

Pravděpodobnost a Statistika 1

Poznámky z přednášek

Letní semestr 2020/2021

Viktor Soukup, Lukáš Salak

Obsah

1 První přednáška	2
-------------------	---

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

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

... 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 reálném počítači)...

Důvody:

1. fyzikální vlastnost přírody?
2. komplikovaný proces (počasí, medicína, molekuly plynu...)
3. neznáme vlivy (působení dalších lidí, programů...)
4. randomizované algoritmy (test prvočíselnosti, quicksort)
5. náhodné grafy (Ramsayovy čísla)

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

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

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

Definice: Prostor jevů

$$\mathbb{F} \subseteq \mathbb{P}(\Omega)$$

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

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

$$1. \emptyset \in \mathcal{F} \implies \Omega \setminus A \in \mathcal{F}, a$$

2. TODOOT

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

$$1. P(\emptyset) = 0, P(\Omega) = 1, a$$

$$2. P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i), \text{ pro libovolnou posloupnost po dvou } \textit{disjunktních} \text{ jevů } \textbf{TODOOT}$$

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.

Příklad: "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 $P(A) = 0$.

$$P(A) = 0 \implies ? A = \emptyset$$

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

$$1. P(A) + P(A^c) = 1$$

$$2. A \subseteq B \implies P(A) \leq P(B)$$

$$3. P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

$$4. P(A_1 \cup A_2 \cup \dots) \leq \sum_i P(A_i) \text{ (subaditivita, Booleova nerovnost) (nevyžadujeme disjunktnost, pak by platila rovnost)}$$

Dk:

$$1. \Omega = A \cup A^c; A, A^c \text{ disj.}, \\ 1 = P(\Omega) = P(A) + P(A^c)$$

$$2. P(B) = P(A) + P(B \setminus A) \geq P(A)$$

3. cvičení **TODOOT**

4. trik zdisjunktnení: 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**

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

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

Definition: 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.

Definition: 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

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

The End