NMAI059 Pravděpodobnost a statistika 1
1. přednáška

Robert Šámal

Přehled

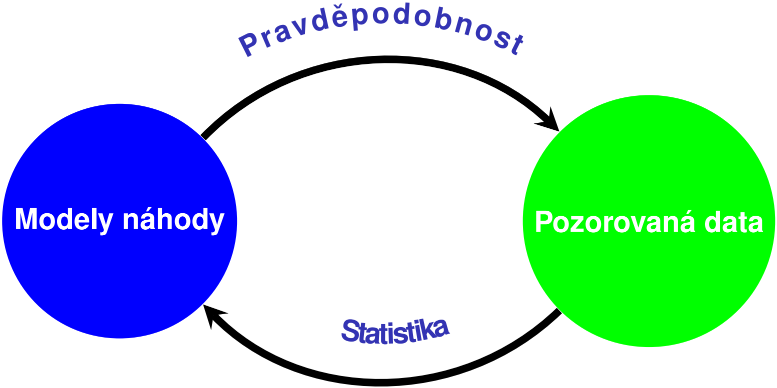
Organizace přednášky

- Přednášky v Zoomu, dokud to bude potřeba (možná celý semestr!). Přednáška má svoji stránku v Moodle (odkaz v SISu). Tam bude všechno.
- Nedojde-li k technickým komplikacím, bude video přednášky dostupné (po přihlášení do SISu).
- Kdyby vám vadilo být nahráni, můžete vypnout svoji kameru, případně dotazy klást v chatu.
- Používejte též funkce Zoomu – přihlásit se, zpomalit-zrychlit, atd.
- Během přednášky budeme používat krátké ankety.
- Pdf verze „tabule“ bude též k dispozici – už před přednáškou.
- Zkouška bude v ideálním případě prezenční písemka s možností ústního dozkoušení.

Organizace cvičení

- Též setkání v Zoomu, z rozhodnutí fakulty se bude nahrávat.
- Kdyby to někomu vadilo, ozvěte se, můžeme mít např. jedno cvičení nahrávané a druhé ne.
- Nicméně budu vyžadovat aktivní účast, analogickou přítomnosti na normálním cvičení. Pokud tomu u někoho brání technické či jiné důvody, ozvěte se. 1/4
- Pravidelné domácí úkoly (jedno krátké cvičení). Odevzdávání v Moodlu. Ideálně pdf (sken papíru, nebo „sken“ pomocí vhodné mobilní aplikace). 1/4
- Zápočtová písemka cca v polovině semestru (z pravděpodobnostní části přednášky). 1/4
- Zápočtová práce – pro statistickou část přednášky, zpracovaná pomocí programu v jazyce R (příp. Python). 25 %
- V Moodlu je také prostor pro diskuzi, jak (ne)funguje technologie. Případně se ozvěte emailem.

Plán přednášky



Přehled

Organizace

Pravděpodobnost – úvod

Podmíněná pravděpodobnost

Bonusy

Pravděpodobnost – intuice, definice

Některé jevy neumíme/nechceme popsat kauzálně:

- hod kostkou $\Omega = \{1, 2, \dots, 6\}$
- tři hody kostkou, nekonečně mnoho hodů kostkou $\{1, \dots, 6\}^3$, $\{1, \dots, 6\}^{\mathbb{N}}$
- hod šipkou na terč $\subseteq \mathbb{R}^3$
- počet emailů za den $\mathbb{N}_0$
- dobu běhu programu (v reálném počítači) $\mathbb{R}$

Důvody:

- fyzikální vlastnost přírody?
- komplikovaný proces (počasí, medicína, molekuly plynu)
- neznámé vlivy (působení dalších lidí, programů, . . .)
- randomizované algoritmy (test prvočíselnosti, quicksort)
- náhodné grafy (odhady Ramseyových čísel)

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

Prostor jevů

Dále vybereme *prostor jevů (event space)* $\mathcal{F} \subseteq \mathcal{P}(\Omega)$, u kterých budeme měřit jejich pravděpodobnost.
Často $\boxed{\mathcal{F} = \mathcal{P}(\Omega)}$, to je možné vždy, když Ω je spočetná. Ale např. pro $\Omega = \mathbb{R}$ to už nejde. $\Omega = [0,1]$

Definice

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

- $\emptyset \in \mathcal{F}$ *a* $\Omega \in \mathcal{F}$, A^c
- $A \in \mathcal{F} \Rightarrow \Omega \setminus A \in \mathcal{F}$, *a*
- $A_1, A_2, \ldots \in \mathcal{F} \Rightarrow \bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$.

$(A_3 = A_4 = \ldots = \emptyset)$

$A_1 \cup A_2 \in \mathcal{F}$

$A_1 \cap A_2 \in \mathcal{F}$

$A_1 \setminus A_2 \in \mathcal{F}$

Axiomy pravděpodobnosti

Definice

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

- $P(\emptyset) = 0$, $P(\Omega) = 1$, *a*
- $P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$, *pro libovolnou posloupnost po dvou disjunktních jevů $A_1, A_2, \ldots \in \mathcal{F}$.*

Definice

Pravděpodobnostní prostor (probability space) je trojice $(\Omega, \mathcal{F}, P)$ taková, že

- $\Omega \neq \emptyset$ *je libovolná množina,*
- $\mathcal{F} \subseteq \mathcal{P}(\Omega)$ *je prostor jevů, a*
- P *je pravděpodobnost.*

Názvosloví

pravděpodobnost

- *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 $1/3$; šance, že na kostce padne šestka je $1/5$.

- „*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 $\boxed{P(A) = 0}$.

$\Leftarrow \; A = \emptyset$

$\not\Rightarrow$

Ω = kruh v rovině

S

$A = \{$střed kruhu$\}$

$P(A) = 0$

$P(S)$ malé

$\neq 0$

Základní vlastnosti

$A_1, A_2, \dots \in \mathcal{F}$

Věta

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

1. $P(A) + P(A^c) = 1 \qquad (A^c = \Omega \setminus A)$
2. $A \subseteq B \Rightarrow P(A) \leq P(B)$
3. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ (A_i nemusí být disj.)
4. $P(A_1 \cup A_2 \cup \dots) \leq \sum_i P(A_i)$ (subaditivita, Booleova nerovnost)

Dk

① $\Omega = A \cup A^c$ A, A^c disj.

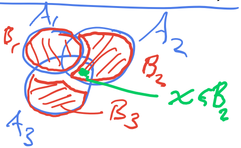
$1 = P(\Omega) = P(A) + P(A^c)$

② $B = A \cup (B \setminus A)$

$P(B) = P(A) + \underbrace{P(B \setminus A)}_{\geq 0} \geq P(A)$

③ podobně, cvič.

④ trik zdisjunktnění



$B_1 = A_1, \; B_2 = A_2 \setminus A_1, \dots$

$\boxed{B_i = A_i \setminus \bigcup_{j<i} A_j}$

- $B_i \subseteq A_i$ $\Rightarrow B_i \cap B_j \subseteq B_i \cap A_j = \emptyset$
- $B_i \cap B_j = \emptyset$ když $j < i$
- $\bigcup B_i = \bigcup A_i$ ($\subseteq$, $\supseteq$)

$\boxed{P(\bigcup A_i)} = P(\bigcup B_i) = \sum P(B_i) \leq \sum P(A_i)$

Příklady pravděpodobnostních prostorů 1

- **Konečný s uniformní pravděpodobností**
  Ω je libovolná konečná množina, $\mathcal{F} = \mathcal{P}(\Omega)$,
  $P(A) = |A|/|\Omega|$.
- **Diskrétní**
  $\Omega = \{\omega_1, \omega_2, \dots\}$ je libovolná spočetná množina. Jsou dány $p_1, p_2, \dots \in [0, 1]$ se součtem 1.
  $$P(A) = \sum_{i:\omega_i \in A} p_i$$

konečná nebo velká jako $\mathbb{N}$

$$\sum_{i=1}^{\infty} p_i = 1$$

$$p_i = \frac{1}{N}$$

$$\sum_{i} p_i = 1$$

$$\sum_{i=1}^{N} \qquad \sum_{i=1}^{\infty}$$

Příklady pravděpodobnostních prostorů 2

- **Spojitý**
  $\Omega \subseteq R^d$ pro vhodné d (Ω např. uzavřená nebo otevřená)
  $\mathcal{F}$ vhodná (obsahuje např. všechny otevřené množiny)
  $P(A) = V_d(A)/V_d(\Omega)$,
  kde $V_d(A) = \int_A 1$ je d-rozměrný objem A.
- **Bernoulliho krychle – nekonečné opakování**
  $\Omega = S^{\mathbb{N}}$, kde S je diskrétní s pstí Q,
  $\mathcal{F}$ vhodná (obsahuje např. všechny množiny tvaru
  $A = A_1 \times \cdots \times A_k \times S \times S \times \cdots$
  $P(A) = Q(A_1) \cdots Q(A_k)$

  Př.: $\{0, 1\}^{\mathbb{N}}$ nekonečné házení mincí

$P($první tři hody PPOP$) = \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2}$

Nepříklady

- **Náhodné přirozené číslo** můžeme vybrat mnoha způsoby. V přednášce poznáme geometrické a Poissonovo rozdělení. Nemůžeme ale požadovat, aby všechna přirozená čísla měla stejnou pravděpodobnost. (Proč?) „Náhodné přirozené číslo je sudé s pravd. $1/2$.“ ???
- **Náhodné reálné číslo** Opět není žádný preferovaný způsob, jak definovat pravděpodobnost pro $\Omega = \mathbb{R}$. Typicky bude každé reálné číslo mít pravděpodobnost 0! Navíc nejde definovat pravděpodobnost tak, aby nezáležela na posunu, tj. $P([0,1]) = P([1,2]) = \dots$

  $\Omega = [0,1]$
- **Náhodná tětiva kružnice – Bertrandův paradox** Vybereme náhodnou tětivu zadané kružnice. Jaká je pravděpodobnost, že její délka je větší, než strana vepsaného rovnostranného trojúhelníku?

Přehled

Podmíněná pravděpodobnost

Definice

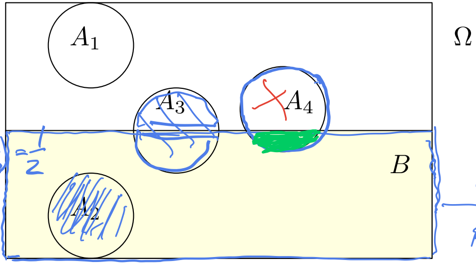
Pokud $A, B \in \mathcal{F}$ a $\boxed{P(B) > 0}$, pak definujeme podmíněnou pravděpodobnost A při B (probability of A given B) jako

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

$= A_2 \cap B$

$P(A_2 \mid B) = \frac{P(A_2)}{P(B)} \leq 1$

$P(B \mid A_2) = 1 = \frac{P(A_2 \cap B)}{P(A_2)}$



$P(B) = \frac{1}{2}$

$P(A_2 \mid B) = 2 \cdot P(A_2)$

$P(B \mid A_3) = \frac{1}{2}$ ALE

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

$P(A_3 \mid B) = \frac{1}{20}$

Zřetězené podmiňování

- $P(A \cap B) = P(B)P(A \mid B)$ ... důsledek defce

Věta

Pokud $A_1, \ldots, A_n \in \mathcal{F}$ *a* $P(A_1 \cap \cdots \cap A_n) > 0$, *tak*

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

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

Dk

$$\cancel{P(A_1)} \frac{\cancel{P(A_1 \cap A_2)}}{\cancel{P(A_1)}} \frac{P(A_1 \cap A_2 \cap A_3)}{\cancel{P(A_1 \cap A_2)}} \cdots \frac{P(A_1 \cap \cdots \cap A_n)}{P(A_1 \cap \cdots \cap A_{n-1})}$$

Pří balíček 32 karet, táhneme 3 karty,

A_i := "i-tá karta je srdce"

$$P(A_1 \cap A_2 \cap A_3) = P(A_1) \cdot P(A_2 \mid A_1)\, P(A_3 \mid A_1 \cap A_2)$$

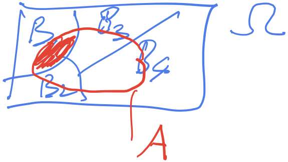
$$\frac{8}{32} \cdot \frac{7}{31} \cdot \frac{6}{30}$$

Rozbor všech možností

Definice

Spočetný systém množin $B_1, B_2, \ldots \in \mathcal{F}$ je rozklad (partition) Ω, pokud

- $B_i \cap B_j = \emptyset$ *pro* $i \neq j$ *a*
- $\bigcup_i B_i = \Omega$.



Věta

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

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

$A = \bigcup_i (A \cap B_i)$

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

$P(A) = \sum_i P(A \cap B_i) = \sum_i P(A / B_i) \cdot P(B_i)$

Rozbor všech možností

Př: 3 mince

1. P&O
2. P&P
3. O&O

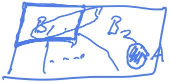
Vybereme náhodně jednu z mincí a tou pak hodíme.

M_i := vybereme i-tou minci

A : padla panna

$$P(A) = P(A|M_1)\cdot P(M_1) + P(A|M_2)P(M_2) + P(A|M_3)P(M_3) = \frac{1}{2}\cdot\frac{1}{3} + 1\cdot\frac{1}{3} + 0\cdot\frac{1}{3} = \boxed{\frac{1}{2}} = \frac{3}{6}$$

Bayesova věta



Ω $\quad A \subseteq B_j,\ j=2$

$P(B_j \mid A) = 1$

Věta

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

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

$P(A \mid B_i) = 0$ pro $i \neq j$

$= 1$

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

Dk

$$P(B_j \mid A) = \frac{P(B_j \cap A)}{P(A)} = \frac{\cancel{P(A \mid B_j)} \cdot P(B_j)}{P(A)}$$

Př

$$\boxed{P(M_1 \mid A)}$$

dos. 2 minulé věty

$$= \frac{P(A \mid M_1) \cdot P(M_1)}{\sum_{i=1}^{3} P(A \mid M_i) \cdot P(M_i)} = \frac{\frac{1}{2} \cdot \frac{1}{3}}{\frac{1}{2}} = \boxed{\frac{1}{3}}$$

Bayesova věta

$p=0.1 \quad \frac{0.06}{0.06+0.01} = \frac{6}{7} = 85\%$

T = vyšel pos. test

C = máme chorobu

$P(T|C)$ = sensitivita pravé pos. $= \underline{0.63}$

$\boxed{P(T^c|C^c)}$ = specificita pravé neg. $= \underline{0.99}$

$$P(C|T) = \frac{P(T|C)\cdot P(C)}{P(T|C)P(C) + \underline{P(T|C^c)}\, P(C^c)}$$

$$= \frac{0.63 \cdot p}{0.63\, p + \cancel{0.99}\,(1-p)}$$

0.01

$(p = 0.0001)$ $\frac{0.000063}{0 + 0.01} \doteq 0$

$(p = 0.01)$ $\frac{0.0063}{\underline{0.0063} + \underline{0.01}\cdot 0.99} = \frac{6}{16}$ $= 36\%$

$p=0.1$

Nezávislost jevů

Definice

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

- Pak také $P(A \mid B) = P(A)$, pokud $P(B) > 0$.

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A)\,P(B)}{P(B)} = P(A)$$

Př.: nez.

A ... poprvé P

B ... podruhé P

C ... právě jednou P

	P	O
P	PP	PO
O	OP	OO

A

C

$$P(A \cap B) = \frac{1}{4} = P(A) \cdot P(B)$$

$$P(A \cap C) = \frac{1}{4} = P(A)\,P(C)$$

Nezávislost více jevů

množ. indexů, lib. velká

Definice

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

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

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

A, B, C z minulé tabule

A, B ... nez.

A, C ... nez.

B, C ... nez.

A, B, C ... $P(A \cap B \cap C) \neq \frac{1}{8}$

Přehled

Spojitost pravděpodobnosti

Věta

Nechť pro množiny z prostoru jevů platí

$$A_1 \subseteq A_2 \subseteq A_3 \subseteq \cdots$$

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

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

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

Borel-Cantelliho lemma

Věta

Nechť jevy $A_1, A_2, \ldots$ splňují $P(A_i) = p_i > 0$ pro každé i. Označme Nic jev „nenastal žádný z jevů $\{A_i\}$“ a Inf jev „nastalo nekonečně mnoho z jevů $\{A_i\}$“.

1. *Pokud $\sum_i p_i < \infty$, tak $P(Inf) = 0$.*
2. *Pokud jsou jevy $A_1, A_2, \ldots$ nezávislé a $\sum_i p_i = \infty$, tak $P(Nic) = 0$, $P(Inf) = 1$.*