

Pravděpodobnost a Statistika 1

Poznámky z přednášek

Letní semestr 2020/2021

Viktor Soukup, Lukáš Salak

Obsah

1	První přednáška	2
2	Druhá přednáška	5

1 První přednáška

Modely náhody \rightarrow Pravděpodobnost \rightarrow Pozorovaná data \rightarrow Modely náhody

Model náhody např. kostka $1, \dots, 6$,

Pozorovaná data : $1, 5, 4, 3, 3$

otázka na pravděpodobnost: jaká je pravděpodobnost... hodně pozorovaných dat \rightarrow statistika na model náhody.

Příklad (Schwartz-Zippel algoritmus): Máme dány dva polynomy $f(x), g(x)$ stupně d . Chceme zjistit, zda jsou stejné, a to co nejrychleji.

Problém: $g(x)$ je součin několik polynomů stupně $\leq \frac{d}{4}$, dostáváme víc než lineární čas.

Řešení: Algoritmus: zvolíme náhodně $x \in \{1, 2, \dots, 100d\}$, ověříme, zda $f(x_1) = g(x_1)$. Když $f \neq g$, tak x_1 je kořen polynomu $f - g$ takových x_1 je $\leq d$.

$$P(f(x_1) = g(x_1) : f \neq g) \leq \frac{1}{100}$$

Pokud jsme spokojeni s 1%, končíme, když ne, volíme $x_2, x_3, \dots \in \{1, 2, \dots, 100d\}$, pak

$$P(\text{Pro } x_1, x_2, x_3 \dots f(x_i) = g(x_i) : f \neq g) \leq \left(\frac{1}{3}\right)^3 = 10^{-6}$$

... aproximační algoritmy

Některé jevy neumíme/nechceme popsat kauzálně

1. hod kostkou
2. tři hody kostkou, nekonečně mnoho hodů kostkou
3. hod šípkou na terč
4. počet emailů za den
5. dobu běhu programu (v reálnem počítači)...

Důvody:

- fyzikální vlastnost přírody
- komplikovaný proces (počasí, medicína, molekuly plynu...)
- neznáme vlivy (působení dalších lidí, programů...)
- randomizované algoritmy (test prvočíselnosti, quicksort)
- náhodné grafy (Ramseyovy čísla)

Pro popis pomocí teorie pravděpodobnosti napřed vybereme množinu elementárních jevů Ω (sample space)

$$\Omega = \{1, 2, \dots, 6\} = [6] \implies \text{hod kostkou}$$

$$\Omega = [6]^3 \implies \text{hod třemi kostkami}$$

Definice (Prostor jevů): $\mathcal{F} \subseteq \mathcal{P}(\Omega)$

$\mathbb{F} \subseteq \mathbb{P}(\Omega)$ je prostor jevů (též σ -algebra), pokud

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

Často $\mathcal{F} = \mathcal{P}(\Omega)$, to je možné vždy, když je Ω spočetná, např. pro $\Omega = \mathbb{R}$ to již nejde.

Definice (Pravděpodobnost): $P : \mathcal{F} \rightarrow [0, 1]$ se nazývá pravděpodobnost (probability), pokud:

1. $P(\emptyset) = 0, P(\Omega) = 1$, a
2. $P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$, pro libovolnou posloupnost po dvou *disjunktních* jevů **TODOOT**

Šance (odds) jevu A je $O(A) = \frac{P(A)}{P(A^c)}$. Např. šance na výhru je 1 ku 2 znamená, že pravděpodobnost výhry je $\frac{1}{3}$; šance, že na kostce padne šestka je 1 ku 5.

Konvence:

- „ A je jistý jev“ znamená $P(A) = 1$. Také se říká, že A nastává skoro jistě (almost surely), zkráceně s.j. (a.s.).
- „ A je nemožný jev“ znamená $P(A) = 0$.

$$P(A) = 0 \Rightarrow^? A = \emptyset$$

\leftarrow axiom

\rightarrow platí často, ne vždy

- Např. $A = \text{střed kruhu (házení šipek na terč)} \implies P(A) = 0$ B spočetná (konečná, velká jako \mathbb{N}) množina:

$$P(B) = 0 + 0 + 0 + \dots = 0$$

B_i je i -tý bod, $B = \bigcup B_i$

Věta (Vlastnosti pravděpodobnostního prostoru): *V pravděpodobnostním prostoru (Ω, \mathcal{F}, P) platí pro $A, B \in \mathcal{F}$:*

1. $P(A) + P(A^c) = 1$
2. $A \subseteq B \implies P(A) \leq P(B)$
3. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
4. $P(A_1 \cup A_2 \cup \dots) \leq \sum_i P(A_i)$ (subaditiva, Booleova nerovnost) (nevyžadujeme disjunktnost, pak by platila rovnost)

Důkaz:

1. $\Omega = A \cup A^c$; A, A^c disj.,
 $1 = P(\Omega) = P(A) + P(A^c)$
2. $P(B) = P(A) + P(B \setminus A) \geq P(A)$
3. cvičení **TODOOT**

4. trik zdisjunktnění: z $A_1, A_2 \dots$ uděláme disjunktní množiny

$$B_1 = A_1, B_2 = A_2 \setminus A_1, B_3 = A_3 \setminus A_1 \cup A_2 \dots$$

$$B_i \subseteq A_i \implies P(B_i) \leq P(A_i)$$

$$B_i \cap B_j = \emptyset : j < i \dots B_i \cap B_j \subseteq B_i \cap A_j = \emptyset$$

$$\bigcup_{i=1}^{\infty} B_i = \bigcup_{i=1}^{\infty} A_i$$

$$\subseteq \text{ok}$$

opačná inkluze **TODOOT**

$$P\left(\bigcup A_i\right) = P\left(\bigcup B_i\right) = \sum P(B_i) \leq \sum P(A_i).$$

□

Příklad (Pravděpodobnostní prostory):

1. Konečný s uniformní pravděpodobností

Ω je libovolná konečná množina, $\mathcal{F} = \mathcal{P}(\Omega)$,

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

2. Diskrétní

$\Omega = \{\omega_1, \omega_2 \dots\}$ je libovolná spočetná množina. Jsou dány $p_1, p_2 \dots \in [0, 1]$ se součtem 1.

$P(A) = \sum_{i: \omega_i \in A} p_i$ (cinknutá loterie, nějaké možnosti mají jiné procenta)

3. Spojitý

$\Omega \subseteq \mathbb{R}^d$ pro vhodné d (Ω např. uzavřená nebo otevřená)

\mathcal{F} vhodná (obsahuje např. všechny otevřené množiny)

$f: \Omega \rightarrow [0, 1]$ je funkce taková, že $\int_{\Omega} f(x) dx = 1$.

$$P(A) = \int_A f(x) dx$$

Speciální případ: $f(x) = 1/V_d(\Omega)$

$P(A) = \frac{V_d(A)}{V_d(\Omega)}$, kde $V_d(A) = \int_A 1$ je d -rozměrný objem A .

4. Bernoulliho krychle - nekonečné opakování

$\Omega = S^{\mathbb{N}}$, kde S je diskrétní s pravděpodobností Q ,

\mathcal{F} vhodná (obsahuje např. všechny množiny tvaru

$$A = A_1 \times \dots \times A_k \times S \times S \times \dots)$$

$$P(A) = Q(A_1) \dots Q(A_k)$$

Příklad (Nepříklady):

1. Náhodné přirozené číslo: můžeme si vybrat mnoha způsoby, Ale všechna přirozená čísla nemají stejnou pravděpodobnost.

není možné, aby měly všechny stejnou nenulovou pravděpodobnost, protože pokud $P(0) = P(1) = P(2) \dots = P$ tak $P(\mathbb{N}) = p + p + p \dots = \infty$.

2. Náhodné reálné číslo

3. Betranův paradox

Definice (Podmíněná pravděpodobnost): Pokud $A, B \in \mathcal{F}$ a $P(B) > 0$, pak definujeme podmíněnou pravděpodobnost A při B (probability of A given B) jako

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

$Q(A) := P(A|B)$. Pak (Ω, \mathcal{F}, Q) je pravděpodobnostní prostor.

Definice (Zřetězené podmínování): $P(A \cap B) = P(B)P(A|B)$

Věta: Pokud $A_1, \dots, A_n \in \mathcal{F}$ a $P(A_1 \cap \dots \cap A_n) > 0$, tak

$$P(A_1 \cap A_2 \cap \dots \cap A_n) = P(A_1)P(A_2|A_1) \dots \text{TODOOT}$$

2 Druhá přednáška

Poznámka: Opakování

1. definice pravděpodobnostního prostoru (Ω, \mathcal{F}, P) : dva axiomy,
2. **naivní** pravděpodobnostní prostor: Ω konečná, $\mathcal{F} = \mathcal{P}(\Omega)$
 $P(A) := |A|/|\Omega|$
3. **diskrétní** pravděpodobnostní prostor: $\Omega = \omega_1, \omega_2, \dots$,
 $\mathcal{F} = \mathcal{P}(\Omega), \sum_i p_i = 1$
 $P(A) := \sum_{i: \omega_i \in A} p_i$
4. **geometrický** pravděpodobnostní prostor:
 $\omega \subseteq \mathbb{R}^d$ s konečným objemem,
 $P(A) := V_d(A)/V_d(\Omega)$
5. pravděpodobnostní prostor **spojitý s hustotou**:
 $\Omega \subseteq \mathbb{R}^d$ s funkcí f , kde $\int_{\Omega} f = 1$,
 $P(A) := \int_A f$

V pravděpodobnostním prostoru (Ω, \mathcal{F}, P) platí pro $A, B \in \mathcal{F}$

1. $P(A^c) = 1 - P(A) \dots (A^c = \Omega \setminus A)$
2. $A \subseteq B \implies P(A) \leq P(B)$
3. $P(A \cup B) = P(A) + P(B) - P(A \cap B) \dots$ PIE
4. $P(A_1 \cup A_2 \cup \dots) \leq \sum_i P(A_i)$ (subaditivita, Booleova nerovnost)
5. Definujeme podmíněnou pravděpodobnost (pro $P(B) > 0$).

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

6. $Q(A) = P(A|B)$ splňuje axiomy pro pravděpodobnost.

$$P(\emptyset|B) = 0$$

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

$$P(A_1 \circ A_2|B) = \frac{P((A_1 \circ A_2) \cup B)}{P(B)} = \frac{P((A_1 \cap B) \cup (A_2 \cap B))}{P(B)} = \frac{P(A_1 \cap B)}{P(B) + \frac{P(A_2 \cap B)}{P(B)}} = P(A_1|B) + P(A_2|B)$$

Definice: Zřetězené podmínování

$$P(A \cup B) = P(B)P(A|B)$$

Věta: Pokud $A_1, \dots, A_n \in \mathcal{F}$ a $P(A_1 \cap \dots \cap A_n) > 0$, tak

$$P(A_1 \cap A_2 \cap \dots \cap A_n) =$$

$$P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2) \dots P(A_n | \bigcap_{i=1}^{n-1} A_i)$$

Důkaz: indukcí

□

Příklad: Příklad: vytáhneme 3 karty z balíčku 52 karet. Jaká je $P(\text{žádné srdce})$?

A_i = i-tá karta není srdce

$$P(A_1 \cap A_2 \cap A_3) = P(A_1) \times P(A_2|A_1)P(A_3|A_2 \cap A_1) = \frac{13 * 3}{52} \times \frac{13 * 3 - 1}{51} \times \frac{13 * 3 - 2}{50}$$

$$\frac{\# \text{dobrých}}{\# \text{všech}} = \frac{\binom{39}{3}}{\binom{52}{3}}$$

Definice: Spočetný systém množin $B_1, B_2, \dots \in \mathcal{F}$ je rozklad (partition) Ω , Pokud

1. $B_i \cap B_j = \emptyset$ pro $i \neq j$ a
2. $\bigcup_i B_i = \Omega$.

Věta: Věta o úplné pravd. = Rozbor všech možností

Pokud B_1, B_2, \dots je rozklad Ω a $A \in \mathcal{F}$, tak

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

(sčítance s $P(B_i) = 0$ považujeme za 0).

$$A = (A \cap B_1) \cup (A \cap B_2) \cup \dots$$

(sjednocení disjunktních množin)

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

Příklad: Máme tři mince: P+O, P+P, O+O. Jaká je pravděpodobnost, že padne orel?

Označíme M_1, M_2, M_3 pro P+O, P+P, O+O.

$$\begin{aligned} P(O) &= P(M_1)P(O|M_1) + P(M_2)P(O|M_2) + P(M_3)P(O|M_3) \\ &= \frac{1}{3} \times \frac{1}{2} + \frac{1}{3} \times 0 + \frac{1}{3} \times 1 = \frac{1}{2} \end{aligned}$$

Rychlejší je vypsát si strom a pak posčítat výsledné jevy

Příklad: Gambler's ruin - zbankrotování hazardního hráče.

Máme a korun, náš protihráč b korun. Hrajeme opakovaně spravedlivou hru o 1kč, dokud někdo nepřijde o všechny peníze. Jaká je pravděpodobnost, že vyhraje? **Důkaz:**

$$P_a = P(\text{z této pozice vyhraje})$$

$$P_0 = 0, P_n = 1 \dots (a + b = n)$$

$$P(\text{výhra} | 1. \text{ kolo výhra}) P(1. \text{ kolo výhra})$$

$$+ P(\text{výhra} | 1. \text{ kolo prohra}) P(1. \text{ kolo prohra})$$

$$\text{výhra} \implies P_{a+1}, \text{prohra} \implies P_{a-1}$$

$$P_a = \frac{P_{a+1}}{2} + \frac{P_{a-1}}{2}$$

$$\text{ekvivalencia } \iff \text{pls}$$

$$P_a - P_{a-1} = P_{a+1} - P_a = \Delta$$

$$1 = P_n = P_0 + n * \Delta \implies \Delta = \frac{1}{n}$$

$$P_a = \frac{a}{a+b} = \frac{a}{n}$$

□

Definition: Bayesova Věta

Pokud B_1, B_2, \dots je rozklad Ω , $A \in \mathcal{F}$, $P(A) > 0$ a $P(B_j) > 0$, tak

$$P(B_j|A) = \frac{P(B_j)P(A|B_j)}{P(A)} = \frac{P(B_j)P(A|B_j)}{\sum_i P(B_i)P(A|B_i)}.$$

(sčítance s $P(B_i) = 0$ považujeme za 0).

Důkaz:

$$P(B_j|A)P(A) = P(B_j)P(A|B_j)$$

$$P(A \cap B_j) = P(B_j \cap A)$$

□

Příklad: N = nemocný, T = testovaný, specif. $P(N|T)$, sens. $P(T|N)$.

$$P(N|T) = \frac{P(N)P(T|N)}{P(N)P(T|N) + P(N^c)P(T|N^c)} = \frac{p * 0.8}{p * 0.8 + (1 - p) * 0.01}$$

$$p = 0.001 \dots 7\%$$

$$p = 0.0016 \dots 56\% \dots \text{momentální stav testování}$$

$$p = 0.05 \dots 80\%$$

Definition: Jevy $A, B \in \mathcal{F}$ jsou nezávislé (independenent) pokud $P(A \cap B) = P(A)P(B)$.

Pak také platí $P(A|B) = P(A)$, pokud $P(B) > 0$.

Definition: Jevy $\{A_i : i \in I\}$ jsou (vzájemně) nezávislé, pokud pro každou konečnou množinu $J \subseteq I$

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

Pokud podmínka platí jen pro dvouprvkové množiny J , nazýváme jevy $\{A_i\}$ po dvou nezávislé (pairwise independenent).

Definice: Necht' pro množiny z prostoru jevů platí

$$A_1 \subseteq A_2 \subseteq A_3 \subseteq \dots$$

a $A = \bigcup_{i=1}^{\infty} A_i$. Pak platí

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

Důkaz:

$$A = A_1 \cup (A_2 \setminus A_1) \cup (A_3 \setminus A_2) \cup \dots$$

$$P(A) = P(A_1) + P(A_2 \setminus A_1) + P(A_3 \setminus A_2) + \dots$$

$$\lim_{i \rightarrow \infty} (P(A_1) + \dots + P(A_i \setminus A_{i-1})) = \lim_{i \rightarrow \infty} P(A_i).$$

□

$A_n \subset P, O^{\mathbb{N}}, A_n =$ mezi prvními n hody padl aspoň jednou orel.

$$P(A) = P(\geq 1 \text{ orel v } \infty \text{ hodech}) = \lim_{n \rightarrow \infty} \dots = 1$$

Definice: Náhodná veličina

Mějme pravděpodobnostní prostor (Ω, \mathcal{F}, P) . Funkci $X : \Omega \rightarrow \mathbb{R}$ nazveme diskrétní náhodná veličina, pokud $I_m(X)$ je spočetná množina a pokud pro všechna reálna x platí

$$\{x \in \Omega : X(\omega) = x\} \in \mathcal{F}.$$

Definice: Pravděpodobnostní funkce diskrétní náhodné veličiny X je funkce $p_X : \mathbb{R} \rightarrow [0, 1]$ taková, že

$$p_X(x) = P(X = x) = P(\{x \in \Omega : X(\omega) = x\})$$

Definice: $\sum_{x \in I_m(X)} p_X(x) = 1$

Definice: $S := I_m(X)1, Q(A) := \sum_{x \in A} p_X(x)$

$(S, \mathcal{P}(S), Q)$ je diskrétní pravděpodobnostní prostor.

Definice: Pro $S = \{s_i : i \in I\}$ spočetnou množinu reálných čísel a $c_i \in [0, 1]$ $\sum_{i \in I} c_i = 1$ existuje pravděpodobnostní prostor a diskrétní n.v. X na něm taková, že $p_X(s_i) = c_i$ pro $i \in I$.

The End