# Základy teorie pravděpodobnosti

Zdeněk Dvořák

6. listopadu 2018

**Příklad 1.** *Mějme ve třídě $n$ náhodných lidí. Jaká je pravděpodobnost, že dva z nich mají stejné narozeniny? (Ignorujeme 29.2. a fakt, že víc dětí se rodí v létě.)*

$365^n$ *všech možných dat narození,* $365 \cdot 365 \cdots (366-n)$ *navzájem různých dat. Pravděpodobnost, že všechna nejsou navzájem různá, je*

$$1 - \frac{365 \cdot 365 \cdots (366-n)}{365^n}.$$

*Pro* $n = 23$ *přibližně* $.507$*, pro* $n = 57$ *přibližně* $0.990$.

Nematematická otázka: co je vlastně pravděpodobnost?

## 1 Pravděpodobnostní prostory

Motivace: Jak interpretovat např. „zvolme náhodně bod z $\langle 0,1\rangle^2$“?

- Pro každý konkrétní bod je nulová pravděpodobnost, že se do něj trefíme.
- Idea: Můžeme ale o každé podmnožině $\langle 0,1\rangle^2$ říct, jaká je pravděpodobnost, že do ní vybraný bod patří – $P[\langle 0,1\rangle^2] = 1$, „pravděpodobnost, že druhá souřadnice je alespoň tak velká jako první“ je $P[\{(x,y) : x,y \in \langle 0,1\rangle, x \le y\}] = 0.5$, ...
- Nefuguje, akceptujeme-li axiom výběru; pravděpodobnost můžeme přidělit pouze „měřitelným“ podmnožinám.

**Definice 1.** *Pravděpodobnostní prostor je trojice* $(\Omega, \mathcal{F}, P)$*, kde* $\Omega$ *je množina elementárních jevů,* $\mathcal{F}$ *je* $\sigma$*-algebra na* $\Omega$*, a* $P : \mathcal{F} \to \langle 0,1\rangle$ *je* $\sigma$*-aditivní funkce tž.* $P[\Omega] = 1$.

Ve výše diskutovaném případě $\Omega = \langle 0,1\rangle^2$, $\mathcal{F}$ = měřitelné podmnožiny $\langle 0,1\rangle^2$. Nemáme nástroje (teorie míry) na to pracovat v takovéto obecnosti. Budeme se proto zabývat pouze speciálním případem, že $\Omega$ je konečná; pak není problém přiřadit pravděpodobnost každé podmnožině.

**Definice 2.** *Konečný pravděpodobnostní prostor je trojice $(\Omega, \mathcal{F}, P)$, kde*

- *$\Omega$ je konečná množina elementárních jevů,*
- *$\mathcal{F} = 2^\Omega$ je množina jevů,*
- *$P : \mathcal{F} \to \langle 0,1\rangle$ přiřazuje pravděpodobnost každému jevu a splňuje*
  - *Jsou-li $A, B \in \mathcal{F}$ disjunktní, pak $P[A \cup B] = P[A] + P[B]$.*
  - *$P[\Omega] = 1$.*

Pro elementární jev $x \in \Omega$ pišme $P[x]$ místo $P[\{x\}]$. Pro $A \in \mathcal{F}$ je doplňkový jev $\bar{A}$ definován jako $\Omega \setminus A$.

**Pozorování 1.** *Pro konečný pravděpodobnostní prostor $(\Omega, \mathcal{F}, P)$ a jev $A \in \mathcal{A}$ platí*

- *jestliže $B \subseteq A$, pak $P[B] \leq P[A]$,*
- *$P[\bar{A}] = 1 - P[A]$,*
- *$P[\emptyset] = 0$,*
- *$P[A] = \sum_{x \in A} P[x]$ pro každé $A \in \mathcal{F}$.*

Jev $A$ je nemožný, jestliže $P[A] = 0$, možný, jestliže $P[A] > 0$, jistý, jestliže $P[A] = 1$.

Konečný pravděpodobnostní prostor je uniformní, jestliže $P[x] = \frac{1}{|\Omega|}$ pro každé $x \in \Omega$; pak $P[A] = \frac{|A|}{|\Omega|}$ pro každý jev $A \in \mathcal{F}$.

**Příklad 2** (Pravděpodobnostní prostory)**.**

*Hod spravedlivou mincí:* $\Omega = \{rub, líc\}$, $P[rub] = P[líc] = 1/2$.

*Hod falešnou mincí, pravděpodobnost $p$ že padne rub:* $\Omega = \{rub, líc\}$, $P[rub] = p$, $P[líc] = 1 - p$

*Hod kostkou:* $\Omega = \{1, \ldots, 6\}$, $P[i] = 1/6$.

*Hod $k$ kostkami:* $\Omega = \{1, \ldots, 6\}^k$, $P[(i_1, \ldots, i_k)] = 1/6^k$.

*Výběr náhodné karty:* $\Omega$ = *karty, pro každou kartu $k$ je* $P[k] = 1/52$.

*Náhodné zamíchání balíčku karet:* $\Omega$ = *permutace karet, pro každou permutaci $\pi$ je* $P[\pi] = 1/52!$.

*Náhodný občan ČR:* $\Omega$ = *občané ČR, pro každého občana $o$ je* $P[o] = 1/|\Omega|$.

**Příklad 3** (Jevy).

*Hodíme dvěma kostkami, jev „větší z čísel je 3“ je*

$$\{(1,1),(1,2),(1,3),(2,1),(2,2),(2,3),(3,1),(3,2),(3,3)\},$$

*jeho pravděpodobnost je* $9/36 = 1/4$.

*Jev „náhodně vybraný občan ČR je muž“:* $M =$ *množina všech občanů ČR, kteří jsou muži,* $P[M] \approx 0.492$.

*Jev „v náhodně zamíchaném balíčku je na vrchu eso“:* $E =$*všechny permutace, v nichž je na vrchu eso,* $P[E] = \frac{4 \cdot 51!}{52!} = 1/13$.

*Hodíme 10-krát falešnou mincí (pravděpodobnost* $p$ *že padne rub).* $\Omega = \{rub, líc\}^{10}$, *pro* $x \in \Omega$ *obsahující* $r(x)$ *rubů a* $l(x)$ *líců je* $P[x] = p^{r(x)}(1-p)^{l(x)}$. *Jev „4-krát po sobě padne rub“:* $H =$*10-prvkové posloupnosti z* $\{rub, líc\}$ *obsahující souvislou podposloupnost 4 rubů.*

$$\begin{aligned} P[H] &= \sum_{x \in H} P[x] = \sum_{x \in H} p^{r(x)}(1-p)^{10-r(x)} \\ &= 7p^4(1-p)^6 + 36p^5(1-p)^5 + 75p^6(1-p)^4 + 80p^7(1-p)^3 + \\ &\quad 42p^8(1-p)^2 + 10p^9(1-p) + p^{10}. \end{aligned}$$

*Pro spravedlivou minci* $p = 0.5$ *máme* $P[H] \approx 0.245$, *pro falešnou minci s* $p = 0.55$ *máme* $P[H] \approx 0.329$.

# 2 Podmíněná pravděpodobnost

Pozorování: Pravděpodobnost, že zároveň nastanou jevy $A$ a $B$, je $P[A \cap B]$.

Provedli jsme nějakou náhodnou volbu a víme, že o ní platí $B$. Jaká je pravděpodobnost, že o ní platí i $A$?

**Definice 3.** *Nechť* $(\Omega, \mathcal{F}, P)$ *je pravděpodobnostní prostor a* $A, B \in \mathcal{F}$ *jsou jevy, kde* $B$ *je možný. Pak* <u>*pravděpodobnost* $A$ *za podmínky* $B$</u> *je*

$$P[A|B] = \frac{P[A \cap B]}{P[B]}.$$

**Příklad 4.** *Hodili jsme kostkou a padlo nám číslo menší než* 4. *Jaká je pravděpodobnost, že nám padlo sudé číslo?* $A = \{2,4,6\}$, $B = \{1,2,3\}$, $P[A \cap B] = P[\{2\}] = 1/6$, $P[B] = 1/2$,

$$P[A|B] = \frac{1/6}{1/2} = 1/3.$$

**Příklad 5.** *V krabici jsou 3 karty, jedna má rub i líc černý, druhá má rub i líc červený, třetí má černý rub a červený líc. Vytáhnu náhodně kartu, zeshora je červená. Jaká je pravděpodobnost, že je i zezdola červená?*

$\Omega = \{(k,s\} : k \in \{1,2,3\}, s \in \{rub, líc\}\}$*: číslo vytažené karty, je nahoře rub nebo líc? Uniformní,* $P[(k,s)] = 1/6$.

*Vytažená karta je zeshora červená:* $B = \{(2, rub), (2, líc), (3, líc)\}$, $P[B] = 1/2$.

*Vytažená karta je zdola červená:* $A = \{(2, rub), (2, líc), (3, rub)\}$.

*Vytažená karta je zeshora i zdola červená:* $A \cap B = \{(2, rub), (2, líc)\}$, $P[A \cap B] = 1/3$.

$$P[A|B] = \frac{P[A \cap B]}{P[B]} = \frac{1/3}{1/2} = 2/3.$$

**Lemma 2** (Bayes)**.** *Nechť* $(\Omega, \mathcal{F}, P)$ *je pravděpodobnostní prostor a* $A, B \in \mathcal{F}$ *jsou možné jevy. Pak*

$$P[A|B] = \frac{P[B|A] \cdot P[A]}{P[B]}.$$

*Důkaz.* Z definice

$$P[A|B] \cdot P[B] = P[A \cap B] = P[B|A] \cdot P[A].$$

□

**Lemma 3.** *Nechť* $B_1$, …, $B_k$ *jsou disjunktní možné jevy tž. jev* $\bigcup_{i=1}^k B_i$ *je jistý. Pak pro libovolný jev* $A$ *platí*

$$P[A] = \sum_{i=1}^{k} P[A|B_i] \cdot P[B_i].$$

*Speciálně, jestliže jev* $B$ *není nemožný ani jistý, pak*

$$P[A] = P[A|B] \cdot P[B] + P[A|\bar{B}] \cdot P[\bar{B}].$$

*Důkaz.* Z definice je $P[A|B_i] \cdot P[B_i] = P[A \cap B_i]$. Označme $X = \bigcup_{i=1}^k B_i$. Jelikož jevy $A \cap B_1$, …, $A \cap B_k$ jsou disjunktní, máme

$$P[A \cap X] = P\Big[\bigcup_{i=1}^{k} (A \cap B_i)\Big] = \sum_{i=1}^{k} P[A \cap B_i] = \sum_{i=1}^{k} P[A|B_i] \cdot P[B_i].$$

Dále

$$P[A \cap \bar{X}] \le P[\bar{X}] = 1 - P[X] = 0,$$

jelikož jev $X$ je jistý. Proto

$$P[A] = P[(A \cap X) \cup (A \cap \bar{X})] = P[A \cap X] + P[A \cap \bar{X}] = \sum_{i=1}^{k} P[A|B_i] \cdot P[B_i].$$

$\square$

**Příklad 6.** *Mamograf (vyšetření na rakovinu prsu) má false positive rate (pravděpodobnost, že o zdravé ženě prohlásí, že má rakovinu) asi* 10%. *Rakovinu prsu má cca* 0.2% 40*-letých žen. Jestliže náhodná* 40*-letá žena jde na mamografii a výsledek je pozitivní, jaká je pravděpodobnost, že přesto nemá rakovinu prsu?*

*$A = \{40$-leté ženy, nemající rakovinu prsu$\}$; $P[A] = 0.998$.*

*$B = \{40$-leté ženy, jimž by mamograf dal pozitivní výsledek$\}$.*

*False positive rate je $P[B|A] = 0.1$.*

*Dle Lemma 3 máme*

$$\begin{aligned} P[B] &= P[B|A] \cdot P[A] + P[B|\bar{A}] \cdot P[\bar{A}] \leq P[B|A] \cdot P[A] + P[\bar{A}] \\ &= P[B|A] \cdot P[A] + 1 - P[A] \approx 0.102. \end{aligned}$$

*Pravděpodobnost, že* 40*-letá žena nemá rakovinu prsu, za podmínky, že má pozitivní mamografii, je dle Bayesova vzorce*

$$P[A|B] = \frac{P[B|A] \cdot P[A]}{P[B]} \geq 0.980.$$