NMAI059 Pravděpodobnost a statistika 1
1. přednáška

Robert Šámal

Přehled

Organizace

Pravděpodobnost – úvod

Podmíněná pravděpodobnost

Bonusy

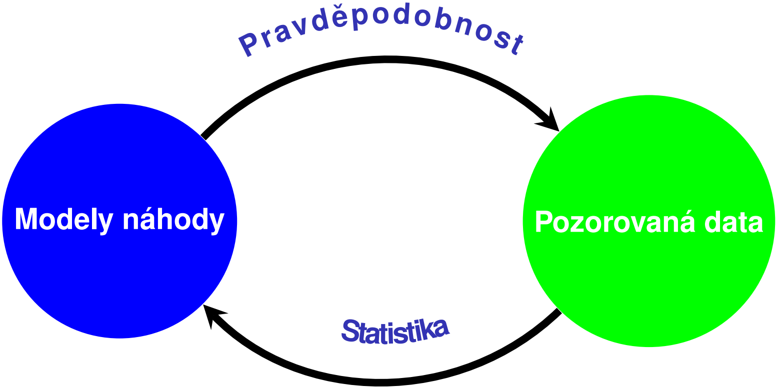
Organizace přednášky

- 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.
- Pravidelné domácí úkoly (jedno krátké cvičení). Odevzdávání v Moodlu. Ideálně pdf (sken papíru, nebo „sken“ pomocí vhodné mobilní aplikace).
- Zápočtová písemka cca v polovině semestru (z pravděpodobnostní části přednášky).
- Zápočtová práce – pro statistickou část přednášky, zpracovaná pomocí programu v jazyce R (příp. Python).
- 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
- tři hody kostkou, nekonečně mnoho hodů kostkou
- hod šipkou na terč
- počet emailů za den
- dobu běhu programu (v reálném počítači)

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

Definice

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

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

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

- *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 $P(A) = 0$.

Základní vlastnosti

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. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
4. $P(A_1 \cup A_2 \cup \dots) \leq \sum_i P(A_i)$ *(subaditivita, Booleova nerovnost)*

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.
  $P(A) = \sum_{i:\omega_i \in A} p_i$

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

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. $P([0,1]) = P([1,2]) = \dots$
- **Náhodná tětiva kružnice – Bertrandův paradox** Vybereme náhodnou tětivu zadané kružnice. Jaká je pravděpodobnost, že její délka je větší, než strana vepsaného rovnostranného trojúhelníku?

Přehled

Organizace

Pravděpodobnost – úvod

Podmíněná pravděpodobnost

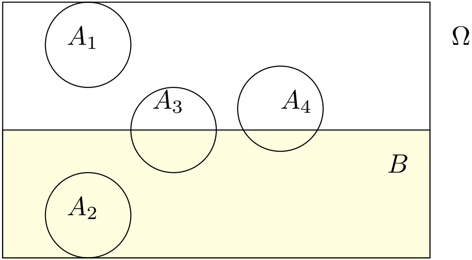
Bonusy

Podmíněná pravděpodobnost

Definice

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

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



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

Zřetězené podmiňování

- $P(A \cap B) = P(B)P(A \mid B)$

Věta

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

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

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

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

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

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

Rozbor všech možností

Bayesova věta

Věta

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

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

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

Bayesova věta

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 $P(B) > 0$.

Nezávislost více jevů

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).*

Přehled

Organizace

Pravděpodobnost – úvod

Podmíněná pravděpodobnost

Bonusy

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 $P(Nic) = 0$, $P(Inf) = 1$.*