

**NMAI059 Pravděpodobnost a statistika**  
**Příručka k přednášce.**  
13. listopadu 2017

**Jak používat tuto příručku.** Jde o postupně vznikající text, který má obsahovat všechny definice a věty z přednášky. Až na výjimky bude obsahovat jen stručná vysvětlení, doplňující příklady, důkazy nemusí být všechny a úplné. Proto **tato příručka nemůže nahradit účast na přednášce** a není tak zamýšlena.

Definice a věty se snažím pojmenovat a strukturovat co nejvíce. Na přednášce se pokusím tohoto držet, ale zároveň se budu snažit vysvětlovat význam jednotlivých pojmů a jejich používání. Zde by vše mělo být korektně (přestože překlepy se vyskytnou s pravděpodobností 1), zatímco na přednášce občas dojde k chybě (nejsem robot) a případně použiji intuitivní výklad k lepšímu pochopení. Text má sloužit hlavně k tomu, abyste měli k dispozici základní poznámky a ty si poté porovnali s poznámkami z přednášek.

Některé části textu jsou psány **malým písmem**. To jsou části, které pro první čtení představují doplňkovou informaci. Je to proto, že plně matematicky rigorózní výklad by vyžadoval větší rozsah přednášky a více pojmů, než je pro základní pochopení v danou chvíli nutné. Zájemci o matematickou přesnost a důkladnost tak nejsou ochuzeni, tyto poznámky si přečtou a budou o nich přemýšlet, ostatní se k nim vrátí při druhém čtení příručky.

1. AXIOMY; PRAVDĚPODOBNOSTNÍ PROSTOR; NÁHODNÉ JEVY

**1.1. Axiomatika.** Intuitivní výklad pravděpodobnosti: Nemůžeme předem říci, jak nějaký pokus dopadne, ze zkušenosti víme, že výsledek závisí na mnoha faktorech a nemáme možnost jej přesně vypočítat. Výsledek tedy chápeme jako *náhodný*. Pravděpodobnost je míra „četnosti“: *budeme-li opakovat stejný pokus za stejných podmínek a výsledky jednotlivých pokusů se navzájem neovlivňují, pak podíl jednotlivých výsledků na celkovém počtu pokusů se blíží teoretické pravděpodobnosti*.

Takto o pravděpodobnosti uvažujeme intuitivně. Rigorózní matematický model v roce 1933 zavedl Andrej N. Kolmogorov.

**Definice 1** (Pravděpodobnostní prostor). Mějme neprázdnou množinu  $\Omega$  a na ní systém jevů (podmnožin)  $\mathcal{F}$ . Na  $\mathcal{F}$  uvažujme pravděpodobnostní míru  $P$ , tedy zobrazení  $P : \mathcal{F} \rightarrow [0, 1]$  splňující

- (1)  $P(\Omega) = 1$
- (2) pro libovolné po dvou disjunktní  $A_1, A_2, \dots, A_n \in \mathcal{F}$  platí

$$P\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n P(A_i).$$

(konečná aditivita)

- (3) pro libovolné po dvou disjunktní  $A_1, A_2, \dots \in \mathcal{F}$  platí

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i).$$

(spočetná, neboli  $\sigma$ -aditivita)

Pro tuto chvíli jsme ponechali otázku  $\mathcal{F}$  otevřenou. Z definice však vidíme, že je třeba, aby sjednocení po dvou disjunktních množin z  $\mathcal{F}$  byla opět množina v  $\mathcal{F}$ . Jsou i další požadavky na  $\mathcal{F}$ , které splňuje  $\sigma$ -algebra množin.

**Definice** ( $\sigma$ -algebra). Nechť  $\Omega$  je neprázdna množina. Třídu podmnožin  $\Omega$  označená  $\mathcal{F}$  nazveme  $\sigma$ -algebrou pokud

- (1)  $\emptyset \in \mathcal{F}, \Omega \in \mathcal{F}$ .
- (2)  $A \in \mathcal{F} \Rightarrow \Omega \setminus A \in \mathcal{F}$ .
- (3)  $A_1, A_2, \dots \in \mathcal{F} \Rightarrow \bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$ .

Vidíte, že kromě spočetné aditivity chceme ještě přítomnost doplňku a celé množiny.

**Definice 2** (Klasický pravděpodobnostní prostor). Buď  $\Omega$  neprázdna konečná množina,  $\mathcal{F} = 2^\Omega$ . Definujme  $P(A) = \frac{|A|}{|\Omega|} \forall A \subset \Omega$ . Potom  $(\Omega, \mathcal{F}, P)$  nazýváme *klasický pravděpodobnostní prostor*.

**Definice 3** (Diskrétní pravděpodobnostní prostor). Buď  $\Omega$  neprázdná konečná nebo spočetná množina,  $\mathcal{F} = 2^\Omega$ . Buď  $p : \omega \rightarrow \mathbb{R}$  taková funkce, že  $p(\omega) \in [0, 1] \forall \omega \in \Omega$  a  $\sum_{\omega \in \Omega} p(\omega) = 1$ . Dále definujeme  $P(A) = \sum_{\omega \in A} p(\omega)$ . Potom  $(\Omega, \mathcal{F}, P)$  nazýváme *diskrétní pravděpodobnostní prostor*.

**Definice 4** (Reálný spojitý pravděpodobnostní prostor). Buď  $\Omega$  omezený či neomezený reálný interval,  $\mathcal{F}$  nechť obsahuje všechny otevřené i uzavřené podmnožiny  $\Omega$  a jejich spočetná sjednocení a doplňky. Buď  $f : \omega \rightarrow \mathbb{R}$  taková funkce, že  $f(\omega) \geq 0$  a  $\int_{\Omega} f(\omega) d\omega = 1$ . Dále pro  $A \in \mathcal{F}$  definujeme  $P(A) = \int_A f(\omega) d\omega$ . Potom  $(\Omega, \mathcal{F}, P)$  nazýváme *reálný spojitý pravděpodobnostní prostor*.

V posledním případě může být  $\Omega$  i podmnožinou  $\mathbb{R}^k$  pro  $k > 1$ . V tom případě je nutné uvažovat násobné integrály. Pro účely tohoto seznámení s pravděpodobností nám budou stačit jen dvojnásobné integrály, ale rozšíření není tak složité, jak by se na první pohled mohlo zdát.

## 1.2. Základní názvosloví a početní věty pro pravděpodobnost.

- $\omega \in \Omega$  se nazývá *elementární jev*.  $A \in \mathcal{F}$  se nazývá *náhodný jev*.
- Jestliže  $P(A) = 1$ , potom  $A$  je jev *jistý*.
- Jestliže  $P(A) = 0$ , potom  $A$  je jev *nemožný*.
- Je-li  $A$  náhodný jev, pak  $A^C = \Omega \setminus A$  je *doplňkový jev* jevu  $A$ .

**Věta 1** (Základní výpočetní vzorce). *Buď  $P$  pravděpodobnostní míra na  $\mathcal{F}$ , tedy  $P$  každému náhodnému jevu  $A$  přiřadí jeho pravděpodobnost.*

- (1)  $P(A^c) = 1 - P(A) \forall A \in \mathcal{F}$
- (2)  $A, B \in \mathcal{F} : A \subset B$ , pak  $P(A) \leq P(B)$  a  $P(B \setminus A) = P(B) - P(A)$

*Důkaz.* Jednotlivé body:

- (1)  $P(A \cup A^c) = P(A) + P(A^c)$ . Jelikož  $A \cup A^c = \Omega$  a podle vlastností pravděpodobnostní míry  $P(\Omega) = 1$ ,  $P(A) + P(A^c) = 1$ .
- (2)  $A \subset B \Rightarrow B = A \cup (B \setminus A)$ . To jsou dvě disjunktní množiny. Potom  $P(B) = P(A) + P(B \setminus A) \geq P(A)$ .

□

Důležitou vlastností pravděpodobnostní míry je její spojitost. My ji využijeme jenom v důkazu spojitosti distribuční funkce zprava, proto je zde uvedena jako poznámka.

Vezměme si systém  $\{A_i\}_{i=1}^{\infty} \subset \mathcal{F}$  takový, že  $A_i \subset A_{i+1}$ . Potom pro  $A = \bigcup_{i=1}^{\infty} A_i$  píšeme  $A_i \nearrow A$  a říkáme, že systém  $\{A_i\}_{i=1}^{\infty}$  konverguje monotónně (vzhůru) k  $A$ .

Analogicky systém, pro který platí  $A_i \supset A_{i+1}$  konverguje monotónně (dolů) k  $A = \bigcap_{i=1}^{\infty} A_i$ , což zapisujeme  $A_i \searrow A$ . Uvědomme si, že v obou případech je  $A \in \mathcal{F}$ , protože  $\mathcal{F}$  je  $\sigma$ -algebra.

**Věta** (Spojitost pravděpodobnostní míry). *Necht' máme  $\{A_i\}_{i=1}^{\infty} \subset \mathcal{F}$  takovou, že  $A_i \searrow \emptyset$ . Pak*

$$\lim_{i \rightarrow \infty} P(A_i) = 0.$$

*Důkaz.* Víme, že  $P(\bigcap_{i=1}^{\infty} A_i) = P(\emptyset) = 0$ . To je ekvivalentní s  $1 - P((\bigcap_{i=1}^{\infty} A_i)^c) = 1 - P(\bigcup_{i=1}^{\infty} A_i^c) = 1$ . Dále víme, že  $A_i \supseteq A_{i+1}$  a díky tomu  $A_i^c \subseteq A_{i+1}^c$ , tedy  $A_i^c \nearrow \Omega$ . Nadefinujme si posloupnost  $B_1 = A_1^c, B_{i+1} = A_{i+1}^c \setminus A_i^c$ . Zřejmě  $B_i$  jsou disjunktní,  $\bigcup_{i=1}^{\infty} B_i = \Omega$  a  $P(\bigcup_{i=1}^{\infty} B_i) = P(\Omega) = 1$  a proto  $\sum_{i=1}^{\infty} P(B_i) = 1$ . Poslední součet rozložíme na dva a pro  $n \rightarrow \infty$  dostaneme

$$1 = \underbrace{\sum_{i=1}^n P(B_i)}_{\rightarrow 1} + \underbrace{\sum_{i=n+1}^{\infty} P(B_i)}_{\rightarrow 0}$$

což je důsledkem konečnosti a  $\sigma$ -aditivnosti pravděpodobnosti. Z definice  $B_i$  a z vlastností  $A_i$  plyne  $P(\bigcup_{i=1}^n B_i) = P(\bigcup_{i=1}^n A_i^c) = 1 - P(\bigcap_{i=1}^n A_i) = 1 - P(A_n)$ . Víme, že levá strana konverguje k 1, tedy proto  $P(A_n) \rightarrow 0$ . □

**Věta 2** (Inkluze a exkluze). *Mějme  $A_1, \dots, A_n$  náhodné jevy. Pak*

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

*Důkaz.* Matematickou indukcí.

První krok pro  $n = 2$ : zřejmě  $A = (A \setminus B) \cup (A \cap B)$ ,  $B = (B \setminus A) \cup (A \cap B)$  a proto

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

Indukční krok pro  $n - 1 \rightarrow n$ :

$$\begin{aligned} P\left(\bigcup_{i=1}^n A_i\right) &= P\left(\left(\bigcup_{i=1}^{n-1} A_i\right) \cup A_n\right) = P\left(\bigcup_{i=1}^{n-1} A_i\right) + P(A_n) - P\left(\bigcup_{i=1}^{n-1} (A_i \cap A_n)\right) \\ &= \sum_{i=1}^{n-1} P(A_i) - \sum_{1 \leq i < j \leq n-1} P(A_i \cap A_j) + \dots + (-1)^{n-2} P\left(\bigcap_{i=1}^{n-1} A_i\right) + P(A_n) \\ &\quad - \sum_{1 \leq i < j \leq n-1} P(A_i \cap A_n) + \dots + (-1)^{n-2} P\left(\bigcap_{i=1}^n A_i\right) \end{aligned}$$

odkud po vhodném přeuspořádání a sečtení dostaneme

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

□

Než přikročíme k dalším třem významným výpočetním větám, musíme zavést podmíněnou pravděpodobnost.

Řekněme, že jev  $A$  nastane s pravděpodobností  $P(A)$  nebo nenastane s  $P(A^c) = 1 - P(A)$ . Nyní dostaneme informaci o tom, že nastal jev  $B$ . Pomůže nám tato informace zpřesnit představu o pravděpodobnosti jevu  $A$ ? Triviálně ano v těchto případech:

- Jestliže  $B \subset A$ .
- Jestliže  $B \cap A = \emptyset$ .

V prvním případě jev  $A$  nastat musí, protože víme, že se stal jev  $B$ . Ve druhém případě jev  $A$  nemůže nastat, jev  $B$  jej vylučuje. V ostatních případech to však může být všelijaké.

**Definice 5** (Podmíněná pravděpodobnost). Mějme jevy  $A, B \in \mathcal{F}$ ,  $P(B) > 0$ . Definujeme podmíněnou pravděpodobnost jevu  $A$  při  $B$  jako  $P(A|B) = \frac{P(A \cap B)}{P(B)}$ .

*Rozmyslete.* Ukažte, že podmíněná pravděpodobnost splňuje vlastnosti předepsané pro pravděpodobnosti míry.

Neplatí ale  $P(A|B \cup C) \neq P(A|B) + P(A|C)$ . Najděte protipříklad.

Podmíněná pravděpodobnost za podmínky  $B$  udává pravděpodobnost při dodatečné informaci, že pozorovaný (i když neznámý) elementární jev  $\omega$  splňuje  $\omega \in B$ . Všimněte si, že platí  $P(A|\Omega) = P(A)$ . Intuitivně, znalost faktu  $\omega \in \Omega$  je totiž triviální a nedává žádnou další informaci.

v\_nasobeni

**Věta 3** (O postupném podmiňování, o násobení pravděpodobností). Mějme  $A_1, \dots, A_n$  náhodné jevy (prvky  $\mathcal{F}$ ) takové, že  $P(\bigcap_{i=1}^n A_i) > 0$ . Potom

$$P\left(\bigcap_{i=1}^n A_i\right) = P(A_1 | \bigcap_{i=2}^n A_i) \cdot P(A_2 | \bigcap_{i=3}^n A_i) \cdots P(A_{n-1} | A_n) \cdot P(A_n).$$

*Důkaz.* Matematickou indukcí. □

**Definice 6** (Disjunktní rozklad). Spočetný systém náhodných jevů  $\{B_i\}_{i=1}^{\infty} \subset \mathcal{F}$  nazveme *disjunktním rozkladem*  $\Omega$ , pokud:

- (1)  $B_i \cap B_j = \emptyset \forall i \neq j$
- (2)  $\bigcup_i B_i = \Omega$  (stačí  $P(\bigcup_i B_i) = 1$ ).
- (3)  $P(B_i) > 0 \forall i$

**Věta 4** (O úplné pravděpodobnosti). *Mějme  $A$  náhodný jev a  $\{B_i\}$  disjunkttní rozklad. Pak*

$$P(A) = \sum_i P(A|B_i) \cdot P(B_i)$$

*Důkaz.* Víme, že  $A \cap B_i, i = 1, 2, \dots$  jsou po dvou disjunkttní množiny, proto  $\bigcup_i A \cap B_i = A \cap \bigcup_i B_i = A \cap \Omega = A$ . Odtud

$$P(A) = P\left(\bigcup_i A \cap B_i\right) = \sum_i P(A \cap B_i) = \sum_i P(A|B_i) \cdot P(B_i)$$

□

**Věta 5** (Bayesova). *Mějme  $A$  náhodný jev a  $\{B_i\}_i$  disjunkttní rozklad  $\Omega$ , nechť také  $P(A) > 0$ . Potom*

$$P(B_i|A) = \frac{P(A|B_i) \cdot P(B_i)}{\sum_j P(A|B_j) \cdot P(B_j)}$$

*Důkaz.* Využijeme definici podmíněné pravděpodobnosti a větu o úplné pravděpodobnosti. □

Věta 5 nám dává přesný výpočet pravděpodobnosti současného výskytu  $n$  náhodných jevů. V praxi však může být výpočet podmíněných pravděpodobností ve větě 5 velmi náročný. V teorii pravděpodobnosti se proto používají i různé odhady pravděpodobností složitých jevů. Uvedeme zde jen jednu z nich.

**Věta 6** (Bonferroniho nerovnost). *Mějme  $A_1, \dots, A_n$  náhodné jevy. Pak*

$$P\left(\bigcap_{i=1}^n A_i\right) \geq 1 - \sum_{i=1}^n (1 - P(A_i))$$

*Důkaz.*  $P\left(\bigcap_{i=1}^n A_i\right) = 1 - P\left(\left(\bigcap_{i=1}^n A_i\right)^C\right) = 1 - P\left(\bigcup_{i=1}^n A_i^C\right) \geq 1 - \sum_{i=1}^n P(A_i^C) = 1 - \sum_{i=1}^n (1 - P(A_i))$ . □

**1.3. Nezávislost.** Nezávislost (stochastická nezávislost) náhodných jevů znamená, že výskyt jednoho náhodného jevu neovlivní pravděpodobnost výskytu druhého náhodného jevu.

**Definice 7** (Nezávislost). *Jevy  $A, B \in \mathcal{F}$  jsou nezávislé, pokud  $P(A \cap B) = P(A) \cdot P(B)$ .*

Tato definice je v souladu s intuitivním popisem nezávislosti. Když jevy  $A, B$  jsou nezávislé, potom

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

a naopak, tato rovnost implikuje nezávislost. O nezávislosti však můžeme hovořit i u jevů s nulovou pravděpodobností (které jsme zatím z podmiňování vyloučili).

Nezávislost ale potřebujeme definovat i pro více než dva jevy.

**Definice 8** (Vzájemná nezávislost). *Buďte  $A_1, \dots, A_n$  náhodné jevy, pak  $A_i$  jsou vzájemně nezávislé, pokud  $\forall i_1, \dots, i_k \subset \{1, \dots, n\}$  platí*

$$P\left(\bigcap_{j=1}^k A_{i_j}\right) = \prod_{j=1}^k P(A_{i_j})$$

V této definici si všimněte, že rovnost *musí platit pro všechny podmnožiny indexů*.

*Rozmyslete.* Najděte trojici náhodných jevů  $A, B$  a  $C$  takových, že každé dva jevy jsou nezávislé, ale přitom  $P(A \cap B \cap C) \neq P(A)P(B)P(C)$ .

K tomu stačí uvažovat jen *klasický* pravděpodobnostní prostor s několika elementárními jevy  $\omega_1, \dots, \omega_n$  a vhodně volit  $A, B$  a  $C$ . Jakou nejmenší mohutnost  $\Omega$  jste ke svému příkladu potřebovali?

Někdy potřebujeme rozšířit pojem nezávislosti na *libovolný* počet náhodných jevů. Pak samozřejmě nemůžeme požadovat platnost definující rovnosti z věty <sup>df\_nezavislost\_jevu</sup> 8 pro nekonečné (zvláště pak i nespočetné) průniky. Nezávislost je ale dobře definována pomocí *všech konečných* indexových podmnožin. Ještě se k tomuto vrátíme u nezávislosti náhodných veličin.

Uvědomte si, jaké (zjednodušující) důsledky má nezávislost na výpočty pravděpodobností—viz výše uvedené výpočetní věty.

## 2. NÁHODNÉ VELIČINY A VEKTORY

**2.1. Náhodná veličina a její rozdělení.** Spočítat pravděpodobnost nějakého jevu je jedna věc. Jiný problém je ale pracovat s obecně neznámým, ale náhodným výsledkem. K tomu slouží náhodná veličina (a je třeba si dobře rozmyslet, co znamená).

df\_velicina

**Definice 9.** Mějme  $(\Omega, \mathcal{F}, P)$  pravděpodobnostní prostor. Zobrazení  $X : \Omega \rightarrow \mathbb{R}$  takové, že  $X^{-1}(-\infty, a] = \{\omega, X(\omega) \leq a\} \in \mathcal{F} \forall a \in \mathbb{R}$  se nazývá *náhodná veličina*.

Díky definici náhodné veličiny a díky vlastnostem pravděpodobnostní míry  $P$  tak dovedeme pro *jakýkoliv* interval říci, s jakou pravděpodobností je hodnota náhodné veličiny v *tomto* intervalu.

Matematicky jde o požadavek *měřitelnosti* zobrazení  $X$ . Tím dostáváme do rukou silný nástroj, který nám umožní odvozování nejrůznějších vlastností náhodné veličiny, korektní definice charakteristik a další. Pro základní pochopení to však není nezbytné. Dokud se budeme věnovat diskrétní náhodné veličině, můžeme se bez pojmu měřitelnost obejít (téměř) úplně.

df\_rozdeleni1

**Definice 10** (Rozdělení náhodné veličiny). Bud'  $X : \Omega \rightarrow \mathbb{R}$  náhodná veličina. Pravděpodobnostní míra  $P_X$  definovaná na  $\mathbb{R}$  předpisem  $P_X(-\infty, a] = P[X \leq a]$  se nazývá *rozdělení náhodné veličiny*  $X$ .

Pravděpodobnostní prostor je tak světem, ve kterém se uskutečňuje náhoda a náhodná veličina je modelem, který zviditelňuje náhodu výsledky v reálných číslech. **Na jednom pravděpodobnostním prostoru můžeme definovat více náhodných veličin.** K tomu se dostaneme zanedlouho. Díky náhodné veličině nám vzniká *kanonický* pravděpodobnostní prostor  $(\mathbb{R}, \mathcal{B}, P_X)$ , kde  $\mathcal{B}$  obsahuje všechny otevřené i uzavřené reálné množiny, jejich spočetná sjednocení a doplňky.

*Příklad.* Mějme

$$\Omega = \{1, 2, 3, 4, 5, 6\}^2, \mathcal{F} = 2^\Omega, P(\{\omega\}) = \frac{1}{36}.$$

Definujme

$$X_1(\omega_1, \omega_2) = \omega_1, X_2(\omega_1, \omega_2) = \omega_2, Y(\omega_1, \omega_2) = \omega_1 + \omega_2.$$

Potom  $X_1$  a  $X_2$  mají stejné rozdělení, ale jsou různé, a  $X_i$  a  $Y$  mají různá rozdělení. Na jednom pravděpodobnostním prostoru tedy můžeme definovat více různých modelů, případně různých náhodných veličin se stejným rozdělením.

Určete rozdělení  $X_i$  a  $Y$ .

Pro jaké množiny  $A \subset \mathbb{R}$  umíme najít  $P_X$ ?

- (1)  $(-\infty, a]$
- (2)  $(a, b], a < b = (-\infty, b] \setminus (-\infty, a]$
- (3)  $(a, b) = \bigcup_{n=1}^{\infty} (a, b - \frac{1}{n}]$
- (4) Všechny otevřené množiny
- (5)  $A \in \mathcal{B} \Rightarrow X^{-1} \in \mathcal{F}$  pro všechny  $A \in \mathcal{B}$ ,  $X$  je *Borelovsky měřitelná*

**Definice** (Náhodné jevy generované  $X$ ). Mějme  $X$  náhodnou veličinu, označme množinu  $\mathcal{F}_X$  takovou, že  $\mathcal{F}_X = \{B : B = X^{-1}(A) \text{ pro nějakou } A \in \mathcal{B}\}$ .

**Věta.**  $\mathcal{F}_X$  je také  $\sigma$ -algebra náhodných jevů generovaných v  $X$ .

$\mathcal{F}_X$  je  $\sigma$ -algebra a  $\mathcal{F}_X \subset \mathcal{F}$ . Potom

$$P_X(A) = P[X \in A] = P(\underbrace{X^{-1}(A)}_{\in \mathcal{F}_X \subset \mathcal{F}}).$$

Na levé straně je míra na  $(\mathbb{R}, \mathcal{B})$ , na pravé straně je míra na  $(\Omega, \mathcal{F})$ .

Z definice <sup>df\_rozdeleni1</sup>  $\mathbb{P}_0$  je vidět, že k jednoznačnému určení (pravděpodobnostního) rozdělení  $P_X$  náhodné veličiny  $X$  stačí znát pravděpodobnosti množin typu  $(-\infty, a]$  pro všechna  $a \in \mathbb{R}$ . To nás vede k definici distribuční funkce.

**Definice 11.** Buď  $X$  náhodná veličina a  $P_X$  její rozdělení. Funkce  $F_X : \mathbb{R} \rightarrow [0, 1]$  definovaná jako  $F_X(x) = P_X[-\infty, x] = P[X \leq x]$  se nazývá *distribuční funkce* náhodné veličiny  $X$ .

Distribuční funkce  $F$  plně popisuje rozdělení  $P$  v tom smyslu, že mají-li dvě náhodné veličiny *stejnou distribuční funkci*, jsou *stejně rozdělené*.

**Věta 7** (Vlastnosti distribuční funkce). *Mějme náhodnou veličinu  $X$  a její distribuční funkci  $F_X$ . Pak*

- (1)  $\lim_{x \rightarrow -\infty} F_X(x) = 0$
- (2)  $\lim_{x \rightarrow \infty} F_X(x) = 1$
- (3)  $F_X$  je neklesající a zprava spojitá

*Důkaz.* Plyne z vlastností pravděpodobnostní míry.

K důkazu vlastností distribuční funkce ovšem potřebujeme spojitost pravděpodobnostní míry v prázdné množině a proto se zde omezíme na konstatování, že tomu tak je. Zájemci si najdou poznámku o spojitosti pravděpodobnostní míry výše.  $\square$

Je tomu ale i naopak. Každá funkce, která má vlastnosti distribuční funkce je distribuční funkcí.

**Věta 8.** *Nechť  $F$  splňuje vlastnosti z věty 7. Pak existuje pravděpodobnostní prostor  $(\Omega, \mathcal{F}, P)$  a náhodná veličina  $X$  takové, že  $F$  je distribuční funkcí náhodné veličiny  $X$ .*

*Důkaz.* Volme  $\Omega = \mathbb{R}$ ,  $\mathcal{F}$  systém obsahující všechny otevřené množiny, jejich spočetná sjednocení i průniky a všechny doplňky,  $P$  volme jako míru na  $\mathcal{F}$  definovanou předpisem  $P(-\infty, x] = F(x)$ .

Nyní definujme  $X : \Omega \rightarrow \mathbb{R}$  takové, že  $X(\omega) = \omega$ . Potom  $F_X(a) = P[X \leq a] = P(-\infty, a] = F(a)$ .

Jde o takzvanou kanonickou konstrukci, zde jsme ukázali hlavní krok.  $\square$

**2.2. Náhodný vektor a jeho rozdělení.** V následujícím budeme užívat pro jakékoliv dva vektory  $\mathbf{a}, \mathbf{b} \in \mathbb{R}^d$  úmluvu

$$\begin{aligned} \mathbf{a} \leq \mathbf{b} &\Leftrightarrow a_i \leq b_i \quad \forall i = 1, 2, \dots, d \\ \mathbf{a} < \mathbf{b} &\Leftrightarrow a_i \leq b_i \quad \forall i = 1, 2, \dots, d, \text{ a existuje } j : a_j < b_j. \end{aligned}$$

**Definice 12** (Náhodný vektor). Zobrazení  $\mathbf{X} : \Omega \rightarrow \mathbb{R}^d$ , kde  $(\Omega, \mathcal{F}, P)$  je pravděpodobnostní prostor,  $d \in \mathbb{N}$ ,  $d \geq 2$  takové, že  $\{\omega : \mathbf{X}(\omega) \leq \mathbf{a}\} \in \mathcal{F} \quad \forall \mathbf{a} \in \mathbb{R}^d$  nazveme *náhodný vektor*.

Opět jde o požadavek měřitelnosti zobrazení  $\mathbf{X}$ .

**Definice 13** (Rozdělení a distribuční funkce náhodného vektoru). Buď  $\mathbf{X}$   $d$ -rozměrný náhodný vektor. Pravděpodobnostní míra  $P_{\mathbf{X}}$  definovaná na  $\mathbb{R}^d$  předpisem

$$P_{\mathbf{X}}\left(\prod_{i=1}^d (-\infty, a_i]\right) = P[\mathbf{X} \leq \mathbf{a}] = P\left(\bigcap_{i=1}^d [X_i \leq a_i]\right)$$

nazveme *rozdělení náhodného vektoru*.

Funkce  $F_{\mathbf{X}} : \mathbb{R}^d \rightarrow [0, 1]$  definovaná  $F_{\mathbf{X}}(\mathbf{a}) = P[\mathbf{X} \leq \mathbf{a}]$  se nazývá *distribuční funkce* náhodného vektoru  $\mathbf{X}$ .

Pravděpodobnostní míra  $P_{\mathbf{X}}$  samozřejmě musí být definovaná i na všech otevřených podmnožinách  $\mathbb{R}^d$ . Toto rozšíření z uvedených intervalů typu  $(-\infty, \mathbf{a}]$  je však jednoznačné a proto je postačující k popisu rozdělení právě distribuční funkce. Pro náhodné vektory však bývá i distribuční funkce někdy velmi složitá.

Zavedme si značení: Pro  $\mathbf{a} < \mathbf{b}$  buď  $\Delta_k(\mathbf{a}, \mathbf{b})$  množina těch  $\mathbf{c}$ , pro které existuje právě  $k$  indexů  $i_1, i_2, \dots, i_k$  takových, že  $c_{i_j} = a_{i_j}$  a pro zbytek indexů je  $c_l = b_l$ .

Tedy například pro  $\mathbf{a} = (a_1, a_2, a_3)$ ,  $\mathbf{b} = (b_1, b_2, b_3)$  je  $\Delta_1(\mathbf{a}, \mathbf{b}) = \{(a_1, b_2, b_3), (b_1, a_2, b_3), (b_1, b_2, a_3)\}$ .

**Věta 9** (Vlastnosti distribuční funkce). *Buď  $F_{\mathbf{X}}$  distribuční funkce náhodného vektoru  $\mathbf{X}$ . Pak:*

- (1)  $\lim_{a_i \rightarrow -\infty} F_{\mathbf{X}}(\mathbf{a}) = 0$  pro libovolné  $i$ .
- (2)  $\lim_{a_i \rightarrow \infty} \forall i F_{\mathbf{X}}(\mathbf{a}) = 1$ .
- (3) V každé složce argumentu je  $F_{\mathbf{X}}$  zprava spojitá a neklesající
- (4)  $\forall \mathbf{a} < \mathbf{b}$  platí  $\sum_{k=0}^d (-1)^k \sum_{\mathbf{c} \in \Delta_k(\mathbf{a}, \mathbf{b})} F_{\mathbf{X}}(\mathbf{c}) \geq 0$

Podívejme se na čtvrtou podmínku pro  $\mathbf{a} = (a_1, a_2)$ ,  $\mathbf{b} = (b_1, b_2)$ . Pak tato podmínka říká, že pro  $\mathbf{a} < \mathbf{b}$  platí, že  $F_{\mathbf{X}}(b_1, b_2) - F_{\mathbf{X}}(a_1, b_2) - F_{\mathbf{X}}(b_1, a_2) + F_{\mathbf{X}}(a_1, a_2) \geq 0$ . Jedná se tedy o pravděpodobnost obdélníku  $(a_1, b_1] \times (a_2, b_2]$  a ta musí být pochopitelně nezáporná.

I zde platí, že funkce splňující vlastnosti (1)–(4) věty 9 je distribuční funkcí nějaké náhodné veličiny. Rozmyslete si, že vlastnost (4) neplyne z vlastnosti (3) a je pro distribuční funkci podstatná—tedy funkce splňující jen (1)–(3) nemusí odpovídat pravděpodobnostní míře, neboť může dávat nějaké množině zápornou míru.

**Definice 14** (Marginální rozdělení). Buď  $\mathbf{X}$  náhodný vektor a  $P_{\mathbf{X}}$  jeho rozdělení takové, že  $P_{X_i}(-\infty, a] = \lim_{a_j \rightarrow \infty, j \neq i} P_{\mathbf{X}}(X_{j=1}^d(-\infty, a_j])$  se nazývá *marginální rozdělení*  $X_i$  a  $F_{X_i} = \lim_{a_j \rightarrow \infty, j \neq i} F_{\mathbf{X}}(\mathbf{a})$  se nazývá jeho marginální distribuční funkce  $X_i$ .

**Terminologie:** Náhodný vektor (veličina)  $\mathbf{X}$ , je *diskrétní*, pokud nabývá nejvýše spočetně mnoha hodnot. V tom případě existuje nejvýše spočetná množina  $\mathbb{S}$  a nezáporné hodnoty  $p_{\mathbf{s}}$ ,  $\mathbf{s} \in \mathbb{S}$  takové, že  $P[\mathbf{X} = \mathbf{s}] = p_{\mathbf{s}}$  a  $P[\in A] = \sum_{\mathbf{s} \in A} p_{\mathbf{s}}$ . Jinými slovy  $\mathbf{X}(\omega) \in \mathbb{S} \forall \omega$ .

Náhodný vektor (veličina)  $\mathbf{X}$  je *spojitý*, pokud pro libovolnou hodnotu platí  $P[\mathbf{X} = \mathbf{a}] = 0$  a existuje-li nezáporná funkce  $f$  taková, že  $P[\mathbf{X} \in A] = \int_A f(\mathbf{x})d\mathbf{x}$ .

*Rozmyslete.* Jistě musí platit

$$\sum_{\mathbf{x} \in \mathbb{S}} p_{\mathbf{x}} = 1, \quad \int_{\mathbb{R}^d} f(\mathbf{x})d\mathbf{x} = 1.$$

Uvědomte si, že uvedený integrál je násobný, což je podobné, jako násobné sumy. Nejprve se musí vyčíslit vnitřní integrál a postupně se jde až ke vnějšímu.

Všimněte si také, že pro spojitý vektor nezávisí na tom, změníme-li hustotu  $f$  v konečně mnoha bodech.

Dohodněme se, že pro naše účely postačí, pokud diskrétní náhodný vektor nabývá hodnoty vždy v podmnožině  $\mathbb{N}_0^d$ , neřekneme-li jinak.

Rozdělení diskrétního náhodného vektoru  $\mathbf{X}$  je zcela charakterizováno souborem hodnot  $\{p_{\mathbf{s}}\}_{\mathbf{s} \in \mathbb{S}}$ , rozdělení spojitého náhodného vektoru je plně charakterizováno funkcí  $f$ . Hodnoty  $p_{\mathbf{s}}$  v případě diskrétního náhodného vektoru a funkci  $f$  v případě spojitého náhodného vektoru nazveme *hustotou* náhodného vektoru  $\mathbf{X}$ .

V angličtině se distribuční funkce nazývá *cummulative distribution function* a zkracuje obvykle *cdf*. U hustoty obvykle dodáváme, vůči jaké referenční míře tato hustota je definována. Zde jde o aritmetickou = číselnou míru u diskrétních a o Lebesgueovu míru v případě spojitých náhodných vektorů. V angličtině se hustota nazývá *probability density function* a zkracuje *pdf*.

Svět však není tak přehledný. Náhodné vektory mohou mít jak kombinovaně spojitě i diskrétní složky, dokonce i náhodná veličina může mít diskrétní a spojitou část—například denní úhrn srážek. Ještě jinou možností je náhodná veličina, která není diskrétní, ani spojitá: nabývá nespočetně mnoha hodnot, ale neexistuje žádná  $f$  taková, že by  $P[\mathbf{X} \in A] = \int_A f(\mathbf{x})d\mathbf{x}$ .

*Příklad.* Základními modely rozdělení diskrétní náhodné veličiny jsou:

- (1) Alternativní (Bernoulliho).  $X$  nabývá hodnot z množiny  $\{0, 1\}$ ,  $P[X = 1] = P_{\mathbf{X}}(\{1\}) = p \in (0, 1)$  a  $P[X = 0] = 1 - p$ . Parametr  $p$  je interpretován jako pravděpodobnost úspěchu. Značíme  $\text{Alt}(p)$ .
- (2) Binomické. Rozdělení počtu úspěchů do  $n$  nezávislých pokusů.  $P[X = k] = \binom{n}{k} p^k (1-p)^{n-k}$ , značíme  $\text{Bi}(n, p)$ .
- (3) Geometrické. Počet neúspěchů před prvním úspěchem v nezávislých pokusech.  $P[X = k] = p(1-p)^k$ , značíme  $\text{Geom}(p)$ .
- (4) Poissonovo. Počet událostí v jednotkovém časovém intervalu.  $P[X = k] = \exp(-\lambda) \lambda^k / k!$ , kde  $\lambda > 0$ ,  $k = 0, 1, \dots$ . Značíme  $\text{Po}(\lambda)$ .

*Příklad.* Základními modely rozdělení spojitě náhodné veličiny jsou:

- (1) Rovnoměrné rozdělení na intervalu  $(a, b)$ ,  $-\infty < a < b < \infty$ . Jeho hustota je

$$f(x) = \begin{cases} \frac{1}{b-a} & x \in (a, b) \\ 0 & x \notin (a, b). \end{cases}$$

Toto rozdělení modeluje nulovou informaci o výskytu hodnoty v daném intervalu a značíme jej  $U(a, b)$ .

- (2) Exponenciální rozdělení je dáno hustotou

$$f(x) = \begin{cases} \lambda \exp(-\lambda x) & x > 0 \\ 0 & x < 0, \end{cases}$$

kde  $\lambda > 0$  je parametrem tohoto rozdělení. Toto rozdělení je základním modelem pro dobu do události. Jeho označení je  $\text{Exp}(\lambda)$ .

- (3) Normální rozdělení je dáno hustotou

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

Zde  $\mu \in \mathbb{R}$  a  $\sigma^2 > 0$  jsou parametry. Normální rozdělení se též nazývá Gaussovo a je základním modelem pro spojité náhodné veličiny (někdy až nadužívaným). Značíme jej  $N(\mu, \sigma^2)$ .

V případě, že  $\mu = 0$  a  $\sigma^2 = 1$ , mluvíme o *standardním* či *normovaném* normálním rozdělení, značíme  $N(0, 1)$  a jeho distribuční funkce a hustota se označují  $\Phi$ , případně  $\phi$ . Tím se dává najevo důležitost postavení tohoto rozdělení v pravděpodobnosti a statistice.

*Příklad.* Pro náhodné vektory uveďme jedno diskrétní a jedno spojité rozdělení.

- (1) Multinomické rozdělení je mnohorozměrnou obdobou binomického. Buď  $n$  počet pokusů v nichž může dojít k jednomu z  $k$  výsledků. Výsledek  $i$  má v každém pokusu pravděpodobnost  $p_i$ ,  $\sum_{i=1}^k p_i = 1$ , výsledky pokusů se nijak neovlivňují. Náhodný vektor  $\mathbf{X} = (X_1, X_2, \dots, X_k)$  zachycující počty výsledků v  $n$  pokusech má rozdělení s hustotou

$$P[\mathbf{X} = (n_1, n_2, \dots, n_k)] = \begin{cases} \frac{n!}{n_1!n_2!\dots n_k!} p_1^{n_1} p_2^{n_2} \dots p_k^{n_k} & n_i \in \mathbb{N}_0, \sum_{i=1}^k n_i = n \\ 0 & \text{jinak.} \end{cases}$$

Toto rozdělení označujeme zkratkou  $\text{Mult}(n, p_1, \dots, p_k)$ .

- (2) Mnohorozměrné (zde  $d$ -rozměrné) normální rozdělení je dáno hustotou

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{d/2} \sqrt{\det \Sigma}} \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T \Sigma^{-1}(\mathbf{x} - \boldsymbol{\mu})\right), \quad \mathbf{x} \in \mathbb{R}^d,$$

a parametry rozdělení jsou  $\boldsymbol{\mu} \in \mathbb{R}^d$  a  $\Sigma$  symetrická pozitivně definitní matice typu  $d \times d$ . Označení tohoto rozdělení je  $N_d(\boldsymbol{\mu}, \Sigma)$ .

Pro další účely postačí, pokud se budeme věnovat dvourozměrným spojitým náhodným vektorům. Vše podstatné na nich lze pochopit a omezíme se tak „jen“ na dvounásobné integrování (kterého se nebojte!!). V tom případě je možné omezit se na dvourozměrné normální rozdělení a zkoumat jeho vlastnosti.

*Příklad* (Dvourozměrné normální rozdělení). Uveďme si základní tři varianty tohoto rozdělení.

- (1) *Standardní normované* dvourozměrné normální rozdělení je charakterizováno  $\boldsymbol{\mu} = \mathbf{0}$  a  $\Sigma = I_2$ , jednotková matice. Jeho hustota má tedy jednoduchý tvar

$$f(\mathbf{x}) = \frac{1}{2\pi} \exp\left(-\frac{1}{2}(x_1^2 + x_2^2)\right).$$

- (2) V případě, kdy  $\boldsymbol{\mu} = \mathbf{0}$  a matice  $\Sigma$  má speciální tvar

$$\Sigma = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}, \quad \text{kde } |\rho| < 1$$

pak mluvíme často také o normovaném normálním rozdělení a jeho hustota má tvar (odvodte si z obecného normálního rozdělení)

$$f(\mathbf{x}) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(-\frac{x_1^2 - 2\rho x_1 x_2 + x_2^2}{2(1-\rho^2)}\right).$$



(3) V obecném případě, je hustota dána vzorcem

$$f(\mathbf{x}) = \frac{1}{2\pi\sqrt{\det \Sigma}} \exp\left(-\frac{1}{2}(x_1 - \mu_1, x_2 - \mu_2)\Sigma^{-1}\begin{pmatrix} x_1 - \mu_1 \\ x_2 - \mu_2 \end{pmatrix}\right).$$

**2.3. Nezávislost, náhodný výběr a empirické rozdělení.** Začneme jednoduchým pozorováním.

**Věta 10** (O marginálním rozdělení). *Máme-li náhodný vektor  $\mathbf{X}$  a jeho rozdělení  $P_{\mathbf{X}}$ , pak marginální rozdělení složek  $X_1, \dots, X_d$  jsou jednoznačně určeny rozdělením  $P_{\mathbf{X}}$ .*

*Důkaz.* Očividné. □

*Rozmyslete.* Naopak ale tvrzení **neplatí**.

Hoďme dvěma kostkami. Jejich výsledky jsou  $A, B$ . Dále mějme  $C = A - 1$  pro  $B$  sudé a  $C = A + 1$  pro  $B$  liché. Zapišme si je do tabulky:

$A \setminus B$	1	2	3	4	5	6		$A \setminus C$	1	2	3	4	5	6	
1	1/36	1/36	1/36	1/36	1/36	1/36	1/6	1	0	1/12	0	0	0	1/12	1/6
2	1/36	1/36	1/36	1/36	1/36	1/36	1/6	2	1/12	0	1/12	0	0	0	1/6
3	1/36	1/36	1/36	1/36	1/36	1/36	1/6	3	0	1/12	0	1/12	0	0	1/6
4	1/36	1/36	1/36	1/36	1/36	1/36	1/6	4	0	0	1/12	0	1/12	0	1/6
5	1/36	1/36	1/36	1/36	1/36	1/36	1/6	5	0	0	0	1/12	0	1/12	1/6
6	1/36	1/36	1/36	1/36	1/36	1/36	1/6	6	1/12	0	0	0	1/12	0	1/6
	1/6	1/6	1/6	1/6	1/6	1/6			1/6	1/6	1/6	1/6	1/6	1/6	

Vidíme, že pro náhodné vektory  $(A, B)$  a  $(A, C)$  jsou jejich marginální rozdělení stejná, ale sdružená rozdělení jsou velmi odlišná.

Je krásným matematickým výsledkem, že jakékoliv mnohorozměrné rozdělení je složeno ze svých marginálních pomocí (v jistém smyslu jednoznačně dané) funkce zvané *kopula* a to následovně. Bud'  $F_{\mathbf{X}}$  sdružená distribuční funkce a buďte  $F_{X_i}, i = 1, \dots, d$  její marginální distribuční funkce. Pak existuje (v jistém smyslu jediná, pro spojitě  $F_{\mathbf{X}}$  opravdu jediná) funkce  $C : [0, 1]^d \rightarrow [0, 1]$  taková, že

$$F_{\mathbf{X}}(\mathbf{x}) = C(F_1(x_1), F_2(x_2), \dots, F_d(x_d)).$$

Funkce  $C$  je sama distribuční funkcí, jejíž marginální rozdělení jsou rovnoměrná na intervalu  $(0, 1)$ . Modelování a odhadování funkce  $C$  je předmětem mnoha studií.

Sdružené rozdělení náhodného vektoru  $\mathbf{X}$  určuje, jaký *stochastický* vztah mezi sebou mají jednotlivé náhodné veličiny v  $\mathbf{X}$ . Při jednom speciálním vztahu ale platí, že sdružené rozdělení z marginálního určíme vždy.

**Definice 15** (Nezávislost náhodných veličin). Mějme  $X_1, X_2, \dots, X_k$  náhodné veličiny definované na  $(\Omega, \mathcal{F}, P)$ . Tyto veličiny jsou (stochasticky) *vzájemně nezávislé*, pokud platí

$$F_{\mathbf{X}}(\mathbf{x}) = \prod_{i=1}^k F_{X_i}(x_i) \quad \forall \mathbf{x} = (x_1, \dots, x_k) \in \mathbb{R}^k,$$

kde  $\mathbf{X} = (X_1, \dots, X_k)$ .

**Věta 11** (Rozdělení nezávislých náhodných veličin). *Náhodné veličiny  $X_1, \dots, X_k$  jsou nezávislé, když  $\forall A_1, \dots, A_k$  platí*

$$P\left(\bigcap_{i=1}^k [X_i \in A_i]\right) = \prod_{i=1}^k P[X_i \in A_i].$$

*Důkaz.* Plyne z charakterizace rozdělení náhodného vektoru pomocí distribuční funkce. □

Správně bychom měli uvažovat jen  $A_i \in \mathcal{B}$ , protože pro neměřitelné množiny nejsou hodnoty pravděpodobnosti definovány a dostali bychom se do velkých potíží. Ale s neměřitelnými množinami se zde nebudeme potýkat.

**Věta 12** (Ekvivalentní podmínky nezávislosti). *Diskrétní náhodné veličiny  $X_1, \dots, X_k$  jsou nezávislé právě tehdy, když  $p_{\mathbf{X}}(a_1, \dots, a_k) = \prod_{i=1}^k p_{X_i}(a_i)$ .*

*Spojitě náhodné veličiny  $X_1, \dots, X_k$  jsou nezávislé právě tehdy, když  $f_{\mathbf{X}}(x_1, \dots, x_k) = \prod_{i=1}^k f_{X_i}(x_i)$ .*

*Důkaz.* Plyne ze vztahu mezi hustotou a distribuční funkcí. □

**Problém:** Modely pro náhodné veličiny a vektory jsou jen teoretické konstrukty. Někdy mohou být velmi blízké (binomické rozdělení, pokud máme „zaručenu“—ale jak?—nezávislost a homogenitu pokusů), někdy prostě „jen rozumné“.

Obvykle se snažíme model najít, navrhnout, posoudit, odhadnout, otestovat podle *empirických* zkušeností. K tomu ale potřebujeme zkušenost získat a to lze opakovaným pozorováním náhodného jevu či vektoru o kterém chceme něco zjistit.

Základním modelem je *náhodný výběr*. Jeho podstatou je, že stejnou „náhodnost“ pozorujeme v nezávislých opakováních a z těchto pak můžeme zevšeobecňovat.

**Definice 16** (Náhodný výběr). Posloupnost  $X_1, X_2, \dots, X_n$  náhodných veličin či vektorů takových, že jsou nezávislé a mají stejné rozdělení  $P$ , nazvěme náhodný výběr velikosti  $n$  z rozdělení  $P$ . Ve zkratce říkáme, že  $X_1, \dots, X_n$  jsou iid (independent and identically distributed).

Vraťme se k problematice nezávislosti a porovnejme nezávislost náhodných veličin a jevů. Pro náhodné jevy jsme potřebovali rovnost mezi pravděpodobnostmi průniku a součinem pravděpodobností *pro všechny podmnožiny indexů*. Pro náhodné veličiny je toto nahrazeno požadavkem na distribuční funkci.

I zde se můžeme ptát na nezávislost *libovolné množiny* náhodných veličin. A i zde ji lze definovat pomocí nezávislosti *všech konečných podmnožin*. Pro nás ale bude stačit jediné: nezávislost náhodných veličin  $X_1, X_2, \dots$ , což budeme potřebovat při limitních větách a jejich použití. K tomu ale stačí, aby pro jakékoliv konečné  $n$  byly nezávislé náhodné veličiny  $X_1, \dots, X_n$ .

Na základě náhodného výběru můžeme hledat vhodné modely a odhady pro neznámé rozdělení náhodné veličiny. Například jednoduše lze definovat odhad distribuční funkce.

df\_edf

**Definice 17** (Empirická distribuční funkce). Bud'  $X_1, \dots, X_n$  náhodný výběr z rozdělení  $P$  s distribuční funkcí  $F$ . Pak  $\widehat{F}_n(x) = \frac{1}{n} \sum_{i=1}^n \chi(X_i \leq x)$ , kde  $\chi$  je indikátorová funkce, se nazývá *empirická distribuční funkce*.

Empirická distribuční funkce je docela dobrým odhadem skutečné distribuční funkce náhodné veličiny. Než však budeme moci říci, co to znamená *býti dobrým odhadem* a proč tomu tak je, musíme si pro náhodné veličiny zavést další charakteristiky.

*Poznámka.* Zde vidíme význam pojmu náhodná veličina. Empirická distribuční funkce je funkcí náhodných veličin a proto je sama náhodnou veličinou (to si rozmyslete!). Máme tedy vzorec, který pracuje s předem neznámými hodnotami (náhodnou veličinou) a při opakovaných experimentech (náhodných výběrech) dává různé výsledky, jejichž rozdělení se také dá odvodit. Což částečně hned uděláme.

**Věta 13** (Rozdělení empirické distribuční funkce I). Bud'  $X_1, \dots, X_n$  náhodný výběr z rozdělení  $P$  s distribuční funkcí  $F$  a zvolme pevné  $x$ . Pak  $\widehat{F}_n(x)$  má rozdělení dané předpisem

$$P\left[\widehat{F}_n(x) = \frac{k}{n}\right] = \binom{n}{k} (F(x))^k (1 - F(x))^{n-k}.$$

*Důkaz.* Jistě platí, že  $n\widehat{F}_n(x) = k$ , když právě  $k$  z  $n$  hodnot v náhodném výběru je nejvýše  $x$ . Pro všechny náhodné veličiny  $X_i$  z náhodného výběru platí

$$P[\chi(X_i \leq x) = 1] = P[X_i \leq x] = F(x) = 1 - P[\chi(X_i \leq x) = 0],$$

proto součet  $\sum_{i=1}^n \chi(X_i \leq x)$  má binomické rozdělení a z toho již tvrzení věty snadno plyne.  $\square$

### 3. STŘEDNÍ HODNOTA A DALŠÍ MOMENTY

Definujme číselné charakteristiky, které dostatečně vypovídají o chování náhodné veličiny. K čemu je potřebujeme, když chování náhodné veličiny *plně popisuje* její rozdělení? Některá použití uvidíme později, ale jeden z důvodů je i ten, že rozdělení (pravděpodobnostní míra či distribuční funkce) jsou poměrně nepřehledné a hodně složité pojmy. Číselné charakteristiky umožňují snažší porovnávání a interpretaci náhodných veličin.

def:EX

**Definice 18** (Střední hodnota, matematická definice). Bud'  $X$  náhodná veličina definovaná na pravděpodobnostním prostoru  $(\Omega, \mathcal{F}, P)$ . Hodnotu

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

nazveme *střední hodnotou*  $X$ , má-li integrál vpravo smysl.

V praxi se však střední hodnota počítá pomocí rozdělení náhodné veličiny.

**Věta 14** (Střední hodnota diskrétní náhodné veličiny). Bud'  $X$  diskrétní náhodná veličina s hodnotami v  $\mathbb{S}$  a s hustotou  $p_X$ . Pak

$$E X = \sum_{s \in \mathbb{S}} s P[X = s] = \sum_{s \in \mathbb{S}} = \sum_{s \in \mathbb{S}} s p_X(s),$$

má-li součet vpravo smysl.

*Důkaz.* Využijeme definici:  $E X$ . Množiny  $A_s := \{\omega : X(\omega) = s\}$  tvoří disjunktní rozklad  $\Omega$ . Proto

$$\begin{aligned} E X &= \int_{\bigcup A_s} X(\omega) dP(\omega) = \sum_{s \in \mathbb{S}} \int_{A_s} X(\omega) dP(\omega) \\ &= \sum_{s \in \mathbb{S}} s \int_{A_s} dP(\omega) = \sum_{s \in \mathbb{S}} s P(A_s) \\ &= \sum_{s \in \mathbb{S}} s P[X = s]. \end{aligned}$$

□

**Věta 15** (Střední hodnota spojité náhodné veličiny). Bud'  $X$  spojitá náhodná veličina s hustotou  $f_X$ . Pak

$$E X = \int_{-\infty}^{\infty} x f_X(x) dx,$$

má-li integrál vpravo smysl.

*Poznámka.* Střední hodnota **má smysl**, i když je nekonečná. Vyjde-li tedy integrál v definici <sup>def:EX</sup> 118 nekonečný ( $\pm\infty$ ), má náhodná veličina  $X$  nekonečnou střední hodnotu. Integrál nemá smysl, je-li typu „ $\infty - \infty$ “. Tedy  $\int_{\mathbb{R}} 1 dx$  smysl má (i když je integrál nekonečný), zatímco  $\int_{\mathbb{R}} x^{-1} dx$  smysl nemá.

*Rozmyslete.* Zkuste najít funkci  $p_X(s)$ ,  $s \in \mathbb{Z}$  takovou, že:

- (1)  $p_X(s) \geq 0$
- (2)  $\sum_{s \in \mathbb{Z}} p_X(s) = 1$
- (3)  $\sum_{s > 0} s p_X(s) = \infty$ ,  $\sum_{s < 0} s p_X(s) = -\infty$

Pro tuto funkci neexistuje střední hodnota, jelikož máme nedefinovanou sumu ( $\infty - \infty$ ). Najít hustotu  $f_X$  tak, aby neexistovala střední hodnota je podobně snadné.

*Poznámka* (Terminologie a triviální pozorování).

- Pokud  $E X$  existuje a je konečná, pak mluvíme o náhodné veličině s konečnou střední hodnotou.
- Pokud  $P[X \geq b] = 1$  pro  $b$  nějakou konečnou konstantu, pak  $E X$  existuje a  $E X > b$ . Speciálně, nezáporná náhodná veličina má nezápornou střední hodnotu. (Analogicky pro opačnou nerovnost.)
- Pokud  $\exists a, b$  konečné a  $P[a \leq X \leq b] = 1$ , pak  $E X$  existuje konečná a  $a \leq E X \leq b$ .
- Definujme  $E |X| = \int_{\Omega} |X(\omega)| dP(\omega)$ , která existuje vždy. Pokud  $E |X| < \infty$ , pak  $X \in L_1(P)$  a v tom případě existuje i konečná střední hodnota  $X$ , tedy  $|E X| < \infty$ .

Můžeme definovat i momenty funkce náhodné veličiny a momenty vyšších řádů.

**Definice 19** (Obecné momenty náhodné veličiny). Bud'  $X$  náhodná veličina a  $g : \mathbb{R} \rightarrow \mathbb{R}$ . Pak  $Eg(X) = \int_{\Omega} g(X(\omega)) dP(\omega)$ , má-li integrál vpravo smysl.

**Věta 16** (Výpočet střední hodnoty). *Bud'  $X$  diskrétní náhodná veličina s hodnotami v  $\mathbb{S}$  a s hustotou  $p_X s$ , bud'  $g : \mathbb{R} \rightarrow \mathbb{R}$ . Pak*

$$E g(X) = \sum_{s \in \mathbb{S}} g(s) P[X = s] = \sum_{s \in \mathbb{S}} g(s) p_X(s),$$

*má-li součet vpravo smysl.*

*Bud'  $X$  spojitá náhodná veličina s hustotou  $f_X x$ . Pak*

$$E g(X) = \int_{-\infty}^{\infty} g(x) f_X(x) dx,$$

*má-li integrál vpravo smysl.*

**Věta 17** (Linearita střední hodnoty). *Bud'  $X$  náhodná veličina s konečnou střední hodnotou. Pak*

$$E(a + bX) = a + b E X, \forall a, b \in \mathbb{R}.$$

*Důkaz.* Plyne přímo z definice střední hodnoty a jejího výpočtu. □

**Věta 18** (Jensenova nerovnost). *Bud'  $X$  náhodná veličina s konečnou střední hodnotou  $E X$ . Bud'  $\varphi$  konvexní funkce. Pak  $E \varphi(x) \geq \varphi(E X)$ .*

*Důkaz.* Protože je funkce  $\varphi$  konvexní, pro každé zvolené  $a$  existuje konstanta  $\lambda$  taková, že  $\varphi(x) \geq \varphi(a) + \lambda(x - a)$ . Volbou  $a = E X$  tak dostáváme

$$\varphi(X) \geq \varphi(E X) + \lambda(X - E X).$$

Vzhledem k tomu, že střední hodnota nezáporné náhodné veličiny je nezáporná a  $E(X - E X) = 0$ , tvrzení věty již snadno platí. □

Uvědomme si, že (reálná) funkce náhodné veličiny je opět náhodnou veličinou, jen pokud splňuje požadavky z definice [9](#). To ale běžně používané funkce splňují, v jazyce teorie míry požadujeme, aby uvažovaná funkce byla měřitelná.

**Definice 20** (Významné momenty a momentová vytvořující funkce). *Bud'  $X$  náhodná veličina a uvažujme  $r \in \mathbb{N}$ . Pak*

- (1)  $E X^r$  je  $r$ -tý moment  $X$ .
- (2)  $E |X|^r$  je  $r$ -tý absolutní moment  $X$ .
- (3) pro  $r \in \mathbb{N}$   $E(X - E X)^r$  je  $r$ -tý centrální moment  $X$ .
- (4) pro  $r = 2$  máme  $\text{var } x = E(X - E X)^2$  rozptyl (variance)  $X$ .
- (5)  $\mu_3 = \frac{E(X - E X)^3}{(E(X - E X)^2)^{\frac{3}{2}}}$  je šikmost rozdělení, měří asymetrii.
- (6)  $\Psi_X(t) = E e^{tX}$  je momentová vytvořující funkce definovaná pro ta  $t$ , pro která má smysl pravé strana. Vždy platí  $\Psi_X(0) = 1$ .

Všechny uvedené charakteristiky jsou definovány tehdy, mají-li příslušné integrály a součty smysl.

*Poznámka* (Význam a základní vlastnosti momentů). Momenty shrnují některé vlastnosti náhodných veličin a jejich rozdělení. Existují mezi nimi i některé základní vztahy.

- Střední hodnota  $E X$  je charakteristikou polohy náhodné veličiny.
- Rozptyl je charakteristikou variability, jde o střední čtvercovou odchylku od střední hodnoty.
- Jinou charakteristikou variability je například první absolutní centrální moment  $E |X - E X|$ .
- Buďte  $0 < s < r$ . Jestliže  $E |X|^r < \infty$ , pak  $E |X|^s < \infty$  a  $|E X^r| < \infty$  má-li druhý výraz smysl (například pro  $r \in \mathbb{N}$ ). Dokonce platí  $(E |X|^s)^{1/s} \leq (E |X|^r)^{1/r}$ , speciálně  $E |X| \leq (E |X|^2)^{1/2} \leq (E |X|^3)^{1/3}$ .
- Existuje-li  $\delta > 0$  takové, že momentová vytvořující funkce  $\psi_X(t)$  existuje konečná  $\forall |t| < \delta$ . Pak  $\forall r \in \mathbb{N} E X^r = \psi_X^{(r)}(0)$  ( $r$ -tá derivace  $\psi_X(t)$  v bodě 0).

**Věta 19** (Výpočet a vlastnosti rozptylu). *Bud'  $X$  náhodná veličina s konečným rozptylem. Pak*

$$\text{var } X = \text{E } X^2 - (\text{E } X)^2 = \text{E}(X(X-1)) - \text{E } X(\text{E } X - 1).$$

*Pro jakékoliv konstanty  $a$  a  $b$  platí*

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

*Důkaz.* Přímý výpočet. □

*Poznámka.* Rozptyl je střední hodnotou nezáporné funkce, čili je nezáporný. Proto musí být  $\text{E } X^2 \geq (\text{E } X)^2$ . To plyne také z Jensenovy nerovnosti.

Nyní přidáme momenty náhodného vektoru. V podstatě je důležitý jen jeden, ostatní již máme připravené.

**Definice 21** (Střední hodnota náhodného vektoru). *Bud'  $\mathbf{X}$  náhodný vektor. Potom definujeme:*

- (1)  $\text{E } \mathbf{X} = (\text{E } X_1, \dots, \text{E } X_d)^T$
- (2) Necht'  $g : \mathbb{R}^d \rightarrow \mathbb{R}$  (taková, že  $g(\mathbf{X})$  je náhodná veličina). Pak definujeme  $\text{E } g(\mathbf{X}) = \int_{\Omega} g(\mathbf{X}(\omega)) dP(\omega)$ , pokud všechny integrály existují.

**Věta 20.** *Je-li  $\mathbf{X}$  diskrétní s hodnotami v  $\mathbb{N}_0^d$ , pak  $\text{E } g(\mathbf{X}) = \sum_{\mathbf{z} \in \mathbb{N}_0^d} g(\mathbf{z})P[\mathbf{X} = \mathbf{z}]$ , existuje-li řada napravo.*

*Je-li  $\mathbf{X}$  spojitý s hodnotami v  $\mathbb{R}^d$  a s hustotou  $f_{\mathbf{X}}$ , pak  $\text{E } g(\mathbf{X}) = \int_{\mathbb{R}^d} g(\mathbf{z})f_{\mathbf{X}}(\mathbf{z})d\mathbf{z}$ , existuje-li integrál napravo.*

V obou uvedených případech jde o násobné sčítání či integrování, které se provádí postupně po složkách  $\mathbf{z}$ .

**Věta 21** (Linearita střední hodnoty II). *Bud'  $\mathbf{X} = (X_1, \dots, X_d)$  náhodný vektor a konstanty  $a \in \mathbb{R}, b_1, \dots, b_d \in \mathbb{R}$  a  $\text{E } |X_i| < \infty \forall i$ . Pak*

$$\text{E} \left( a + \sum_{i=1}^d b_i X_i \right) = a + \sum_{i=1}^d b_i \text{E } X_i.$$

*Důkaz.* Jen pro diskrétní náhodný vektor, indukci.

Vezměme si  $(X_1, X_2)$  a  $a, b_1, b_2$ . Potom

$$\begin{aligned} \text{E}(a + b_1 X_1 + b_2 X_2) &= \sum_{\mathbf{z} \in \mathbb{N}_0^2} (a + b_1 z_1 + b_2 z_2) P[X_1 = z_1, X_2 = z_2] \\ &= \sum_{\mathbf{z} \in \mathbb{N}_0^2} a P[X_1 = z_1, X_2 = z_2] + \sum_{\mathbf{z} \in \mathbb{N}_0^2} b_1 z_1 P[X_1 = z_1, X_2 = z_2] + \sum_{\mathbf{z} \in \mathbb{N}_0^2} b_2 z_2 P[X_1 = z_1, X_2 = z_2] \\ &= a \underbrace{\sum_{\mathbf{z} \in \mathbb{N}_0^2} P[X_1 = z_1, X_2 = z_2]}_1 + b_1 \sum_{z_1 \in \mathbb{N}_0} z_1 \underbrace{\sum_{z_2 \in \mathbb{N}_0} P[X_1 = z_1, X_2 = z_2]}_{P[X_1 = z_1]} + b_2 \sum_{z_2 \in \mathbb{N}_0} z_2 \underbrace{\sum_{z_1 \in \mathbb{N}_0} P[X_1 = z_1, X_2 = z_2]}_{P[X_2 = z_2]} \\ &= a + b_1 \sum_{z_1 \in \mathbb{N}_0} z_1 P[X_1 = z_1] + b_2 \sum_{z_2 \in \mathbb{N}_0} z_2 P[X_2 = z_2] = a + b_1 \text{E } X_1 + b_2 \text{E } X_2. \end{aligned}$$

Takto můžeme pokračovat indukci. □

**Definice 22** (Kovariance a korelace). *Bud'  $(X_1, X_2)$  náhodný vektor takový, že  $\text{var } X_1 < \infty, \text{var } X_2 < \infty$ .*

- (1) *Kovariance  $X_1$  a  $X_2$  je definována  $\text{cov}(X_1, X_2) = \text{E}(X_1 - \text{E } X_1)(X_2 - \text{E } X_2)$ .*
- (2) *Korelace  $X_1$  a  $X_2$  je definována*

$$\text{corr}(X_1, X_2) = \frac{\text{cov}(X_1, X_2)}{\sqrt{\text{var } X_1 \text{var } X_2}}$$

Význam kovariance:  $X_1 - EX_1, X_2 - EX_2$  jsou odchylky od příslušných středních hodnot. Proto součin  $(X_1 - EX_1)(X_2 - EX_2)$  je kladný, když násobíme odchylky stejného znaménka. Součin je naopak záporný, když odchylky mají různé znaménko. Takže střední hodnota je kladná, pokud „častěji“ mají odchylky od středních hodnot shodné znaménko a naopak. Ale pozor, kovariance je jen indikátor, rozdělení náhodného vektoru je složité a jedna hodnota nemůže postihnout všechny možnosti závislosti  $X_1$  a  $X_2$ . Korelaci zavádíme proto, abychom získali hodnotu nezávislou na „jednotkách“. Platí totiž

$$\text{cov}(aX, Y) = a \text{cov}(X, Y), \text{ ale } \text{corr}(aX, Y) = \text{sign}(a) \text{corr}(X, Y).$$

*Rozmyslete* (Výpočet kovariance). K výpočtu kovariance můžeme využít přímou definici. Ta však nemusí být úplně vhodná.

Podobně jako pro rozptyl můžeme tento vzoreček zjednodušit:

$$E((X_1 - EX_1)(X_2 - EX_2)) = E(X_1X_2 - X_1EX_2 - X_2EX_1 + EX_1EX_2) = E(X_1X_2) - EX_1EX_2$$

**Věta 22** (Nezávislost a korelace). *Nechť náhodné veličiny  $X$  a  $Y$  jsou nezávislé a mají konečný rozptyl. Pak  $\text{cov}(X, Y) = \text{corr}(X, Y) = 0$ , neboli  $X$  a  $Y$  jsou nekorelované.*

### Opačná implikace NEPLATÍ!!!

*Důkaz.* Uvedeme pro diskrétní náhodný vektor  $(X, Y)$ , pro spojitý jde o analogický postup. Přímou z výpočtu střední hodnoty funkce  $g(X, Y) = XY$  a z nezávislosti  $X$  a  $Y$  plyne

$$\begin{aligned} EXY &= \sum_m \sum_n mnP[X = m, Y = n] = \sum_m \sum_n mnP[X = m]P[Y = n] \\ &= \sum_m mP[X = m] \sum_n nP[Y = n] = EXEY. \end{aligned}$$

Kovariance je tedy nulová a proto i korelace je nulová. □

v\_var\_soucet

**Věta 23** (Rozptyl součtu náhodných veličin). *Bud'  $\mathbf{X} = (X_1, \dots, X_d)$  náhodný vektor s konečnými rozptyly. Pak*

$$\text{var} \left( \sum_{i=1}^d X_i \right) = \sum_{i=1}^d \text{var} X_i + \sum_{1 \leq i \neq j \leq d} \text{cov}(X_i, X_j) = \sum_{i=1}^d \text{var} X_i + 2 \sum_{1 \leq i < j \leq d} \text{cov}(X_i, X_j).$$

*Speciálně, jsou-li  $X_1, \dots, X_d$  nezávislé, pak*

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

*Důkaz.* Podle definice rozepíšeme a postupnou úpravou:

$$\begin{aligned} &\text{var} \left( \sum_{i=1}^d X_i \right) \\ &= E \left( \sum_{i=1}^d X_i - E \left( \sum_{i=1}^d X_i \right) \right)^2 = E \left( \sum_{i=1}^d (X_i - EX_i)^2 + 2 \sum_{1 \leq i < j \leq d} (X_i - EX_i)(X_j - EX_j) \right) \\ &= \sum_{i=1}^d \text{var} X_i + 2 \sum_{1 \leq i < j \leq d} \text{cov}(X_i, X_j). \end{aligned}$$

□

**Definice 23** (Varianční a korelační matice). *Bud'  $\mathbf{X} = (X_1, \dots, X_d)$  náhodný vektor,  $\text{var} X_i < \infty \forall i = 1, \dots, d$ . Potom varianční matice je  $\text{Var} \mathbf{X} = \{\text{cov}(X_i, X_j)\}_{i,j}$  a korelační matice je  $\text{Corr} \mathbf{X} = \{\text{corr}(X_i, X_j)\}_{i,j}$ .*

Varianční matice  $\text{Var} \mathbf{X}$  má díky definici na diagonále právě  $\text{var} X_i$ . Podobně korelační matice má na diagonále samé jedničky.

**Věta 24** (Vlastnosti kovariance a korelace). *Bud'te  $\mathbf{X}$  náhodný vektor s variační maticí  $\text{Var } \mathbf{X}$  (která existuje a má konečné prvky). Dále mějme  $X, Y$  náhodné veličiny s konečným rozptylem. Potom:*

- (1)  $-1 \leq \text{corr}(X, Y) \leq 1, \text{corr}(X, X) = 1$
- (2)  $|\text{corr}(X, Y)| = 1 \Leftrightarrow \exists a \neq 0, b \in \mathbb{R} : X = aY + b$  (Korelace měří míru lineární závislosti  $X$  a  $Y$ .)
- (3)  $\text{cov}(aX + c, bY + d) = ab \text{cov}(X, Y)$
- (4)  $\text{corr}(aX + c, bY + d) = \text{sign}(ab) \text{corr}(X, Y)$
- (5)  $\text{Var } \mathbf{X}, \text{Corr } \mathbf{X}$  jsou symetrické pozitivně semidefinitní
- (6)  $\forall A \in \mathbb{R}^{l \times d}, B \in \mathbb{R}^l, \mathbf{X} \in \mathbb{R}^d : \text{Var}(A\mathbf{X} + B) = A(\text{Var } \mathbf{X})A^T$

Podíváme-li se na vzorec pro rozptyl součtu náhodných veličin, vidíme, že platí

$$\text{var} \sum X_i = \mathbf{1}^T \text{Var } \mathbf{X} \mathbf{1},$$

kde  $\mathbf{1}$  je vektor složený ze samých jedniček. Ve stejném duchu je možné psát maticově i

$$\text{var} \sum a_i X_i = \mathbf{a}^T \text{Var } \mathbf{X} \mathbf{a},$$

pro libovolný vektor  $\mathbf{a}$ . Protože výraz na levé straně je nezáporný, plyne odtud i pozitivní semidefinitnost matice  $\text{Var } \mathbf{X}$ .

Odtud již snadno plyne závěr věty pro  $|\text{corr}(X, Y)| = 1$ . V tomto případě platí, že  $\text{cov}^2(X, Y) = \text{var } X \text{ var } Y$ . Proto

$$\text{Var}(X, Y) = \begin{pmatrix} \text{var } X & \pm \sqrt{\text{var } X \text{ var } Y} \\ \pm \sqrt{\text{var } X \text{ var } Y} & \text{var } Y \end{pmatrix},$$

je symetrická matice, znaménko na vedlejší diagonále odpovídá znaménku korelace a tato matice je singulární. Existuje tedy vektor, konkrétně například  $\mathbf{a}^T = (\text{var } Y, \mp \sqrt{\text{var } X \text{ var } Y})$ , takový, že

$$\mathbf{a}^T \text{Var}(X, Y) \mathbf{a} = 0.$$

To ovšem znamená, že náhodná veličina  $a_1 X + a_2 Y$  má nulový rozptyl a proto je skoro jistě konstantní.

#### 4. KRÁTKÁ ODBOČKA KE KONVERGENCI

Než pokročíme dále, bude dobré zadefinovat dva typy konvergence náhodných veličin. Připomeňme, že náhodná veličina je zobrazení  $X : \Omega \rightarrow \mathbb{R}$  a tedy pojem konvergence toto musí zohlednit.

**Definice 24** (Konvergence v pravděpodobnosti). *Bud'te  $X$  náhodná veličina a  $X_1, X_2, \dots$  posloupnost náhodných veličin definovaných na tomtéž pravděpodobnostním prostoru. Řekneme, že  $X_n$  konverguje v pravděpodobnosti k  $X$ , pokud*

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

Tuto konvergenci značíme  $X_n \xrightarrow{\text{P}} X$ .

*Poznámka.* Limitní náhodná veličina  $X$  v předchozí definici může být (a často bude) konstantou.

Zapišeme-li předchozí konvergenci s využitím *zamlčených* proměnných  $\omega$ , dostaneme

$$\text{P}\{\omega \in \Omega : |X_n(\omega) - X(\omega)| > \varepsilon\} \xrightarrow{n \rightarrow \infty} 0.$$

To znamená, že se zvětšujícím se  $n$  se „zmenšuje“ (alespoň v pravděpodobnostním smyslu) množina těch elementárních jevů, pro něž se  $X_n$  a  $X$  liší o více než „zanedbatelnou“ hodnotu. Uvědomme si však, že množina  $\Omega_\varepsilon := \{\omega \in \Omega : |X_n(\omega) - X(\omega)| > \varepsilon\}$  závisí na  $n$  a obecně tyto množiny netvoří monotónní do sebe vnořenou posoupnost.

**Definice 25** (Konvergence v rozdělení). *Bud'te  $X$  náhodná veličina a  $X_1, X_2, \dots$  posloupnost náhodných veličin definovaných na tomtéž pravděpodobnostním prostoru. Řekneme, že  $X_n$  konverguje v rozdělení (v distribuci) k  $X$ , pokud*

$$F_{X_n}(x) = \text{P}[X_n \leq x] \xrightarrow{n \rightarrow \infty} \text{P}[X \leq x] = F_X(x) \text{ pro všechna } x, \text{ kde } F_X \text{ je spojitá.}$$

Tuto konvergenci značíme  $X_n \xrightarrow{d} X$ .

Zde vlastně nekonvergují  $X_n$ , ale „jen“ jejich rozdělení. Pravděpodobnostní chování náhodných veličin  $X_n$  se blíží chování náhodné veličiny  $X$ .

## 5. ODHADY MOMENTŮ A ODHADY PARAMETRŮ MOMENTOVOU METODOU

Rozdělení náhodného vektoru obvykle neznáme. Dají se rozlišit dvě základní situace (nedělejme si ovšem iluze o úplnosti tohoto výčtu):

- (1) **Parametrické modelování:** O rozdělení máme nějakou představu, obvykle kladenou na hustotu rozdělení. Předpokládáme, že náhodná veličina má rozdělení, které patří do nějaké třídy popsatelné až na konečný (malý) počet parametrů. Příkladem může být počet úspěchů v nějakém počtu pokusů. Model binomického rozdělení je naprosto vyhovující, můžeme-li považovat úspěchy v jednotlivých pokusech za nezávislé a nastávající se stejnou pravděpodobností. V tomto případě obvykle potřebujeme *něco říci o neznámých parametrech*.
- (2) **Neparametrický přístup:** O rozdělení (hustotě) předpokládáme jen celkem málo. Například existence některých momentů a spojitost rozdělení jsou typické slabé požadavky kladené na rozdělení. V tomto případě obvykle chceme *něco říci o charakteristikách rozdělení*, jeho konkrétní tvar nás moc nezajímá. Častou otázkou tohoto typu je, zda dvě náhodné veličiny mají stejné střední hodnoty (rozptyl, ...), nebo zda jsou nezávislé (nekorelované).

**5.1. Náhodný výběr a odhad.** Nejčastějším nástrojem k hledání odpovědí na otázky týkající se náhodných veličin a vektorů (parametricky i neparametricky) je *náhodný výběr*. Opakovaným pozorováním nezávislých výsledků experimentu (výsledky náhodného výběru) dostáváme *data*, ve kterých máme ukrytou dostupnou informaci o rozdělení. Již víme, jak odhadnout distribuční funkci pomocí empirické distribuční funkce. Nyní budeme odhadovat momenty a parametry.

**Definice 26** (Bodový odhad). Buď  $X_1, X_2, \dots, X_n$  náhodný výběr z rozdělení s distribuční funkcí  $F_X$ . Funkci  $g_n : \mathbb{R}^n \rightarrow \Theta$ , jejíž předpis nezávisí na rozdělení  $F_X$  (ani jeho parametrech), nazveme *bodový odhad*.

*Poznámka.* Bodový odhad, ale čeho? V zásadě čehokoliv, parametru, momentu, hustoty, pravděpodobnosti, distribuční funkce. Důležité je, že předpis  $g_n$  nezávisí na neznámém—rozdělení či parametru, ale je funkcí jen hodnot náhodného výběru. Všimněme si, že předpis  $g_n$  obvykle závisí na rozsahu  $n$  náhodného výběru (připomeňte si definici <sup>def. 17</sup> ~~def. 17~~). Z toho ovšem plyne důležitý závěr: *odhad sám je náhodnou veličinou*.

Odhadovat můžeme buď nějaký parametr (například  $p$  v binomickém rozdělení,  $\mu$  v normálním rozdělení), nebo jinou hodnotu svázanou s rozdělením (pravděpodobnost nějaké hodnoty či intervalu, střední hodnotu, distribuční funkci či hustotu v nějakém bodě). V souladu se zvyklostmi označíme symbolem  $\theta$  *hodnotu, kterou odhadujeme*, pokud nebude specifikována přesněji. Odhad hodnoty  $\theta$  založený na náhodném výběru o rozsahu  $n$  se obvykle značí  $\hat{\theta}_n$ .

*Poznámka.* Čtenář v tuto chvíli může být zmaten a ptát se „Co je tedy vlastně odhad, jak jej mám spočítat, jaký je vzorec?“. Podstatné je *co* odhadujeme a také jak *posuzujeme chybu* odhadu—a k té zákonitě dochází, neboť odhad je náhodná veličina a odhadujeme obvykle reálné číslo (vektor).

**5.2. Kvalita odhadu.** Jednou z žádaných vlastností odhadu je jeho nestrannost. Tedy aby se *průměru rovnal* odhadované hodnotě.

**Definice 27** (Nestrannost odhadu). Buď  $X_1, X_2, \dots, X_n$  náhodný výběr z rozdělení  $F_X$  a buď  $\theta$  odhadovaný parametr. Bodový odhad  $\hat{\theta}_n$  nazveme nestranným, pokud  $\forall \theta \in \Theta$  platí  $E\hat{\theta}_n = \theta$  je-li  $\theta$  skutečnou hodnotou parametru.

Chceme tedy, aby střední hodnota odhadu byla onou neznámou odhadovanou hodnotou. Je jistě žádoucí, aby odhad netrpěl systematickou chybou.

Druhý dobrým požadavkem je, aby se odhad *blížil odhadované hodnotě*. Tomu se říká konzistence.

**Definice 28** (Konzistence odhadu). Buď  $X_1, X_2, \dots, X_n$  náhodný výběr z rozdělení  $F_X$  a buď  $\theta$  odhadovaný parametr. Bodový odhad  $\hat{\theta}_n$  nazveme konzistentním, pokud  $\forall \theta \in \Theta$  platí  $\hat{\theta}_n \xrightarrow{P} \theta$  je-li  $\theta$  skutečnou hodnotou parametru.



Konečně posledním častým požadavkem je asymptotická normalita, tedy aby se chování vhodně upraveného odhadu blížilo standardnímu normálnímu rozdělení. O tom si však povíme až později

Zvědavý čtenář si jistě klade spoustu otázek: jak máme zaručeno, že existuje nekonečně mnoho nezávislých náhodných veličin se stejným rozdělením a na stejném pravděpodobnostním prostoru? Spokojme se s tím, že takový prostor lze bezesporně zkonstruovat a tedy je to možné, pokud dokážeme zkonstruovat nekonečně rozměrný součinný prostor a na něm součinnou míru.

Jinou otázkou je, co znamená „pro každý parametr  $\theta$ , je-li  $\theta$  skutečný parametr“, když přeci  $\theta$  neznáme. Znamená to, že předpis pro  $\hat{\theta}_n$  musí být takový, že nezávisí na  $\theta$ , ale jen na hodnotách náhodného výběru (obvykle na hodnotách v  $\mathbb{R}^n$ , kde  $n$  je rozsah výběru (tedy pro každé  $n$  máme trochu jiný předpis pro  $\hat{\theta}_n$ ). Jelikož rozdělení odhadu  $\hat{\theta}_n$  je odvozené od rozdělení náhodného výběru, závisí toto rozdělení také na  $\theta$ . Proto dosadíme-li libovolně  $\theta$  do rozdělení  $\hat{\theta}_n$ , musí nám vyjít to, co chceme.

**5.3. Výběrové momenty.** V této sturčně kapitole jen ukážeme jak počítat výběrové momenty a proč.

**Definice 29** (Výběrový průměr a výběrový rozptyl). Buď  $X_1, \dots, X_n$  náhodný výběr. Výběrovou střední hodnotou nazveme  $\bar{X}_n$  a výběrovým průměrem  $S_n^2$  definované

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

**Věta 25** (Nestrannost výběrového průměru, rozptylu a empirické distribuční funkce). *Je-li  $X_1, \dots, X_n$  náhodný výběr z rozdělení s distribuční funkcí  $F_X$ , pak  $\hat{F}_n(x)$  je nestranným odhadem  $F(x)$  v každém bodě  $x$ .*

*Je-li  $X_1, \dots, X_n$  náhodný výběr z rozdělení s konečnou střední hodnotou  $EX$ , pak  $\bar{X}_n$  je nestranným odhadem  $EX$ .*

*Je-li  $X_1, \dots, X_n$  náhodný výběr z rozdělení s konečným rozptylem  $\text{var } X$ , pak  $S_n^2$  je nestranným odhadem  $\text{var } X$ .*

**Definice 30** (Výběrové momenty). Buď  $X_1, \dots, X_n$  náhodný výběr. Pak

$$\widehat{EX^r} = \frac{1}{n} \sum_{i=1}^n X_i^r$$

se nazývá  $r$ -tý výběrový moment.

**Věta 26** (Nestrannost výběrových momentů). *Je-li  $X_1, \dots, X_n$  náhodný výběr z rozdělení s konečným  $r$ -tým momentem  $EX^r$ , pak  $\widehat{EX^r}$  je nestranným odhadem  $EX^r$ .*

**5.4. Odhady metodou momentů.** Abychom si přeci jen ukázali nějaký způsob konstrukce odhadu, ukažme momentovou metodu, která nepotřebuje žádné další teoretické kapitoly.

Předpokládejme, že distribuční funkce  $F$  (rozdělení  $P$ ) závisí na neznámém parametru. Pak obvykle i momenty náhodné veličiny s tímto rozdělením závisí na těchto parametrech (vzpomeňte si na příklady jednotlivých rozdělení).

Předpokládejme, že chceme odhadnout neznámý parametr  $\theta$ , jehož hodnoty se nacházejí v množině  $\Theta \subset \mathbb{R}^d$  (často  $d = 1, 2$ ). Necht' momenty náhodné veličiny  $X$  existují alespoň do řádu  $d$  a závisí na parametru  $\theta$ , tedy necht' platí

$$EX^r = \mu_r(\theta), r = 1, 2, \dots, d,$$

kde známe funkce  $\mu_r(\theta)$ . Levou stranu nahradíme výběrovými momenty a řešíme soustavu rovnic

$$\widehat{EX^r} = \mu_r(\hat{\theta}_n), r = 1, 2, \dots, d,$$

čili hledáme  $\hat{\theta}_n$  splňující tyto rovnice. Obvykle je pro  $d$ -rozměrný parametr potřeba právě  $d$  těchto momentových rovnic, někdy však může stačit méně.

Obecně odhady momentovou metodou nemusí být nestranné, někdy je možné je na nestranné upravit. Hlavní výhodou této metody je jednoduchost a pochopitelnost.

## 6. LIMITNÍ VĚTY

V této důležité kapitole si představíme některé nerovnosti a limitní věty pro pravěpodobnost a pro náhodný výběr.

v\_markovova

**Věta 27** (Markovova nerovnost I). *Bud'  $X$  nezáporná náhodná veličina. Potom*

$$P[X \geq \varepsilon] \leq \frac{E X}{\varepsilon} \quad \forall \varepsilon > 0$$

*Důkaz.* Ukážeme si pro  $X$  spojitou náhodnou veličinu s hustotou  $f$ , ale lze i obecně.

$$P[X \geq \varepsilon] = \int_{x \geq \varepsilon} 1 f(x) dx \leq \int_{x \geq \varepsilon} \underbrace{\frac{x}{\varepsilon}}_{\text{vždy kladné}} f(x) dx \leq \int_0^{\infty} \frac{x}{\varepsilon} f(x) dx = \frac{E X}{\varepsilon}.$$

□

v\_markovova\_2

**Věta 28** (Markovova nerovnost II). *Bud'  $X$  nezáporná náhodná veličina. Potom*

$$P[X \geq \varepsilon] \leq \frac{E(X^k)}{\varepsilon^k} \quad \forall \varepsilon > 0, k > 0$$

v\_cebysev

**Věta 29** (Čebyševova nerovnost). *Bud'  $X$  náhodná veličina a  $E|X| < \infty$ . Pak*

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

*Důkaz.* Věta [28](#) použitá na  $|X - E X|$  a  $k = 2$ . □

Čebyševova nerovnost může být zajímavě zesílena Kolmogorovovou nerovností.

**Věta 30** (Kolmogorovova nerovnost). *Bud'te  $X_1, X_2, \dots$  nezávislé náhodné veličiny a  $E|X_i| < \infty \forall i$ . Pak*

$$P\left[\max_{1 \leq k \leq n} \left| \sum_{i=1}^k (X_i - E X_i) \right| \geq \varepsilon\right] \leq \frac{\sum_{i=1}^n \text{var } X_i}{\varepsilon^2}.$$

*Důkaz.* Označme si  $S_k = \sum_{i=1}^k (X_i - E X_i)$ . Potom  $E S_k = 0$  a  $\text{var } S_k = \sum_{i=1}^k \text{var } X_i$  díky nezávislosti. Hledáme tedy pravděpodobnost  $P[\max_{1 \leq k \leq n} |S_k| \geq \varepsilon]$ .

Zadefinujme si množinu  $A_k = \{\omega : |S_j(\omega)| < \varepsilon, j < k, |S_k(\omega)| \geq \varepsilon\}$ , kde  $S_0 = 0$ . Můžeme si všimnout, že  $A_k$  jsou po dvou disjunktní, jde tedy o disjunktní rozklad jevu  $[\max_{1 \leq k \leq n} |S_k| \geq \varepsilon]$ .

Nyní předpokládejme, že náhodné veličiny  $X$  jsou spojitě, tedy i  $|S_k|$  jsou spojitě s hustotou  $s_k$  (opět lze dokázat obecně). Platí

$$P[\max_{1 \leq k \leq n} |S_k| \geq \varepsilon] = \sum_{k=1}^n P[A_k] \leq \sum_{k=1}^n \frac{1}{\varepsilon^2} E \chi(A_k) S_k^2,$$

kde k poslední nerovnosti využijeme Markovovu nerovnost. Použijme dále identitu

$$E \chi(A_k) S_n^2 = E \chi(A_k) (S_n - S_k + S_k)^2 = E \chi(A_k) (S_n - S_k)^2 + 2 E \chi(A_k) S_k (S_n - S_k) + E \chi(A_k) (S_k)^2$$

a zaměříme se na střední člen. Díky nezávislosti  $X_i$  dostáváme nezávislost  $\chi(A_k) S_k$  a  $(S_n - S_k)$ , z čehož plyne

$$E \chi(A_k) 2(S_n - S_k) \cdot S_k = E(S_n - S_k) E S_k \chi_{A_k} = 0.$$

Celkově  $E \chi(A_k) (S_k)^2 \leq E \chi(A_k) S_n^2$  a dosazením do předchozího dostaneme

$$\sum_{k=1}^n P[A_k] \leq \sum_{k=1}^n \frac{1}{\varepsilon^2} E \chi(A_k) S_k^2 \leq \sum_{k=1}^n \frac{1}{\varepsilon^2} E \chi(A_k) S_n^2 \leq \frac{1}{\varepsilon^2} E S_n^2.$$

Postačí si uvědomit, že  $E S_n^2 = \text{var } S_n$ , protože  $E S_n = 0$ . □

Nyní lze ukázat, že odhady momentů i odhad distribuční funkce pomocí empirické distribuční funkce jsou konzistentní. Nejprve použijeme vhodné nerovnosti na součty náhodných veličin a odvodíme obecný výsledek, který pak budeme používat na jednotlivé případy.

thm:slzvc

**Věta 31** ((Slabý) Zákon velkých čísel). *Bud' te  $X_1, X_2, \dots$  iid náhodné veličiny s konečnou střední hodnotou  $E X_1 = \mu$  a s konečným rozptylem  $\text{var } X_1 = \sigma^2$ . Pro libovolné  $\varepsilon > 0$*

$$P \left[ \left| \frac{1}{n} \sum_{i=1}^n X_i - \mu \right| > \varepsilon \right] \xrightarrow{n \rightarrow \infty} 0.$$

*Důkaz.* Z věty [26](#) víme, že aritmetický průměr je nestranným odhadem své střední hodnoty  $\mu$ , existuje-li tato. S využitím Čebyševovy nerovnosti

$$P[|\bar{X}_n - \mu| > \varepsilon] \leq \frac{\text{var } \bar{X}_n}{\varepsilon^2},$$

což spolu se skutečností

$$\text{var } \bar{X}_n = \text{var } \frac{1}{n} \sum X_i = \frac{1}{n^2} \text{var } \sum X_i = \frac{1}{n^2} n \text{var } X_1 = \frac{\text{var } X_1}{n} \xrightarrow{n \rightarrow \infty} 0,$$

kde předposlední rovnost platí díky nezávislosti a stejnému rozdělení  $X_i$ , dává požadovanou konvergenci.  $\square$

**Důsledek 1.** *Bud' te  $X_1, X_2, \dots$  iid náhodné veličiny s konečným  $2r$ -tým momentem  $E X_1^{2r}$ . Pak pro libovolné  $\varepsilon > 0$*

$$P \left[ \left| \frac{1}{n} \sum_{i=1}^n X_i^r - E X_1^r \right| > \varepsilon \right] \xrightarrow{n \rightarrow \infty} 0.$$

Ve skutečnosti lze předchozí větu dokázat i silněji. Zaprvé nepotřebujeme nezávislost a stačí nekorelovanost, protože v tom případě také platí, že

$$\text{var } \sum_i X_i = \sum_i \text{var } X_i + \sum_{i \neq j} \underbrace{\text{cov}(X_i, X_j)}_{=0} = \sum_i \text{var } X_i.$$

Dokonce stačí požadovat ještě méně a to aby kovariance byly dostatečně rychle mizející se vzdáleností indexů  $i$  a  $j$ , aby  $n^{-2} \text{var } \sum_i X_i$  konvergovalo k nule při  $n$  rostoucím do nekonečna. Zadruhé, pokud si *ponecháme* nezávislost, můžeme naopak požadavek na existenci „druhého“ momentu nahradit požadavkem  $E |X| < \infty$  (ve větě [31](#), případně  $E |X|^r < \infty$  v důsledku) a věta zůstane v platnosti. Dokonce lze i zesílit smysl konvergence na konvergenci ve *všech*  $\omega$  až na množinu pravděpodobnosti 0. Takovému výsledku pak říkáme *silný zákon velkých čísel*. Praktici často považují rozdíl mezi silným a slabým ZVČ za zanedbatelný a z hlediska používání a interpretace nezajímavý, teoretici naopak velmi lpějí na jejich rozlišování a jejich snahou je pokud možno dokázat silný zákon (konvergenci skoro všude). V našem textu však nemáme techniky pro důkaz silného zákona, zájemci si laskavě vyhledají příslušné výsledky v bohatě dostupné literatuře.

Dokázali jsme tedy konzistenci odhadů momentů náhodných veličin. Pokud jsou parametry odhadované pomocí metody momentů spojitými funkcemi momentů (a jednoznačně určenými), pak ze spojitosti a konvergence dostaneme i konzistenci odhadů vytvořených momentovou metodou. Základem může být jednoduché tvrzení, které uvedeme bez čísla a bez důkazu.

**Věta** (Konvergence spojitě transformace). *Bud'  $X_1, X_2, \dots$  posloupnost náhodných veličin a a konstanta taková, že  $P[|X_n - a| > \varepsilon] \rightarrow 0$  pro libovolné  $\varepsilon > 0$ . Bud'  $\phi$  spojitá funkce. Pak*

$$P[|\phi(X_n) - \phi(a)| > \varepsilon] \xrightarrow{n \rightarrow \infty} 0 \quad \forall \varepsilon > 0.$$

Další možnou modifikací zákona velkých čísel je vynechání požadavku na stejné rozdělení náhodných veličin. Pak je samozřejmě otázkou, k jaké hodnotě konverguje aritmetický průměr a za jakých podmínek. Opět se vyskytuje více variant odpovědi, ale nám bude stačit jen jednoduchá varianta.

**Věta.** *Bud' te  $X_1, X_2, \dots$  iid náhodné veličiny s konečnou střední hodnotou  $E X_i = \mu_i$  a s omezeným rozptylem  $\text{var } X_i = \sigma_i^2 \leq K$ . Pak*

$$P \left[ \left| \frac{1}{n} \sum_{i=1}^n (X_i - \mu_i) \right| > \varepsilon \right] \xrightarrow{n \rightarrow \infty} 0.$$

*Pokud  $n^{-1} \sum_i \mu_i$  konverguje k hodnotě  $\mu$  při  $n$  rostoucím nade všechny meze, pak platí*

$$P \left[ \left| \frac{1}{n} \sum_{i=1}^n X_i - \mu \right| > \varepsilon \right] \xrightarrow{n \rightarrow \infty} 0.$$

Ted' již můžeme u každého odhadu rozhodovat o jeho nestrannosti (asymptotické nestrannosti) a konzistenci. Mnohem hlubší výsledek nám poskytuje *centrální limitní věta* (CLV). Její důkaz však je buď velice zdlouhavý, nebo vyžaduje pokročilejší techniky, takže se spokojíme se zněním.

**thm:clv**

**Věta 32** (Centrální limitní věta). *Bud' te  $X_1, X_2, \dots$  nezávislé a stejně rozdělené náhodné veličiny s konečnou střední hodnotou  $E X_1 = \mu$  a konečným kladným rozptylem  $\text{var } X_1 = \sigma^2 > 0$ . Pak*

$$P \left[ \sqrt{n} \frac{\overline{X}_n - \mu}{\sigma} \leq x \right] \xrightarrow[n \rightarrow \infty]{} \Phi(x), \quad \forall x \in \mathbb{R},$$

kde  $\Phi$  je distribuční funkce standardního normálního rozdělení  $N(0, 1)$ .

Ekvivalentně můžeme říct

$$\sqrt{n} \frac{\overline{X}_n - \mu}{\sigma} \xrightarrow{d} N(0, 1) \text{ nebo } \frac{\sum_{i=1}^n X_i - n\mu}{\sqrt{n\sigma^2}} \xrightarrow{d} N(0, 1) \text{ nebo } \sqrt{n} (\overline{X}_n - \mu) \xrightarrow{d} N(0, \sigma^2).$$

Použití CLV jsou všude. Využijeme je ke konstrukci (asymptotických) intervalových odhadů, asymptotických predikčních intervalů, testování, ... Zde se naplno ukazuje význam postavení normálního rozdělení: mnoho navzájem nezávislých vlivů se počítá, extrémy se vyruší a výsledkem je normální (Gaussovo) rozdělení.

I zde je možné zeslabit požadavky a tak není třeba ani stejné rozdělení, ani úplná nezávislost. Jde však o velmi specializovaná tvrzení, tak se jim vyhneme a naučíme se používat CLV alespoň v tomto základním tvaru.