Pravděpodobnost a matematická statistika (NMSA202)

Petr Velička *
přednášející: doc. RNDr. Michal Pešta, Ph.D. †

LS 2024/25

*petrvel@matfyz.cz
†pesta@karlin.mff.cuni.cz

1 Náhodné jevy

Začneme nejdříve základními definicemi, bez nichž vůbec nemůžeme mluvit o pravděpodobnosti.

Definice 1.1. *Výběrovým prostorem* rozumíme množinu Ω všech možných výsledků nějakého experimentu. Prvky $\omega \in \Omega$ této množiny nazýváme *elementárními jevy.* Podmnožině $A \subset \Omega$ říkáme *(náhodný) jev.*

Pro ilustraci uvedeme následující motivační příklad, kde podrobně popíšeme souvislosti s právě zadefinovanými pojmy.

Příklad 1.2. Házíme dvakrát férovou mincí. Naším výběrovým prostorem bude množina $\Omega = \{PP, PO, OP, OO\}$. Událost, že první hod je panna, je tedy $A = \{PP, PO\}$. V tomto zápise písmeno P odpovídá tomu, že padla panna, kdežto písmeno O odpovídá orlu.

Dále uvažujme jevy H_1 – při prvním hodu padne panna, a H_2 – při druhém hodu padne panna. Nechť jsou všechny výsledky stejně pravděpodobné (jinými slovy, mince je férová), potom pravděpodobnost, že padne alespoň jedna panna (tj. nastane jev $H_1 \cup H_2$) je $\frac{3}{4}$.

Důkaz. Zřejmě z předchozího máme $H_1 = \{PP, PO\}$ a $H_2 = \{OP, PP\}$. Pravděpodobnost spočteme jako podíl velikosti $|H_1 \cup H_2| = 3$ a velikosti celého prostoru $|\Omega| = 4$. $\square$

Tato jednoduchá intuice však selže v případě nekonečné (nespočetné) množiny Ω, neboť jak již čtenář jistě ví z přednášky základů teorie míry, na nespočetné množině neexistuje "rozumný" způsob, jak měřit množiny. Musíme proto pracovat pouze s jistou třídou podmnožin Ω, které budeme říkat σ-algebra.

Definice 1.3. Nechť $\Omega \neq \emptyset$ je množina a $\mathcal{A} \subset 2^\Omega$ soubor jejích podmnožin. Této množině $\mathcal{A}$ říkáme σ-algebra, jestliže jsou splněny následující podmínky:

(i) $\emptyset \in \mathcal{A}$,

(ii) Pokud $A \in \mathcal{A}$, pak $A^C := \Omega \setminus A \in \mathcal{A}$,

(iii) Pokud $A_1, A_2, \dots \in \mathcal{A}$, pak $\bigcup_{i=1}^\infty A_i \in \mathcal{A}$.

Dvojici $(\Omega, \mathcal{A})$ nazýváme *měřitelný prostor.*

Každé události $A \in \mathcal{A}$ přiřadíme číslo $\mathbb{P}(A)$, které nazýváme *pravděpodobnost* jevu A. Jelikož chceme, aby se zachovala intuice z předchozího příkladu, musíme tuto představu náležitým způsobem formalizovat.

Definice 1.4. Nechť $(\Omega, \mathcal{A})$ je měřitelný prostor. Zobrazení $P : \mathcal{A} \to [0, 1]$ nazýváme *pravděpodobnostní mírou (pravděpodobností)*, jestliže:

(i) $P(\Omega) = 1$,

(ii) Pro libovolné po dvou disjunktní měřitelné množiny $A_i \in \mathcal{A}$, $i \in \mathbb{N}$ platí $P(\bigcup_{i=1}^\infty A_i) = \sum_{i=1}^\infty P(A_i)$.

Trojici $(\Omega, \mathcal{A}, P)$ nazýváme *pravděpodobnostní prostor*.

Přímo z této definice již můžeme odvodit pár základních vlastností pravděpodobnosti, se kterými dále budeme pracovat. Ve všech následujících tvrzeních pracujeme s pravděpodobnostním prostorem $(\Omega, \mathcal{A}, P)$.

Pozorování 1.5. ***(Základní vlastnosti pravděpodobnostní míry)*** *Pro výše jmenovaný pravděpodobnostní prostor platí následující tvrzení:*

1. $P(\emptyset) = 0$,
2. *Pro* $A, B \in \mathcal{A}$ *disjunktní platí* $P(A \cup B) = P(A) + P(B)$.
3. *Pro* $A \in \mathcal{A}$ *platí* $P(A^C) = 1 - P(A)$,
4. *Pro* $A, B \in \mathcal{A}, A \subset B$ *platí* $P(A) \leq P(B)$.

Důkaz.

1. Uvažujme posloupnost $A_1 = \Omega, A_2 = A_3 = \dots = \emptyset$. Potom z vlastnosti (ii) z definice máme, že $P(\Omega) = P(\Omega \cup \emptyset \cup \emptyset \dots) = P(\Omega) + \sum_{n=2}^{\infty} P(\emptyset)$. Tedy $\sum_{n=2}^{\infty} P(\emptyset) = 0$, což může nastat pouze v případě $P(\emptyset) = 0$ (jde o součet nekonečně mnoha nezáporných čísel).
2. Nechť $A_1 = A, A_2 = B, A_i = \emptyset$ pro $i > 2$. Tvrzení plyne přímo z vlastnosti (ii) z definice pravděpodobnostní míry a již dokázané vlastnosti 1.
3. $1 = P(\Omega) = P(A \cup A^C) = P(A) + P(A^C)$. Tato rovnost platí, neboť množina je vždy disjunktní se svým komplementem.
4. $P(B) = P(A \cup B \setminus A) = P(A) + P(B \setminus A)$. Jelikož funkce P je nezáporná, snadno vidíme, že $P(B) \geq P(A)$.

□

Lemma 1.6. ***(Pravděpodobnost sjednocení)*** *Pro libovolné* $A, B \in \mathcal{A}$ *platí* $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

Důkaz. Rozepíšeme $A \cup B = (A \cap B^C) \cup (A \cap B) \cup (A^C \cap B)$. Tyto tři množiny jsou zřejmě po dvou disjunktní. Dále díky aditivitě pravděpodobnosti máme $P(A \cup B) = P(A \cap B^C) + P(A \cap B) + P(A^C \cap B) + P(A \cap B) - P(A \cap B) = P(A) + P(B) - P(A \cap B)$. □

Věta 1.7. ***(Spojitost pravděpodobnosti)*** *Buď* $A_n \uparrow A$ *nebo* $A_n \downarrow A$ *pro* $A_n, A \in \mathcal{A}$. *Potom platí* $P(A_n) \to P(A)$.

Důkaz. Nechť $A_n \uparrow A$. Potom z definice $A_1 \subset A_2 \dots$ a platí $A = \bigcup_{i=1}^{\infty} A_i$. Definujme posloupnost B_n: $B_1 = A_1, B_n = A_n \setminus A_{n-1}$. Potom B_i jsou po dvou disjunktní a platí $A_n = \bigcup_{i=1}^{n} B_i$. Zřejmě také platí $A \equiv \bigcup_{n=1}^{\infty} A_n = \bigcup_{n=1}^{\infty} B_n$. Pak $P(A_n) = P(\bigcup_{i=1}^{n} B_i) = \sum_{i=1}^{n} P(B_i)$. Z toho již můžeme odvodit $\lim_{n\to\infty} P(A_n) = \lim_{n\to\infty} \sum_{i=1}^{n} P(B_i) = \sum_{i=1}^{\infty} P(B_i) = P(\bigcup_{i=1}^{\infty} B_i) = P(A)$.

Případ klesající A_n se dokáže analogicky, stačí uvažovat $C_n = A_n^C$. □

konec 1. přednášky (17.2.2025)

Uvedeme si ještě jeden příklad ilustrující intuitivní chápání pravděpodobnosti a zavedeme první takzvané pravděpodobnostní rozdělení. Uvažujme případ, že prostor Ω je konečný. Nechť všechny výsledky jsou stejně pravděpodobné, pak platí

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

V tomto případě mluvíme o *rovnoměrném rozdělení pravděpodobnosti*.

Příklad 1.8. *(Hod dvěma kostkami)* Výběrový prostor $\Omega = \{(i, j) : i, j \in \{1 \dots 6\}\}$ má 36 prvků. Jestliže všechny výsledky jsou stejně pravděpodobné, pak platí $P(A) = \frac{|A|}{36}$. Například, pravděpodobnost toho, že součet na kostkách je přesně 11, je $2/36$, protože pouze dva výsledky $(5, 6)$ a $(6, 5)$ odpovídají této události.

V praxi často chceme odlišit, zda pravděpodobnost výskytu jedné události nějakým způsobem závisí na výskytu jiné události. K tomu nám poslouží pojem nezávislosti jevů.

Definice 1.9. Dvě události $A, B \in \mathcal{A}$ jsou *nezávislé*, jestliže platí $P(A \cap B) = P(A)P(B)$. Obdobně, množina událostí $\{A_i : i \in I\}$ (kde indexová množina I je nejvýše spočetná) je nezávislá, jestliže platí

$$P\left(\bigcap_{j \in J} A_j\right) = \prod_{j \in J} P(A_j)$$

pro každou konečnou podmnožinu $J \subset I$.

Je důležité si uvědomit, že disjunktní události s kladnou pravděpodobností nejsou nezávislé (neboť součin jejich pravděpodobností není roven 0 – pravděpodobnost výskytu jejich prázdného průniku). Obecně se pracuje se dvěma typy nezávislosti – předpokládanou (plyne z podstaty zkoumané úlohy) a odvozenou (dokázaná pomocí jiných vlastností úlohy). Následující příklad ilustruje praktické použití právě zavedeného pojmu.

Příklad 1.10. Házíme férovou mincí 10krát. Nechť A je událost "padla aspoň jedna panna". Pak platí $P(A) = 1 - (1/2)^{10}$.

Důkaz. Nechť T_j je událost, že při j-tém hodu padne orel. Můžeme psát $P(A) = 1 - P(A^C) = 1 - P(\text{samé orly}) = 1 - P(T_1 \cap \dots \cap T_{10})$. Dále díky nezávislosti (v tomto případě jde o nezávislost předpokládanou) jevů T_j máme $1 - P(T_1 \cap \dots \cap T_{10}) = 1 - P(T_1) \cdots P(T_{10}) = 1 - (1/2)^{10} \approx 0.999$. □

Dalším silným nástrojem v teorii pravděpodobnosti je podmíněná pravděpodobnost, která nám poskytuje odpověď na otázku "Pokud vím, že nastala událost B, jaká je pravděpodobnost události A?".

Definice 1.11. Mějme jevy $A, B \in \mathcal{A}$. Pokud $P(B) > 0$, pak *podmíněná pravděpodobnost* A za podmínky B je definována vztahem

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

Poznamenejme si několik základních vlastností podmíněné pravděpodobnosti, jejichž důkaz snadno plyne z příslušných definic.

Pozorování 1.12. ***(Vlastnosti podmíněné pravděpodobnosti)***

(i) *Pro pevné* $B \in \mathcal{A}, P(B) > 0$ *je* $P(\cdot|B)$ *pravděpodobnostní míra.*

(ii) *Obecně platí* $P(A|B) \neq P(B|A)$, *platí totiž* $P(A|B) = P(B|A)\frac{P(A)}{P(B)}$ *(pokud obě strany rovnosti dávají smysl).*

(iii) *Události* A *a* B *jsou nezávislé právě tehdy, když* $P(A|B) = P(A)$ *(předpokládáme nenulovost* $P(B)$*).*

(iv) $P(A \cap B) = P(A|B)P(B) = P(B|A)P(A)$ *v případě, že* $P(A)P(B) > 0$.

Důkaz. Vlastnosti (iii) a (iv) plynou přímo z definice vynásobením vhodnou konstantou.

Vlastnost (ii) se dokáže následujícím protipříkladem, uvažujme hod dvěma férovými mincemi. Nechť H_1 je událost "padla aspoň jedna panna" a H_2 událost "padly dvě panny". Potom $P(H_1|H_2) = 1$ ale $P(H_2|H_1) = \frac{1}{3}$. Důkaz obecného vztahu je ponechán čtenáři jako snadné (ale užitečné) cvičení.

Nakonec, vlastnost (i) je důsledkem toho, že pro libovolnou množinu $A \in \mathcal{A}$ je $A \cap B$ měřitelná, a navíc pro libovolný systém po dvou disjunktních množin $A_i, i \in \mathbb{N}$ platí $P(\bigcup_{i=1}^{\infty} A_i|B) = \frac{1}{P(B)} P\left((\bigcup_{i=1}^{\infty} A_i) \cap B\right) = \frac{1}{P(B)} P\left(\bigcup_{i=1}^{\infty}(A_i \cap B)\right) = \frac{1}{P(B)} \sum_{i=1}^{\infty} P(A_i \cap B) = \sum_{i=1}^{\infty} P(A_i|B)$. □

Použití podmíněné pravděpodobnosti v praxi však někdy může vést k neintuitivním výsledkům, které ilustruje následující příklad.

Příklad 1.13. Uvažujme nemoc D a test, který má dva možné výsledky. Pravděpodobnosti výsledků tohoto testu jsou uvedeny v následující tabulce. Zde sloupce odpovídají přítomnosti/absenci nemoci a řádky výsledkům testu.

	D	D^C
$+$	0.009	0.099
$-$	0.001	0.891

Z definice spočteme následující podmíněné pravděpodobnosti:

$$P(+|D) = \frac{P(+ \cap D)}{P(D)} = \frac{0.009}{0.009 + 0.001} = 0.9.$$

$$P(-|D^C) = \frac{P(- \cap D^C)}{P(D^C)} = \frac{0.891}{0.891 + 0.099} \approx 0.9.$$

Vychází nám, že test je docela přesný, neboť nemocní lidé mají test v 90% případů pozitivní, stejně tak zdraví lidé jsou v 90% případů negativní.

Dále předpokládejme, že pacient šel na test a získal pozitivní výsledek. Spočteme, s jakou pravděpodobností je opravdu nakažený.

$$P(D|+) = \frac{P(D \cap +)}{P(+)} = \frac{0.009}{0.009 + 0.099} \approx 0.08.$$

Vyšlo nám, že na první pohled zdánlivě precizní test ve skutečnosti má méně než 10% úspěšnost. Jedním z důvodů této diskrepance může být například velký nepoměr zdravých lidí vůči nakaženým (pouze jedno procento) ve zdrojových datech, což je jev který se obecně vyskytuje u většiny nemocí. V praxi se proto často pracuje s domněnkami – například testujeme jen pacienty, kteří vykazují nějaké symptomy apod.

Na závěr uvedeme dvě velmi užitečné věty, které se často používají v nejrůznějších úlohách a týkají se podmíněné pravděpodobnosti. Zformulujeme je pro spočetné rozklady, ale obdobná tvrzení platí i pro konečné rozklady s velmi podobným důkazem.

Věta 1.14. ***(Zákon úplné pravděpodobnosti)*** *Nechť $A_1, A_2, \dots$ je spočetný disjunktní rozklad Ω takový, že $P(A_i) > 0$ pro každé $i \in \mathbb{N}$. Potom pro libovolnou událost $B \in \mathcal{A}$ platí:*

$$P(B) = \sum_{i=1}^{\infty} P(B|A_i)P(A_i).$$

Důkaz. Definujme posloupnost množin $C_i = B \cap A_i$ pro $i \in \mathbb{N}$. Zjevně $\{C_i, i \in \mathbb{N}\}$ je disjunktní pokrytí B. Potom $P(B) = \sum_{i=1}^{\infty} P(C_i) = \sum_{i=1}^{\infty} P(B \cap A_i) = \sum_{i=1}^{\infty} P(B|A_i)P(A_i)$. □

Věta 1.15. ***(Bayes)*** *Nechť $A_1, A_2, \dots$ je spočetný disjunktní rozklad Ω takový, že $P(A_i) > 0$ pro každé $i \in \mathbb{N}$. Mějme událost $B \in \mathcal{A}$ s nenulovou pravděpodobností. Potom platí:*

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{\sum_{j=1}^{\infty} P(B|A_j)P(A_j)}.$$

Důkaz. Přímým výpočtem dostáváme

$$P(A_i|B) = \frac{P(A_i \cap B)}{P(B)} = \frac{P(B|A_i)P(A_i)}{P(B)} = \frac{P(B|A_i)P(A_i)}{\sum_{j=1}^{\infty} P(B|A_j)P(A_j)},$$

kde poslední rovnost získáme aplikací *Věty 1.14*. □

Použití Bayesovy věty si ukážeme na následujícím příkladu.

Příklad 1.16. Uvažujme e-mailovou schránku. Máme tři kategorie e-mailů: A_1 – spam, A_2 – nízká priorita, A_3 – vysoká priorita. Na základě předchozích zkušeností víme, že $P(A_1) = 0.7$, $P(A_2) = 0.2$, $P(A_3) = 0.1$. Nechť B je událost, že daný e-mail obsahuje slovo "zdarma". Platí $P(B|A_1) = 0.9, P(B|A_2) =$

$0.01, P(B|A_3) = 0.01$[1]. Jaká je pravděpodobnost, že příchozí e-mail obsahující slovo "zdarma" je spam?

Přímým výpočtem z Bayesovy věty získáme

$$P(A_1|B) = \frac{0.9 \cdot 0.7}{0.9 \cdot 0.7 + 0.01 \cdot 0.2 + 0.01 \cdot 0.1} = 0.995.$$

Tedy pravděpodobnost, že tento e-mail je spam je přes 99%!

[1] Tyto hodnoty se nutně nemusí sečíst na 1

2 Náhodné veličiny

V této kapitole se budeme věnovat náhodným veličinám, což bude formalizovat (a zobecňovat) jakýsi intuitivní chápání toho, že nějaká proměnná nabývá různých hodnot s určitými pravděpodobnostmi. Začneme ústřední definicí celé statistiky – náhodnou veličinou.

Definice 2.1. Nechť $(\Omega, \mathcal{A})$ je měřitelný prostor. *Náhodná veličina* je měřitelné zobrazení, které přiřazuje každému výsledku ω reálné číslo $X(\omega)$. Jinými slovy, $\{\omega \in \Omega : X(\omega) \leq x\} \in \mathcal{A} \forall x \in \mathbb{R}$.

konec 2. přednášky (18.2.2025)