) = 0, \quad \text{nikoli} \quad \mathsf{E}_\theta|\varphi_n(X_1, \dots, X_n) - g(\theta)| = 0,$$

(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

$$\mathsf{E}_\theta(\varphi_n(X_1, \dots, X_n) - g(\theta))^2. \tag{5.1}$$

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 $\operatorname{var}(\varphi_n(X_1,\dots,X_n))$ odhadu φ_n, tedy (5.1) pro $n\to\infty$ popisuje, jak rychle s rostoucím n klesá variabilita (neboli nejistota) odhadu.

Zde končí předn. 20 (20.4)

K diskusi vlastností bodových odhadů se ještě vrátíme. Teď začneme tím, že prozkoumáme dvě základní statistiky.

Definice 5.8 *Statistikou* nazveme libovolnou borelovsky měřitelnou funkci $T_n:\mathbb{R}^n\to\mathbb{R}$, jejíž definiční obor obsahuje 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

$$\overline{X}_n=\frac{1}{n}\sum_{k=1}^n X_k$$

se nazývá *výběrový průměr.* A statistika

$$S_n^2=\frac{1}{n-1}\sum_{k=1}^n(X_k-\overline{X}_n)^2$$

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 $\mu=\mathsf{E}\,X\in\mathbb{R}$ a rozptylem $\sigma^2=\operatorname{var}X<\infty$. Pak výběrový průměr je nestranný a konzistentní odhad $\mu=\mathsf{E}\,X$ a výběrový rozptyl je nestranný a konzistentní odhad rozptylu $\sigma^2=\operatorname{var}X$.

Důkaz: Z věty 3.14 máme nestrannost $\overline{X}_n$, a protože předpokládáme $\mu\in\mathbb{R}$, tak ze SZVČ pro stejně rozdělené náhodné veličiny máme $\overline{X}_n\xrightarrow[n\to\infty]{\text{sj}}\mathsf{E}\,X=\mu$, tedy $\overline{X}_n$ je silně konzistentní.

Pro S_n^2 si nejdříve upravíme

$$\begin{aligned}S_n^2&=\frac{1}{n-1}\sum_{k=1}^n(X_k-\mu-(\overline{X}_n-\mu))^2\\&=\frac{1}{n-1}\sum_{k=1}^n(X_k-\mu)^2-2\,\frac{1}{n-1}\sum_{k=1}^n(X_k-\mu)(\overline{X}_n-\mu)+\frac{n}{n-1}\,(\overline{X}_n-\mu)^2\\&=\frac{1}{n-1}\sum_{k=1}^n(X_k-\mu)^2-\frac{n}{n-1}\,(\overline{X}_n-\mu)^2.\end{aligned}\tag{5.2}$$

Tedy

$$\mathsf{E}_{(\mu,\sigma^2)}S_n^2=\frac{1}{n-1}\sum_{k=1}^n\mathsf{E}_{(\mu,\sigma^2)}(X_k-\mu)^2-\frac{n}{n-1}\,\mathsf{E}_{(\mu,\sigma^2)}(\overline{X}_n-\mu)^2=\frac{n}{n-1}\,\sigma^2-\frac{n}{n-1}\frac{1}{n}\,\sigma^2=\sigma^2.$$

V předposlední rovnosti jsme využili, že $\mathsf{E}_{(\mu,\sigma^2)}(\overline{X}_n - \mu)^2 = \mathsf{var}_{(\mu,\sigma^2)}\overline{X}_n = \frac{1}{n}\,\mathsf{var}_{(\mu,\sigma^2)}X_1$. Takže S_n^2 je nestranný odhad σ^2.

Abychom ověřili konzistenci, musíme prozkoumat limity obou členů z (5.2). Neboť $\mathsf{var}\,X_k \in \mathbb{R}$, a $(X_k - \mu)^2$ jsou vzájemně nezávislé náhodné veličiny, tak ze SZVČ pro stejně rozdělené náhodné veličiny máme

$$\frac{1}{n}\sum_{k=1}^{n}(X_k - \mu)^2 \xrightarrow[n\to\infty]{\text{sj}} \mathsf{E}_{(\mu,\sigma^2)}(X_k - \mu)^2 = \sigma^2.$$

Protože $\frac{n}{n-1} \xrightarrow[n\to\infty]{} 1$, tak i $\frac{1}{n-1}\sum_{k=1}^{n}(X_k - \mu)^2 \xrightarrow[n\to\infty]{\text{sj}} \sigma^2$, 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 $(\overline{X}_n - \mu) \xrightarrow[n\to\infty]{\text{sj}} 0$, a z věty 4.6 o spojité transformaci pro konvergenci skoro jistě máme $(\overline{X}_n - \mu)^2 \xrightarrow[n\to\infty]{\text{sj}} 0$. Protože $\frac{n}{n-1} \xrightarrow[n\to\infty]{} 1$, dostáváme dohromady $\frac{n}{n-1}(\overline{X}_n - \mu)^2 \xrightarrow[n\to\infty]{\text{sj}} 0$, což zbývalo dokázat. □

Máme tedy rozumné odhady pro momenty $\mathsf{E}\,X$ a $\mathsf{var}\,X$. Ale jak najít rozumné (případně vůbec nějaké) odhady parametrů v modelu $\{\mathsf{P}_\theta : \theta \in \Theta\}$? Ukážeme si dvě metody.

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 $\mathcal{F} = \{\mathsf{P}_\theta : \theta \in \Theta\}$ na hodnotě parametru typicky nějakým jednoduchým způsobem závisejí. 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 $X_1, \ldots, X_n$ náhodný výběr z rozdělení P_θ. Označme si $m_r(\theta) = \mathsf{E}\,X^r$, $r \in \mathbb{N}$, r-tý moment náhodné veličiny X s rozdělením P_θ (pokud existuje). A označme

$$\widehat{m_r} = \frac{1}{n}\sum_{k=1}^{n} X_k^r,$$

r-tý výběrový moment spočítaný z náhodného výběru $X_1, \ldots, X_n$.

Nechť pro $X \sim \mathsf{P}_\theta$ platí $\mathsf{E}\,|X|^r < \infty$. Náhodný výběr z P_θ jsou nezávislé, stejně rozdělené náhodné veličiny s rozdělením P_θ, tedy zřejmě platí

$$\mathsf{E}\,\widehat{m_r} = \frac{1}{n}\sum_{k=1}^{n}\mathsf{E}\,X_k^r = \frac{n}{n}\mathsf{E}\,X^r = \mathsf{E}\,X^r,$$

neboť $\mathsf{E}\,X^r \in \mathbb{R}$. 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 $\mathsf{E}\,X_k^r \in \mathbb{R}$. Tedy splňují předpoklady SZVČ 4.10 pro stejně rozdělené náhodné veličiny a posloupnost odhadů $\widehat{m_r}$ konverguje s rostoucím rozsahem výběru n skoro jistě k $m_r(\theta) = \mathsf{E}\,X^r$. Takže jsme dokázali následující tvrzení:

Tvrzení 5.2 Pokud pro náhodnou veličinu X s rozdělením P_θ platí $\mathsf{E}\,|X|^r < \infty$, pak r-tý výběrový moment $\widehat{m_r}$ je nestranný a silně konzistentní odhad $m_r(\theta) = \mathsf{E}\,X^r$.

Tedy výběrové momenty jsou „rozumnými“ bodovými odhady teoretických momentů rozdělení $X \sim \mathsf{P}_\theta$.

Momentový odhad $\hat{\theta}$ parametru θ najdeme jako řešení soustavy momentových odhadovacích rovnic

$$m_r(\theta) = \widehat{m_r}, \qquad r = 1, 2, \ldots, d.$$

Zde d je typicky dimenze θ, a nebo kolik je potřeba, abychom mohli soutavu vyřešit.

Příklad. Buď $\theta \in \mathbb{R}$, tedy neznámý parametr je jednorozměrný, a buď funkce $m_1(\theta)$ prostá. Pak stačí použít jednu momentovou odhadovací rovnici $m_1(\hat{\theta}) = \widehat{m_1}$, a $\hat{\theta} = m_1^{-1}(\widehat{m_1})$ je momentový odhad parametru θ (pokud existuje).

Příklad. Buď model $\mathcal{F} = \{\mathsf{P}_\theta, \theta \in \Theta\} = \{Rovnom[0, a], a \in \mathbb{R}^+\}$, neboli $\theta = a$ a $\Theta = \mathbb{R}^+$. Výpočtem zjistíme, že pro $X \sim Rovnom[0, a]$ je $\mathsf{E}\,X = \frac{a}{2}$, tedy momentová odhadovací rovnice je

$$\frac{a}{2} = \overline{X}_n.$$

Jejím vyřešením zjistíme $\widehat{a_n} = 2\overline{X}_n$ (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 $a = 2\mathsf{E}\,X$ (což je ta samá lineární funkce střední hodnoty $\mathsf{E}\,X$, kterou $\overline{X}_n$ nestraně a konzistentně odhaduje (věta 5.1)).

Je to vždy tak jednoduché? Ne nutně.

Příklad. Buď model $\mathcal{F} = \{\mathsf{P}_\theta, \theta \in \Theta\} = \{Rovnom[-a, a], a \in \mathbb{R}^+\}$. Tedy opět $\theta = a$ a $\Theta = \mathbb{R}^+$.

Ovšem $\mathsf{E}\,X = 0$ pro $X \sim Rovnom[-a, a]$, tedy první moment nelze pro momentovou odhadovací rovnici použít. Heuristicky: $\mathsf{E}_a X$ na hodnotě parametru $\theta = a$ nezávisí, a tedy ho nelze využít k odhadu θ. Ovšem pozor! Neznamená to, že rozdělení $\overline{X}_n$ 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:

$$\mathsf{E}_a X^2 = \int_{-a}^{a} \frac{x^2}{2a}\, dx = 2\int_0^a \frac{x^2}{2a}\, dx = \left[\frac{1}{a}\,\frac{x^3}{3}\right]_0^a = \frac{a^2}{3}. \tag{5.3}$$

Momentová odhadovací rovnice založená na druhém momentu tedy bude

$$\frac{a^2}{3} = \overline{X_n^2} = \frac{1}{n}\sum_{i=1}^{n} X_i^2,$$

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 $\overline{X_n^2} \xrightarrow[n\to\infty]{\text{sj}} \frac{a^2}{3}$, $\forall a \in \mathbb{R}^+$. A z věty 4.6 o spojité transformaci pro konvergenci skoro jistě dostaneme $\widehat{a_n} = \sqrt{3\,\overline{X_n^2}} \xrightarrow[n\to\infty]{\text{sj}} a$, $\forall a \in \mathbb{R}^+$, tedy $\widehat{a_n}$ je silně konzistentní odhad parametru a.

Z tvrzení 5.2 víme $\mathsf{E}\,\overline{X_n^2} = \frac{a^2}{3}$, ovšem $\mathsf{E}\sqrt{3\,\overline{X_n^2}} \neq a$ (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 $\mathsf{E}\,(\sqrt{3\,\overline{X_n^2}} - a)$, 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 $\mathsf{E}\,X = \mu$ a $\mathsf{E}\,X^2 = \mu^2 + \sigma^2$. Momentové odhadovací rovnice tedy jsou

$$\mu = \overline{X}_n \qquad \sigma^2 + \mu^2 = \overline{X_n^2},$$

jejich vyřešením dostaneme

$$\widehat{\mu}_n = \overline{X}_n, \qquad \widehat{\sigma_n^2} = \overline{X_n^2} - (\overline{X}_n)^2 = \frac{\sum_{i=1}^{n}(X_i - \overline{X}_n)^2}{n}.$$

Už víme, že $\widehat{\mu}_n$ je silně konzistentní a nestranný odhad $\mu = \mathsf{E}\,X$ (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 spojité transformaci pro $\xrightarrow[n\to\infty]{\text{sj}}$ 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.

Zde končí přednáška 21 (26.4.)

5.1.2 Metoda maximální věrohodnosti

Buď $X_1, \dots, X_n$ náhodný výběr z rozdělení P_θ. A nechť 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 spojité rozdělení P_θ je $\nu = \lambda$, pro diskrétní rozdělení P_θ je ν vhodná čítací 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 ale definujme, co je to věrohodnost.

Definice 5.10 Buď $\mathcal{F} = \{\mathsf{P}_\theta, \theta \in \Theta\}$ 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

$$L(\theta; \mathbf{x}) = f_\theta(x_1, \dots x_n) = \prod_{i=1}^n f_\theta(x_i), \quad \mathbf{x} = (x_1, \dots x_n) \in \mathbb{R}^n, \ \theta \in \Theta.$$

$l(\theta; \mathbf{x}) = \log(L(\theta; \mathbf{x}))$ se nazývá *logaritmickou věrohodností.*

Buď $\mathcal{F} = \{\mathsf{P}_\theta, \theta \in \Theta\}$ 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 konktrétní 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 $X \sim \mathsf{P}_{\theta_0}$ než když $X \sim \mathsf{P}_{\theta_1}$. 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

$$\frac{\partial}{\partial \theta_j} \sum_{i=1}^{n} \log f_\theta(x_i) = 0, \qquad j = 1, \dots d, \quad \text{pro } \theta \in \Theta \subseteq \mathbb{R}^d,$$

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 spojité, tedy referenční míra $\nu = \lambda$ a

$$f_\theta(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \, \mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}}, \qquad x \in \mathbb{R},$$

takže

$$L(\theta; \mathbf{x}) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \, \mathrm{e}^{-\frac{(x_i-\mu)^2}{2\sigma^2}}$$

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

$$\frac{\partial l(\theta; \mathbf{x})}{\partial \mu} = \frac{2 \sum_{i=1}^{n} (x_i - \mu)}{2\sigma^2}, \qquad \frac{\partial l(\theta; \mathbf{x})}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{\sum_{i=1}^{n} (x_i - \mu)^2}{2(\sigma^2)^2},$$

takže věrohodnostní rovnice jsou

$$\sum_{i=1}^{n} x_i - n\,\mu = 0 \qquad \sum_{i=1}^{n} (x_i - \mu)^2 - n\,\sigma^2 = 0.$$

Vyřešením získáme $\widehat{\mu}_n = \overline{X}_n$, $\widehat{\sigma^2_n} = \frac{\sum_{i=1}^{n} (X_i - \overline{X}_n)^2}{n}$ (kdybychom dopočetli matici druhých parciálních derivací $l(\theta; \mathbf{x})$ a dosadili $(\widehat{\mu}_n, \widehat{\sigma^2_n})$, zjistili bychom, že $l(\theta; \mathbf{x})$ opravdu nabývá v $(\widehat{\mu}_n, \widehat{\sigma^2_n})$ svého maxima). Získali jsme tedy stejné odhady jako metodou momentů. Takže oba odhady jsou konzistentní a $\widehat{\mu}_n$ 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 $\mathcal{F} = \{\mathsf{P}_\theta, \theta \in \Theta\} = \{Rovnom[0, a], a \in \mathbb{R}^+\}$, neboli $\theta = a$ a $\Theta = \mathbb{R}^+$. Rovnoměrné rozdělení je absolutně spojité, tedy referenční míra $\nu = \lambda$ a

$$f_a(x) = \frac{1}{a} \mathbb{1}_{[0,a]}(x), \qquad x \in \mathbb{R}.$$

Tedy věrohodnost je

$$L(a;\mathbf{x}) = \prod_{i=1}^{n} \frac{1}{a} \mathbb{1}_{[0,a]}(x_i) = \frac{1}{a^n} \mathbb{1}\,(\max_{i=1,\dots,n} x_i \leq a) \mathbb{1}(0 \leq \min_{i=1,\dots,n} x_i).$$

Ta je zřejmě maximalizována pro $\max_{i=1,\dots,n} x_i$, tedy $\widehat{a}_n = \max_{i=1,\dots,n} X_i$. 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

$$\begin{aligned} \mathsf{P}(|U_n - a| > \epsilon) &= \mathsf{P}(U_n > a + \epsilon) + \mathsf{P}(U_n < a - \epsilon) = \mathsf{P}\left(\bigcap_{i=1}^{n} \{X_i < a - \epsilon\} \right) \\ &= \prod_{i=1}^{n} \mathsf{P}(X_i < a + \epsilon) = \left(\frac{a - \epsilon}{a} \right)^n \quad \xrightarrow[n\to\infty]{} 0, \qquad \forall a \in \mathbb{R}^+. \end{aligned}$$

Využili jsme, že $\mathsf{P}(\max_{i=1,\dots,n} X_i > a) = 0$ ve druhé rovnosti, a nezávislost X_i ve třetí rovnosti.

Abychom ukázali silnou konzistenci, musíme zjistit, jestli $\mathsf{P}(\lim_{n\to\infty} \widehat{a}_n = a) = 1$. Můžeme využít speciální vlastnost odhadu $\widehat{a}_n$, že je neklesající v n, $\forall \omega \in \Omega$, $a \in [0,1]$. Takže platí

$$\mathsf{P}\left(\lim_{n\to\infty} \widehat{a}_n \leq c \right) = \mathsf{P}\left(\bigcap_{n=1}^{\infty} \{\widehat{a}_n \leq c\} \right) = \lim_{n\to\infty} \mathsf{P}(\widehat{a}_n \leq c), \tag{5.4}$$

ze spojitosti pravděpodobnosti, neboť jevy $B_n = \{\widehat{a}_n \leq c\}$ jsou také monotónní. Počítejme tedy

$$\begin{aligned} \mathsf{P}\left(\lim_{n\to\infty} \widehat{a}_n = a \right) &= 1 - \mathsf{P}\left(\lim_{n\to\infty} \widehat{a}_n < a \right) = 1 - \mathsf{P}\left(\bigcup_{m=1}^{\infty} \left\{ \lim_{n\to\infty} \widehat{a}_n \leq a - \frac{1}{m} \right\} \right) \\ &= 1 - \lim_{m\to\infty} \mathsf{P}\left(\lim_{n\to\infty} \widehat{a}_n \leq a - \frac{1}{m} \right) = 1 - \lim_{m\to\infty} \lim_{n\to\infty} \mathsf{P}\left(\widehat{a}_n \leq a - \frac{1}{m} \right) = 1 - \lim_{m\to\infty} 0 = 1. \end{aligned}$$

Využili jsme, že když je nerovnost $\lim_{n\to\infty} \widehat{a}_n(\omega) < a$ ostrá, musí existovat nějaké $m \in \mathbb{N}$ takové, že platí i $\lim_{n\to\infty} \widehat{a}_n(\omega) \leq a - \frac{1}{m}$. Proto platí druhá rovnost. Jevy $\{\lim_{n\to\infty} \widehat{a}_n \leq a - \frac{1}{m}\}$ jsou monotónní v m, takže ve třetí rovnosti jsme použili spojitost pravděpodobnosti. Ve čtvrté jsme použili (5.4). Dokázali jsme tedy, že $U_n = \widehat{a}_n$ je i silně konzistentní odhad a.

Pro střední hodnotu můžeme spočítat

$$\mathsf{E}_a U_n = \mathsf{E}_a \max_{i=1,\dots,n} X_i = \int_0^a x\; n\; x^{n-1}\; a^{-n}\; \mathrm{d}x = \frac{n}{n+1}\, a,$$

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. Definujme 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 $T_n = 2\overline{X}_n$ parametru a. Ten byl také nestranný a silně konzistentní. Tedy oba odhady jsou centrované kolem „správného“ a a pro $n \to \infty$ k němu konvergují skoro jistě. Nabízí se otázka, který odhad je lepší. Když spočítáme rozptyl obou odhadů, zjistíme, že

$$\operatorname{var} V_n = \frac{a^2}{n(n+2)} \qquad \operatorname{var} T_n = \frac{a^2}{3n}.$$

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* $\mathsf{E}_\theta(\widehat{\theta}_n - \theta)^2$ (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 $\widehat{\theta}_n$ a tím, co má odhadovat. Samozřejmě závisí na hodnotě neznámého θ. Můžeme si ji rozepsat:

$$\mathsf{E}_\theta\left(\widehat{\theta}_n - \theta\right)^2 = \mathsf{E}_\theta\left(\widehat{\theta}_n - \mathsf{E}_\theta\widehat{\theta}_n\right)^2 + \mathsf{E}_\theta\left(\mathsf{E}_\theta\widehat{\theta}_n - \theta\right)^2 = \operatorname{var}_\theta\widehat{\theta}_n + \left(\mathsf{E}_\theta\widehat{\theta}_n - \theta\right)^2.$$

Závisí tedy na obojím - rozptylu $\widehat{\theta}_n$ i vychýlení $\mathsf{E}_\theta\widehat{\theta}_n - \theta$. Extrémní příklad toho, jak by to dopadlo, kdybychom brali ohled jen na rozptyl $\widehat{\theta}_n$, je následující:

Příklad. Buď $X_1, X_2, \ldots, X_n$ náhodný výběr z modelu $\mathcal{F} = \{Alt(p) : p \in [0,1]\}$. A definujme $\widehat{p}_n = \frac{1}{2}$. Pak zřejmě $\operatorname{var}(\widehat{p}_n) = 0$, tedy minimální možný. Ale vychýlení

$$\mathsf{E}_p(\widehat{p}_n - p) = \frac{1}{2} - p, \qquad \forall p \in [0,1],$$

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 potřebujeme odhadnout parametrickou funkci $g(\theta)$? Pak se jako rozumná strategie jeví definovat

$$\widehat{g(\theta)}_n = g(\widehat{\theta}_n).$$

Pokud byl $\widehat{\theta}_n$ „rozumný“, je šance, že i $\widehat{g(\theta)}_n$ by mohl být „rozumný“. Například, pokud je g spojitá, tak se konzistence $\widehat{\theta}_n$ přenese i 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.

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.

Zde končí předn. 22 (27.4.)

5.2 Intervalový odhad

Definice 5.11 Buď P_θ z modelu $\mathcal{F}$, $\mathbf{X} = (X_1, \ldots, X_n)$ 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 $\eta_L : \mathbb{R}^n \to \mathbb{R}$ a $\eta_U : \mathbb{R}^n \to \mathbb{R}$, jejichž předpis nezávisí na θ, a $\forall\theta \in \Theta$ platí

$$\mathsf{P}_\theta(\eta_L(\mathbf{X}) < g(\theta) < \eta_U(\mathbf{X})) \geq 1 - \alpha.$$

Poznámka. $\eta_L(\mathbf{X})$ a $\eta_U(\mathbf{X})$ jsou náhodné veličiny. Interval chceme co nejužší (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 Buď P_θ z modelu $\mathcal{F}$, $\mathbf{X} = (X_1, \ldots, X_n)$ 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 $\eta_D : \mathbb{R}^n \to \mathbb{R}$, jejíž předpis nezávisí na θ a $\forall\theta \in \Theta$ platí

$$\mathsf{P}_\theta(\eta_D(\mathbf{X}) < g(\theta)) \geq 1 - \alpha.$$

Horním intervalovým odhadem parametrické funkce $g(\theta)$ *o spolehlivosti* $(1 - \alpha)$ nazveme borelovskou funkci $\eta_H : \mathbb{R}^n \to \mathbb{R}$, jejíž předpis nezávisí na θ a $\forall\theta \in \Theta$ platí

$$\mathsf{P}_\theta(g(\theta) < \eta_H(\mathbf{X})) \geq 1 - \alpha.$$

A jak zkonstruovat intervalový odhad? Obecný postup je možno popsat následovně:

- Najdeme funkci $H(X_1, \ldots, X_n; \theta)$ takovou, že rozdělení $H(X_1, \ldots, X_n; \theta)$ 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
$$\mathsf{P}_\theta(q_L < H(X_1, \ldots, X_n; \theta) < q_U) \geq 1 - \alpha, \qquad \forall \theta \in \Theta.$$
- Nyní oddělíme v předpisu H od sebe $g(\theta)$ a náhodný výběr $(X_1, \ldots, X_n)$ tak, aby
$$q_L < H(X_1, \ldots, X_n; \theta) < q_U \qquad \Leftrightarrow \qquad \eta_L(X_1, \ldots, X_n) < g(\theta) < \eta_U(X_1, \ldots, X_n).$$
- 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 $\mathcal{F} = \{Rovnom[0, a], a \in \mathbb{R}^+\}$, 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 $X_i \in [0, a]$, $i = 1, \ldots, n$, skoro jistě a tedy $\frac{X_i}{a} \in [0, 1]$, $i = 1, \ldots, n$, skoro jistě. Snadno odvodíme

$$\mathsf{P}_a\left(\frac{X_i}{a} \leq x\right) = \mathsf{P}_a\left(X_i \leq a\, x\right) = \begin{cases} 0, & x < 0, \\ \frac{a\,x}{a} = x, & x \in [0, 1], \\ 1, & x > 1. \end{cases}$$

Tedy $Y_i = \frac{X_i}{a} \sim Rovnom[0, 1]$, 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 $\max\limits_{i=1,\ldots,n} Y_i$ platí $\mathsf{P}\left(\max\limits_{i=1,\ldots,n} Y_i \leq x\right) = (\frac{x}{1})^n = x^n$, $x \in [0, 1]$.

Volíme
$$H(X_1, \ldots, X_n; a) = \frac{1}{a} \max_{i=1,\ldots,n} X_i,$$
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 $H(X_1, \ldots, X_n; a)$, vidíme, že nejkratší (q_L, q_U) získáme volbou $q_U = 1$ a q_L tak, aby
$$\mathsf{P}_a\left(q_l < \frac{\max_{i=1,\ldots,n} X_i}{a} < 1\right) = \mathsf{P}_a\left(\max_{i=1,\ldots,n} X_i < a < \frac{\max_{i=1,\ldots,n} X_i}{q_L}\right) = 1 - \alpha.$$

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
$$\left(\max_{i=1,\ldots,n} X_i, \alpha^{-\frac{1}{n}} \max_{i=1,\ldots,n} X_i\right).$$

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

$$\mathsf{P}(X \geq q_\beta) \geq 1 - \beta \qquad \text{a zároveň} \qquad \mathsf{P}(X \leq q_\beta) \geq \beta.$$

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 $\chi^2_{\beta,n} = \chi^2_n(\beta)$,
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 spojité 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.

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.

Začněme situací se známým rozptylem:

Věta 5.3 Buď $\mathbf{X} = (X_1, \ldots, X_n)$ náhodný výběr z modelu $\mathcal{F} = \{N(\mu, \sigma^2), \mu \in \mathbb{R}\}$, kde $\sigma^2 > 0$ je známá konstanta. Pak

$$\left(\overline{X}_n - u_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \overline{X}_n + u_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right) \tag{5.5}$$

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 $\overline{X}_n \sim N\left(\mu, \frac{\sigma^2}{n}\right)$. Pomocí věty 2.15 o monotónní transformaci pak následně dostaneme, že

$$\frac{\overline{X}_n - \mu}{\sigma} \sqrt{n} \sim N(0, 1)$$

(nebo viz příklad na str. 31).

Volme tedy v obecném postupu pro odvození intervalového odhadu $H(X_1, \dots, X_n; \mu) = \frac{\overline{X}_n - \mu}{\sigma} \sqrt{n}$. Její rozdělení je $N(0, 1)$ a nezávisí na hodnotě μ. Tedy platí

$$\mathsf{P}_\mu \left(u_{\frac{\alpha}{2}} < H(\mathbf{X}, \mu) < u_{1-\frac{\alpha}{2}}\right) = 1 - \alpha, \qquad \forall \mu \in \mathbb{R}.$$

Použijeme symetrii normovaného normálního rozdělení, abychom přepsali $u_{\frac{\alpha}{2}} = -u_{1-\frac{\alpha}{2}}$, a ekvivalentně přepíšeme nerovnosti v závorce:

$$\begin{aligned} 1 - \alpha &= \mathsf{P}_\mu \left(-u_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} < \overline{X}_n - \mu < u_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right) = \mathsf{P}_\mu \left(|\overline{X}_n - \mu| < u_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right) \\ &= \mathsf{P}_\mu \left(\overline{X}_n - u_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} < \mu < \overline{X}_n + u_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right), \qquad \forall \mu \in \mathbb{R}. \end{aligned}$$

Tím je věta dokázána. □

Zde končí předn. 23 (4.5.)

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 Cramérovy-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 Buď $\mathbf{X}$ náhodný vektor s rozdělením $N_n(0, \sigma^2 \mathbb{I}_n)$, kde $\sigma^2 > 0$. Buď C ortonormální matice rozměru $n \times n$ (tj. platí $CC^\top = C^\top C = \mathbb{I}_n$). 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. □

Věta 5.5 Buď $\mathbf{X} = (X_1, \dots, X_n)$ náhodný výběr z rozdělení $N(\mu, \sigma^2)$, $\mu \in \mathbb{R}$, $\sigma^2 \in \mathbb{R}^+$. Pak platí:

(i) $\overline{X}_n$ a $S_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \overline{X}_n)^2$ jsou nezávislé náhodné veličiny,

(ii) náhodná veličina $\frac{n-1}{\sigma^2}S_n^2 = \sum_{i=1}^n \frac{(X_i-\overline{X}_n)^2}{\sigma^2}$ má χ^2_{n-1} rozdělení,

(iii) náhodná veličina $T_n = \frac{\overline{X}_n-\mu}{S_n}\sqrt{n}$ 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 $c_{1,i} = \frac{1}{\sqrt{n}}$, $i = 1, \ldots n$. Zbytek matice volíme libovolně tak, aby celá C byla ortonormální (to lze). Položme $\mathbf{Y} = C\mathbf{X}$. Potom

$$Y_1 = \frac{1}{\sqrt{n}}\sum_{i=1}^n X_i = \sqrt{n}\,\overline{X}_n \sim N(0,\sigma^2).$$

Také platí

$$\sum_{i=1}^n Y_i^2 = \mathbf{X}^\top C^\top C\,\mathbf{X} = \mathbf{X}^\top\mathbf{X} = \sum_{i=1}^n X_i^2,$$

takže

$$\sum_{i=2}^n Y_i^2 = \sum_{i=1}^n X_i^2 - Y_1^2 = \sum_{i=1}^n X_i^2 - n(\overline{X}_n)^2 = \sum_{i=1}^n (X_i - \overline{X}_n)^2 = (n-1)S_n^2.$$

Tedy $\overline{X}_n = \frac{Y_1}{\sqrt{n}}$ a $S_n^2 = \frac{\sum_{i=2}^n Y_i^2}{n-1}$ jsou vzájemně nezávislé, neboť jsou funkcemi Y_1 a $\{Y_2,\ldots,Y_n\}$, 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 $\sum_{i=2}^n \frac{Y_i^2}{\sigma^2} = \frac{n-1}{\sigma^2}S_n^2 \sim \chi^2_{n-1}$ z definice χ^2 rozdělení. Tím je dokázán bod (ii).

Lineární transformací vzájemně nezávislých $\overline{X}_n$ a S_n^2 dostaneme také vzájemně nezávislé náhodné veličiny U a V

$$U = \frac{\sqrt{n}(\overline{X}_n - \mu)}{\sigma} = \frac{Y_1}{\sigma} \sim N(0,1) \qquad V = \frac{(n-1)S_n^2}{\sigma^2} \sim \chi^2_{n-1}.$$

Tedy $T_n = \frac{U}{\sqrt{\frac{V}{n-1}}} \sim t_{n-1}$ z definice Studentova t-rozdělení. Tím je dokázán bod (iii). □

Věta 5.6 Buď $\mathbf{X} = (X_1,\ldots,X_n)$ náhodný výběr z modelu $\mathcal{F} = \{N(\mu,\sigma^2), \mu\in\mathbb{R}, \sigma^2 > 0\}$. Pak

(i)
$$\left(\overline{X}_n - t_{n-1}(1-\alpha/2)\frac{S_n}{\sqrt{n}}, \overline{X}_n + t_{n-1}(1-\alpha/2)\frac{S_n}{\sqrt{n}}\right) \tag{5.6}$$
je oboustranný intervalový odhad parametru μ o spolehlivosti $(1-\alpha)$.

(ii)
$$\left(\frac{(n-1)S_n^2}{\chi^2_{n-1}(1-\frac{\alpha}{2})}, \frac{(n-1)S_n^2}{\chi^2_{n-1}(\frac{\alpha}{2})}\right)$$
je oboustranný intervalový odhad parametru σ^2 o spolehlivosti $(1-\alpha)$.

Používáme značení kvantilů χ^2 a Studentova t-rozdělení ze strany 99.

Důkaz: (i) Postupujeme podle obecného postupu odvození intervalového odhadu a volíme $H(\mathbf{X};\mu) = \frac{\overline{X}_n-\mu}{S_n}\sqrt{n}$. 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

$$\begin{aligned} 1-\alpha &= \mathsf{P}_{\mu,\sigma^2}\left(t_{n-1}(\alpha/2) < \frac{\overline{X}_n-\mu}{S_n}\sqrt{n} < t_{n-1}(1-\alpha/2)\right) \\ &= \mathsf{P}_{\mu,\sigma^2}\left(\overline{X}_n - t_{n-1}(1-\alpha/2)\frac{S_n}{\sqrt{n}} < \mu < \overline{X}_n + t_{n-1}(1-\alpha/2)\frac{S_n}{\sqrt{n}}\right), \quad \forall \mu\in\mathbb{R}, \sigma^2\in\mathbb{R}^+, \end{aligned}$$

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^2_{n-1} rozdělení, které nezávisí na parametrech (μ,σ^2). Proto

$$\begin{aligned} 1-\alpha &= \mathsf{P}_{\mu,\sigma^2}\left(\chi^2_{n-1}(\alpha/2) < \frac{n-1}{\sigma^2}S_n^2 < \chi^2_{n-1}(1-\alpha/2)\right) \\ &= \mathsf{P}_{\mu,\sigma^2}\left(\frac{(n-1)S_n^2}{\chi^2_{n-1}(1-\frac{\alpha}{2})} < \sigma^2 < \frac{(n-1)S_n^2}{\chi^2_{n-1}(\frac{\alpha}{2})}\right), \quad \forall \mu\in\mathbb{R}, \sigma^2\in\mathbb{R}^+, \end{aligned}$$

čímž je (ii) dokázáno. □

Horní a dolní intervalové odhady bychom odvodili obdobně, např. dolní intervalový odhad pro μ v normálním modelu je $\left(\overline{X}_n - t_{n-1}(1-\alpha)\frac{S_n}{\sqrt{n}}, +\infty\right)$.

Poznámka. Délka intervalu (5.6) je $2t_{n-1}(1-\alpha/2)\frac{S_n}{\sqrt{n}}$ a je náhodná. I kdyby se náhodou hodnota statistiky S_n přesně rovnala hodnotě neznámého σ, tak je $2t_{n-1}(1-\alpha/2)\frac{S_n}{\sqrt{n}}$ větší než $2u_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}$, délka intervalového odhadu pro případ známého σ^2, neboť platí

$$u_{1-\frac{\alpha}{2}} < t_{n-1}(1-\alpha/2), \qquad \forall\, 0<\alpha<1,\ n>1.$$

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 $n\to\infty$, 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^2_n 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.

nebylo v LS 22/23 přednášeno

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

$$\lim_{n\to\infty} H_n(x) = \Phi(x),\ \ \forall x \in \mathbb{R} \qquad \text{a} \qquad \lim_{n\to\infty} G_n(n + \sqrt{2n}\, x) = \Phi(x),\ \ \forall x \in \mathbb{R}.$$

Důkaz: Buďte $U, Y_1, \ldots, Y_i, \ldots$ vzájemně nezávislé stejně rozdělené náhodné veličiny s $N(0,1)$ rozdělením. Pak $V_n = \sum_{i=1}^n Y_i^2$ má χ^2_n rozdělení s distribuční funkcí G_n a platí $\mathsf{E}\, V_n = n$ a $\mathsf{var}\, V_n = 2n$. 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 $\frac{V_n}{n} \xrightarrow[n\to\infty]{\mathsf{P}} 1$. Věta o spojité transformaci pro konvergenci v pravděpodobnosti nám při volbě funkce $g(x) = 1/\sqrt{x}$, která je spojitá na okolí bodu 1, dá $1/\sqrt{\frac{V_n}{n}} \xrightarrow[n\to\infty]{\mathsf{P}} 1$. Pro posloupnost $\{U_n\}_{n=1}^\infty$, kde $U(\omega) = U_n(\omega)$, $\omega \in \Omega$, $n \in \mathbb{N}$, zřejmě platí $U_n \xrightarrow[n\to\infty]{\mathsf{d}} N(0,1)$ a Cramérova-Sluckého věta pak dává

$$T_n = U_n \frac{1}{\sqrt{\frac{V_n}{n}}} \xrightarrow[n\to\infty]{\mathsf{d}} N(0,1),$$

což je ekvivalentní rovnosti

$$\lim_{n\to\infty} H_n(x) = \Phi(x),\ \ \forall x \in \mathbb{R},$$

protože $\Phi(x)$ je spojitá všude.

Posloupnost $\{Y_i^2\}_{i=1}^\infty$ splňuje i předpoklady centrální limitní věty (neboť $\mathsf{var}\, Y_i^2 = 2 \in \mathbb{R}^+$) a z té plyne, že

$$\frac{V_n - n}{\sqrt{2n}} \xrightarrow[n\to\infty]{\mathsf{d}} N(0,1). \tag{5.7}$$

Přepišme si nyní postupně, za pomoci definice G_n a faktu, že V_n má χ^2_n rozdělení

$$\lim_{n\to\infty} G_n(n + \sqrt{2n}\, x) = \lim_{n\to\infty} \mathsf{P}\left(\sum_{i=1}^n Y_i^2 \le n + \sqrt{2n}\, x\right) = \lim_{n\to\infty} \mathsf{P}\left(\frac{V_n - n}{\sqrt{2n}} \le x\right) = \Phi(x),\ \ \forall x \in \mathbb{R},$$

z (5.7) a toho, že distribuční funkce $\Phi(x)$ je spojitá všude. □

Důsledek. Platí

$$\lim_{n\to\infty} t_n(\alpha) = u_\alpha, \forall \alpha \in (0,1) \qquad \text{a} \qquad \lim_{n\to\infty} \frac{\chi^2_n(\alpha) - n}{\sqrt{2n}} = u_\alpha, \forall \alpha \in (0,1).$$

Důkaz: Plyne snadno z předchozí věty. Ponecháváme čtenáři jako cvičení. □

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 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 střední hodnota dané charakteristiky v obou výběrech (měření před a po) liší (resp. zvýšila, nebo snížila). Odvodíme proto ještě intervalový odhad pro rozdíl středních hodnot dvou výběrů z normálního rozdělení.

Mějme tedy $\mathbf{X} = (X_1, \dots, X_n)$ náhodný výběr z $N(\mu_1, \sigma_1^2)$ a $\mathbf{Y} = (Y_1, \dots, Y_m)$ 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 předpoklad stejných rozptylů v obou výběrech, tj.

$$\sigma_1^2 = \sigma_2^2 = \sigma^2 \in \mathbb{R}^+.$$

Zopakujme si, co už víme:

$$\overline{X}_n \sim N\left(\mu_1, \frac{\sigma^2}{n}\right) \qquad \overline{Y}_m \sim N\left(\mu_2, \frac{\sigma^2}{m}\right).$$

Protože oba výběry jsou nezávislé, jsou nezávislé i náhodné veličiny $\overline{X}_n$ a $\overline{Y}_m$ a $(\overline{X}_n, \overline{Y}_m)$ 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 $(\overline{X}_n - \overline{Y}_m) \sim N(\mu_1 - \mu_2, \sigma^2/n + \sigma^2/m)$ a dále

$$\frac{\overline{X}_n - \overline{Y}_m - (\mu_1 - \mu_2)}{\sigma}\sqrt{\frac{mn}{m+n}} \sim N(0, 1), \tag{5.8}$$

které nezávisí na hodnotě parametrů. Ovšem ještě musíme vyřešit problém neznámého σ^2. Uděláme to obdobně jako v případě jednoho výběru. Označme

$$S_X^2 = \frac{1}{n-1}\sum_{k=1}^{n}(X_k - \overline{X}_n)^2 \qquad S_Y^2 = \frac{1}{m-1}\sum_{k=1}^{m}(Y_k - \overline{Y}_m)^2. \tag{5.9}$$

Z věty 5.5 bod (ii) víme

$$\frac{(n-1)S_X^2}{\sigma^2} \sim \chi^2_{n-1} \qquad \frac{(m-1)S_Y^2}{\sigma^2} \sim \chi^2_{m-1},$$

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

$$\frac{(n-1)S_X^2 + (m-1)S_Y^2}{\sigma^2} \sim \chi^2_{m+n-2}. \tag{5.10}$$

Náhodné veličiny $\overline{X}_n$, $\overline{Y}_m$, S_X^2, S_Y^2 jsou vzájemně nezávislé, a tedy z (5.8), (5.10) a definice t-rozdělení dostaneme

$$H(X_1,\ldots,X_n,Y_1\ldots,Y_m;\mu_1,\mu_2,\sigma^2) = \frac{\overline{X}_n - \overline{Y}_m - (\mu_1-\mu_2)}{\sqrt{\frac{(n-1)S_X^2+(m-1)S_Y^2}{n+m-2}}}\sqrt{\frac{mn}{m+n}} \sim t_{m+n-2}. \tag{5.11}$$

Nyní už snadno dokážeme větu:

Věta 5.8 Buďte $\mathbf{X} = (X_1,\ldots,X_n)$ náhodný výběr z $N(\mu_X,\sigma^2)$ a $\mathbf{Y} = (Y_1,\ldots,Y_m)$ 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

$$\left(\overline{X}_n - \overline{Y}_m - t_{m+n-2}\left(1-\frac{\alpha}{2}\right)S^*\sqrt{\frac{m+n}{mn}},\ \overline{X}_n - \overline{Y}_m + t_{m+n-2}\left(1-\frac{\alpha}{2}\right)S^*\sqrt{\frac{m+n}{mn}}\right),$$

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: Snadný – z obecného postupu na odvození intervalového odhadu za použití H z (5.11), jejíž rozdělení nezávislé na všech parametrech bylo odvozeno před formulací věty. □

Poznámka. Proč bylo potřeba předpokládat stejné σ^2 pro oba výběry? Abychom našli funkci výběrů $\mathbf{X} = (X_1,\ldots,X_n)$ a $\mathbf{Y} = (Y_1,\ldots,Y_m)$, která má χ^2-rozdělení, a ono neznámé σ, které se vyskytuje v (5.8) a (5.10), se pokrátí tak, že předpis statistiky $H(X_1,\ldots,X_n,Y_1\ldots,Y_m;\mu_1,\mu_2,\sigma^2)$ 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 $\mathbf{X} = (X_1,\ldots,X_n)$ náhodný výběr z $N(\mu_X,\sigma^2)$ a $\mathbf{Y} = (Y_1,\ldots,Y_m)$ 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

$$\left(\overline{X}_n - \overline{Y}_m - u_{1-\frac{\alpha}{2}}\sigma\sqrt{\frac{m+n}{mn}},\ \overline{X}_n - \overline{Y}_m + u_{1-\frac{\alpha}{2}}\sigma\sqrt{\frac{m+n}{mn}}\right).$$

Důkaz: Snadný – z obecného postupu na odvození intervalového odhadu za použití H rovné statistice (5.8), jejíž rozdělení nezávisí parametrech. □

nebylo v LS 22/23 předná-šeno

Ve větě 5.8 i 5.9 je v předpokladech nezávislost náhodných výběrů $\mathbf{X} = (X_1, \dots, X_n)$ a $\mathbf{Y} = (Y_1, \dots, Y_m)$. 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 $\mathbf{X} = (X_1, \dots, X_n)$ a $\mathbf{Y} = (Y_1, \dots, Y_m)$ nezávislé nejsou? Pak musíme něco předpokládat o sdruženém rozdělení výběrů $\mathbf{X} = (X_1, \dots, X_n)$ a $\mathbf{Y} = (Y_1, \dots, Y_m)$. Nejpřirozenější je předpokládat, že $(X_1, Y_1)^\top \dots (X_n, Y_n)^\top$ 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 $(X_1, Y_1)^\top \dots (X_n, Y_n)^\top$ náhodný výběr z dvojrozměrného normálního rozdělení

$$N_2\left((\mu_X, \mu_Y)^\top, \begin{pmatrix} \sigma_X^2 & \rho\sigma_X\sigma_Y \\ \rho\sigma_X\sigma_Y & \sigma_Y^2 \end{pmatrix}\right),$$

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

$$\left(\overline{X}_n - \overline{Y}_n - t_{n-1}\left(1 - \frac{\alpha}{2}\right)\frac{S_{D,n}}{\sqrt{n}},\ \overline{X}_n - \overline{Y}_n + t_{n-1}\left(1 - \frac{\alpha}{2}\right)\frac{S_{D,n}}{\sqrt{n}}\right)$$

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 $\overline{D}_n$ 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). □

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 $\overline{X}_n - \overline{Y}_n$ za předpokladů první věty 5.8 bude $\frac{2\sigma^2}{n}$, neboť výběry $\mathbf{X} = (X_1, \dots, X_n)$ a $\mathbf{Y} = (Y_1, \dots, Y_n)$ jsou nezávislé. Za předpokladů druhé věty 5.10 bude $\operatorname{var}(\overline{X}_n - \overline{Y}_n) = \frac{2(1-\rho)\sigma^2}{n}$, 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 promíchání = 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 $\mathbf{X} = (X_1, \ldots, X_n)$ je náhodný výběr z jiného, než normálního modelu? Můžeme využít speciálních vlastností daného modelu, abychom nalezli funkci $H(\mathbf{X}; \theta)$ použitelnou pro obecný postup odvození intervalového odhadu (jako jsme to udělali v příkladu na str. 93). 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

$$\lim_{n\to\infty} \mathsf{P}_\theta(\eta_L(\mathbf{X}) < g(\theta) < \eta_U(\mathbf{X})) = 1 - \alpha, \qquad \forall \theta \in \Theta.$$

Ukažme si to na intervalovém odhadu pro střední hodnotu $\mathsf{E}\, X_1 = \mu$. Předpokládejme, že $\mathbf{X} = (X_1, \ldots, X_n)$ je náhodný výběr z rozdělení s neznámou střední hodnotou μ a konečným rozptylem $\operatorname{var} X_1 = \sigma^2 \in \mathbb{R}^+$. V tom případě jsou splněny předpoklady CLV a platí $\frac{\overline{X}_n - \mu}{\sigma}\sqrt{n} \xrightarrow[n\to\infty]{\mathsf{d}} N(0,1)$. Tedy

Zde končí předn. 24 (11.5.)

$$\lim_{n\to\infty} \mathsf{P}_\mu \left(\overline{X}_n - u_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} < \mu < \overline{X}_n + u_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \right)$$
$$= \lim_{n\to\infty} \mathsf{P}_\mu \left(\left| \frac{\overline{X}_n - \mu}{\sigma} \sqrt{n} \right| < u_{1-\frac{\alpha}{2}} \right) = \Phi\left(u_{1-\frac{\alpha}{2}}\right) - \Phi\left(u_{\frac{\alpha}{2}}\right) = 1 - \alpha,$$

a

$$\left(\overline{X}_n - u_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}},\ \overline{X}_n + u_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \right)$$

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 $\operatorname{var} X_1 = \sigma^2 \in \mathbb{R}^+$ – viz věta 5.1). S_n^2 je i slabě konzistentní odhad, čili $S_n^2 \xrightarrow[n\to\infty]{\mathsf{P}} \sigma^2$. Věta 4.6 o spojité transformaci pro konvergenci v pravděpodobnosti nám pak pro funkci $g(x) = \frac{\sigma}{\sqrt{x}}$, spojitou na okolí bodu σ^2, dává $\frac{\sigma}{\sqrt{S_n^2}} \xrightarrow[n\to\infty]{\mathsf{P}} 1$. Použitím Cramérovy-Sluckého věty následně dostaneme

$$\sqrt{n}\,\frac{\overline{X}_n - \mu}{\sigma}\,\frac{\sigma}{\sqrt{S_n^2}} = \frac{\overline{X}_n - \mu}{\sqrt{S_n^2}}\sqrt{n} \xrightarrow[n\to\infty]{\mathsf{d}} N(0,1), \qquad \forall \mu \in \mathbb{R}, \sigma^2 \in \mathbb{R}^+. \tag{5.12}$$

Takže můžeme formulovat větu:

Věta 5.11 Buď $\mathbf{X} = (X_1, \ldots, X_n)$ náhodný výběr z rozdělení, jež je prvkem modelu $\mathcal{F} = \{\mathsf{P}_\theta, \theta \in \Theta\}$, ve kterém platí $\operatorname{var}_\theta X_1 \in \mathbb{R}^+$. Potom intervalovým odhadem střední hodnoty, tj. odhadem parametrické funkce $\mu = \mathsf{E}_\theta X_1$, o asymptotické spolehlivosti $(1-\alpha)$ je interval

$$\left(\overline{X}_n - u_{1-\frac{\alpha}{2}} \frac{S_n}{\sqrt{n}}, \overline{X}_n + u_{1-\frac{\alpha}{2}} \frac{S_n}{\sqrt{n}} \right). \tag{5.13}$$

Důkaz: Snadný – z obecného postupu na odvození intervalového odhadu za použití $H(\mathbf{X};\mu)$ z (5.12), jejíž rozdělení nezávisí na hodnotě parametru θ. Ponecháváme čtenáři jako cvičení. □

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 spojité 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 máme k dispozici lepší (méně variabilní) odhad σ^2 než je S_n^2, je dobré ho použít.

Příklad. Buď $\mathbf{X} = (X_1, \dots, X_n)$ náhodný výběr z modelu $\mathcal{F} = \{Pois(\lambda), \lambda \in \mathbb{R}^+\}$. Chceme odvodit intervalový odhad λ. Pro Poissonovo rozdělení platí $\mathsf{E}\, X_1 = \mathsf{var}\, X_1 = \lambda \in \mathbb{R}^+$. Tedy jsou splněny předpoklady věty 5.11 a (5.13) je intervalový odhad parametru λ o asymptotické spolehlivosti $(1-\alpha)$.

Ale protože platí, že $\mathsf{var}\, X_1 = \lambda = \mathsf{E}\, X_1$, můžeme ho odhadnout konzistentně také výběrovým průměrem $\overline{X}_n$. Ten má menší variabilitu než S_n^2 a alternativní intervalový odhad

$$\left(\overline{X}_n - u_{1-\frac{\alpha}{2}}\frac{\sqrt{\overline{X}_n}}{\sqrt{n}}, \overline{X}_n + u_{1-\frac{\alpha}{2}}\frac{\sqrt{\overline{X}_n}}{\sqrt{n}}\right) \tag{5.14}$$

o asymptotické spolehlivosti $(1-\alpha)$ bude stabilnější. (Ověřte, že (5.14) 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

$$\mathsf{P}_\theta\left(\eta_L(\mathbf{X}) < \theta < \eta_U(\mathbf{X})\right) = \mathsf{P}_\theta\left(g(\eta_L(\mathbf{X})) < g(\theta) < g(\eta_U(\mathbf{X}))\right).$$

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 $(g(\eta_U(\mathbf{X})), g(\eta_L(\mathbf{X})))$.

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: *Zde končí předn. 25 (17.5.)*

1. Formulace statistického modelu: Tím se musí vždy začít. Tedy $\mathbf{X} = (X_1, \ldots, X_n)$ je náhodný výběr z $\mathsf{P}_{\theta_0} \in \mathcal{F} = \{\mathsf{P}_\theta : \theta \in \Theta\}$.

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, 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.

Poznámka. Uvědomme si, že zde vzniká asymetrie mezi nulovou a alternativní hypotézou. Nulovou hypotézu H_0 můžeme zamítnout ve prospěch alternativní hypotézy H_1, a tím vlastně potvrdit platnost alternativy H_1 (když jsou naše data v rozporu s H_0 a svědčí ve prospěch H_1). Druhá možnost je, že nezamítneme nulovou hypotézu H_0 – to v tom případě, že naše data nejsou v rozporu s H_0 a neumožňují tedy její zamítnutí. To není to samé, co „prokázání platnosti H_0“. Žádný test nemůže „prokázat platnost H_0“, platnost H_0 je možno pouze nezamítnout.

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. Provedení experimentu: 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. $\mathcal{F} = \{Hypergeom(N = 10000, M = \theta, n = 50), \theta \in \Theta = \{0, \dots, N\}\}$. 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

$$\begin{aligned} \mathsf{P}(X = m) &= \frac{\binom{M}{m}\binom{N-M}{n-m}}{\binom{N}{n}}, \quad && \text{pro } 0 \le m \le M \text{ a } 0 \le n - m \le N - M, \\ &= 0, && \text{jinak.} \end{aligned}$$

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. Chyba I. druhu by v tomto případě znamenala, že importér chybně zamítne nulovou hypotézu o dostatečné kvalitě dodávky, a bude požadovat slevu, přestože je dodávka dostatečně kvalitní. To by pro něj bylo velmi zahanbující, a proto chce udržet pravděpodobnost této chyby ne větší než α. Chyba II. druhu by znamenala, že importér chybně nezamítne nulovou hypotézu, i když ve skutečnosti nulová hypotéza neplatí a platí hypotéza alternativní – kvalita dodávky není dostatečná. Tedy přijme dodávku s nedostatečnou kvalitou a zaplatí za ni plnou cenu.

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 chceme W největší možné (a tedy c nejmenší možné), abychom maximalizovali 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 zamítneme hypotézu H_0. Když chceme splnit hladinu testu α, tak požaduji

$$\sum(M) \leq \alpha \quad \forall M \in \Theta_0 = \{0, \dots, 500\}. \tag{5.15}$$

Stačí ovšem najít nejmenší možné c tak, aby platilo $\sum(500) \leq \alpha$. Pro ostatní $M \in \Theta_0$ bude nerovnost (5.15) 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 $\sum(500) \approx \alpha$, tedy hladině testu. Například pro $\alpha = 0.05$ odvodíme v bodě 4. $c = 5$, neboli při šesti zkažených pomerančích ve vzorku zamítáme 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 $\sum(2000) = 0.95$. Tedy schopnost testu správně odhalit neplatnost nulové hypotézy roste se vzdáleností skutečné hodnoty parametru θ od množiny Θ_0.

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

$$\sup_{\theta \in \Theta_0} \mathsf{P}_\theta(\mathbf{X} \in W) = \sup_{\theta \in \Theta_0} \mathsf{P}_\theta(T_n(\mathbf{X}) \in C) \leq \alpha,$$

a zároveň chceme

$$\forall \theta \in \Theta_1 \quad \mathsf{P}_\theta(\mathbf{X} \in W) = \mathsf{P}_\theta(T_n(\mathbf{X}) \in C) = \text{maximální}.$$

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

$$\beta(\theta) = \mathsf{P}_\theta(\mathbf{X} \in W), \quad \theta \in \Theta_1,$$

se nazývá *síla testu proti alternativě* θ. Test s kritickým oborem W^*, který by splňoval hladinu testu α a zároveň

$$\mathsf{P}_\theta(\mathbf{X} \in W^*) \geq \mathsf{P}_\theta(\mathbf{X} \in V) \quad \forall\theta \in \Theta_1,$$

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 sílou testu. Čím menší hladina, tím menší je obecně i síla testu. Čili čím víc se snažíme 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 nulové 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 nulové 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 $g(\theta)$.

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 $\mathbf{X} = (X_1, \dots, X_n)$ a $\mathbf{Y} = (Y_1, \dots, Y_m)$ 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 $(\overline{X}_n - \overline{Y}_m)$, když se hodně liší, tak svědčí proti H_0. Ale rozdělení $(\overline{X}_n - \overline{Y}_m)$ závisí na neznámém σ^2 (i za nulové hypotézy!). Je tedy potřeba $(\overline{X}_n - \overline{Y}_m)$ upravit tak, aby na neznámém σ^2 nezáviselo. Umíme to? – Ano, víme totiž z odvození před větou 5.8, že

$$\frac{\overline{X}_n - \overline{Y}_m - (\mu_X - \mu_Y)}{\sqrt{\frac{(n-1)S_X^2+(m-1)S_Y^2}{n+m-2}}}\sqrt{\frac{mn}{m+n}} \sim t_{m+n-2},$$

ale za platnosti nulové hypotézy H_0 je $\mu_X - \mu_Y = 0$. Tedy za platnosti H_0 platí

$$T(\mathbf{X}, \mathbf{Y}) = \frac{\overline{X}_n - \overline{Y}_m}{\sqrt{\frac{(n-1)S_X^2+(m-1)S_Y^2}{n+m-2}}}\sqrt{\frac{mn}{m+n}} \sim t_{m+n-2}.$$

Jako vhodný kritický obor lze volit

$$C = \left(-\infty, -t_{m+n-2}\left(1-\frac{\alpha}{2}\right)\right] \cup \left[t_{m+n-2}\left(1-\frac{\alpha}{2}\right), +\infty\right), \tag{5.16}$$

neboť ten obsahuje málo pravděpodobné hodnoty statistiky $T(\mathbf{X}, \mathbf{Y})$, a za platnosti nulové hypotézy H_0 máme $\mathsf{P}_{\mu_X,\mu_Y}(T(\mathbf{X}, \mathbf{Y}) \in C) = \alpha$.

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

$$\mathsf{P}_{\mu_X,\mu_Y}\left((\mu_X - \mu_Y) \in \left(\overline{X}_n - \overline{Y}_m - t_{1-\frac{\alpha}{2},m+n-2}S^*\sqrt{\frac{m+n}{mn}},\ \overline{X}_n - \overline{Y}_m + t_{1-\frac{\alpha}{2},m+n-2}S^*\sqrt{\frac{m+n}{mn}}\right)\right) = 1-\alpha.$$

Za platnosti hypotézy H_0 je $\mu_X - \mu_Y = 0$, tedy postupně dostáváme

$$\begin{aligned}
\mathsf{P}_{H_0}\left(\overline{X}_n - \overline{Y}_m - t_{1-\frac{\alpha}{2},m+n-2}S^*\sqrt{\tfrac{m+n}{mn}} < 0 < \overline{X}_n - \overline{Y}_m + t_{1-\frac{\alpha}{2},m+n-2}S^*\sqrt{\tfrac{m+n}{mn}}\right) &= 1-\alpha\\
\mathsf{P}_{H_0}\left(|\overline{X}_n - \overline{Y}_m| < t_{m+n-2}\left(1-\tfrac{\alpha}{2}\right)S^*\sqrt{\tfrac{m+n}{mn}}\right) &= 1-\alpha\\
\mathsf{P}_{H_0}\left(T(\mathbf{X}, \mathbf{Y}) \in \left(-t_{m+n-2}\left(1-\tfrac{\alpha}{2}\right), t_{m+n-2}\left(1-\tfrac{\alpha}{2}\right)\right)\right) &= 1-\alpha
\end{aligned}$$

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 (5.16). Ekvivalentně vyjádřeno pomocí kritického oboru přímo pro $\mathbf{X}$ a $\mathbf{Y}$

$$W = \left\{(\mathbf{x}, \mathbf{y}) : |\overline{x}_n - \overline{y}_m| \geq t_{m+n-2}\left(1-\frac{\alpha}{2}\right)S^*\sqrt{\frac{m+n}{mn}}\right\},$$

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 $g(\theta)$ se spolehlivostí $(1-\alpha)$ (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 se spolehlivostí $(1-\alpha)$, 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.

Zde končí předn. 26 (17.5.) nebylo v LS 22/23 předná- šeno

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 $\overline{X}_n = 180.6$, $\overline{Y}_m = 164.9$, $S_X^2 = 6.5$, $S_Y^2 = 9.3$. Po dosazení zjistíme, že oboustranný intervalový odhad se spolehlivostí $(1-0.05)$ 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
$H_0 : \mu_X - \mu_Y = 10$ proti $H_1 : \mu_X - \mu_Y > 10$
Kritický obor je podle odstavce před příkladem

$$W = \left\{ (\mathbf{x}, \mathbf{y}) : \overline{x}_n - \overline{y}_m - t_{1-\alpha, m+n-2} S^* \sqrt{\frac{m+n}{mn}} \geq 10 \right\},$$

a pro naše data $\overline{X}_n - \overline{Y}_m - t_{1-\alpha, m+n-2} S^* \sqrt{\frac{m+n}{mn}} = 14.3$. 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.

Literatura

Billingsley, P. (1995). *Probability and Measure*. New York: Willey.

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.

Kallenberg, O. (2002). *Foundations of Modern Probability*. New York: Springer Verlag.

Lachout, P. (2004). *Teorie pravděpodobnosti*. Praha: Karolinum.

Rataj, J. (2022). *Teorie míry a integrálu 1 – text přednášky* [online]. dostupné z https://www2.karlin.mff.cuni.cz/~rataj/TMI/TMI-text_2020.pdf.