NMAI059 Pravděpodobnost a statistika 1
2. přednáška

Robert Šámal

Co už víme

jevy, které pozorujeme

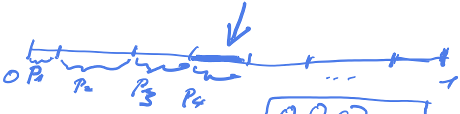
- definice pravděpodobnostního prostoru $(\Omega, \mathcal{F}, P)$: dva axiomy (1) $P(\Omega)=1$, $P(\emptyset)=0$ (2) $P(A_1 \cup A_2 \cup \dots) = P(A_1) + \dots$ A_i disj.
- **naivní** pravděpodobnostní prostor: Ω konečná, $\mathcal{F} = \mathcal{P}(\Omega)$, $P(A) := |A|/|\Omega|$

  $\frac{|A_1 \cup A_2 \cup \dots|}{|\Omega|} = \frac{|A_1|}{|\Omega|} + \frac{|A_2|}{|\Omega|} + \dots$

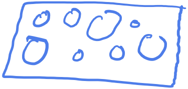
váhy

- **diskrétní** pravděpodobnostní prostor: $\Omega = \{\omega_1, \omega_2, \dots\}$, $\mathcal{F} = \mathcal{P}(\Omega)$, $\sum_i p_i = 1$

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



- **geometrický** pravděpodobnostní prostor: $\Omega \subseteq \mathbb{R}^d$ s konečným objemem, $P(A) := V_d(A)/V_d(\Omega)$



váha

- pravděpodobnostní prostor **spojitý s hustotou**: $\Omega \subseteq \mathbb{R}^d$ s funkcí f, kde $\int_\Omega f = 1$, $P(A) := \int_A f$

Co už víme: Základní vlastnosti

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

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

zob. P.I.E.

- Definujeme podmíněnou pravděpodobnost (pro $P(B) > 0$).

prosecutor's fallacy

$P(\text{důkazy} \mid \text{nevinen})$ malá

$P(\text{nevinen} \mid \text{důkazy})$

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

- $Q(A) = P(A \mid B)$ splňuje axiomy pro pravděpodobnost

B fixn.

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

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

disj.:

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

Přehled

Zřetězené podmiňování

- $P(A \cap B) = P(B)P(A \mid B)$ (dosl. def.) $= P(A) \cdot P(B \mid A)$

Věta

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

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

$A_1 \cap A_2 \cap \cdots A_k \supseteq A_1 \cap \cdots \cap A_n$

$P(\cdots \cap \cdots) > 0$

definováno

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

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

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

$13 \cdot 3 = 39$

- Příklad: vytáhneme 3 karty z balíčku 52 karet. Jaká je P(žádné srdce)? A_i = i-tá karta není srdce

$$P(A_1 \cap A_2 \cap A_3) = P(A_1) \cdot P(A_2 \mid A_1)\, P(A_3 \mid A_1 \cap A_2) = \frac{39}{52} \cdot \frac{38}{51} \cdot \frac{37}{50}$$

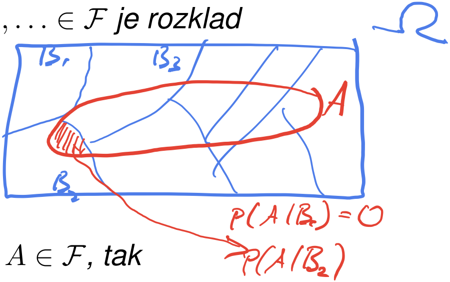
Věta o úplné pravd. = Rozbor všech možností

$B_1, \dots B_n$

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

$$\boxed{P(A) = \sum_i P(B_i) P(A \mid B_i)}$$

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

$A = (A \cap B_1) \cup (A \cap B_2) \cup \ldots$ (tj. disj. množin)

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

Věta o úplné pravd. = Rozbor všech možností

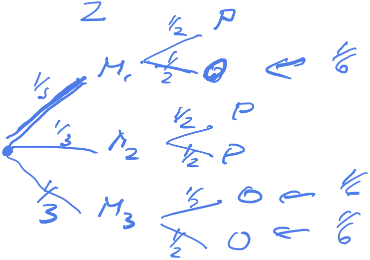
- 1. aplikace. Máme tři mince: $\boxed{P+O}$, P+P, O+O. Jaká je pravděpodobnost, že padne orel?

M_1 M_2 M_3

$$P(O) = P(M_1)\cdot P(O|M_1) + P(M_2)\cdot P(O|M_2) + P(M_3)\cdot P(O|M_3)$$

$$= \frac{1}{3}\cdot\frac{1}{2} + \frac{1}{3}\cdot 0 + \frac{1}{3}\cdot 1$$

$$= \frac{1}{2}$$



Věta o úplné pravd. = Rozbor všech možností

- 2. aplikace. Gambler's ruin – zbankrotování hazardního hráče. $P(\text{prohr.}) = \frac{b}{a+b} \Rightarrow P(\text{nek. hra}) = 0$
  Máme a korun, náš protihráč b korun. Hrajeme opakovaně spravedlivou hru o 1 Kč, dokud někdo nepřijde o všechny peníze. Jaká je pravděpodobnost, že vyhrajeme?

$$\boxed{p_a = \frac{a}{a+b} = \frac{a}{n}}$$

0   naše a protihráč $n = a+b$

$0 \le a \le n$

$p_a = P(\text{z této pozice vyhrajeme})$

$$\boxed{p_0 = 0, \quad p_n = 1} \quad \text{2 rovnice}$$

$$= P(\text{výhra} \mid \text{1. kolo výhra}) \cdot P(\text{1. kolo výhra}) \quad \text{„}p_{a+1}\text{"} \quad = \frac{1}{2}$$

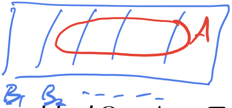
$$+ \underbrace{P(\text{výhra} \mid \text{1. kolo prohra})}_{p_{a-1}} \cdot P(\text{1. kolo prohra})$$

$$\boxed{p_a = \frac{p_{a-1}}{2} + \frac{p_{a+1}}{2}} \quad n-1 \text{ rovnic}$$

$$\boxed{p_a - p_{a-1} = p_{a+1} - p_a} = \Delta$$

$$1 = p_n = p_0 + n \cdot \Delta \Rightarrow \Delta = \frac{1}{n}$$

Bayesova věta



pozorování

Věta

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

„stav světa“ víme

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

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

Dk

$$P(B_j \mid A) \cdot P(A) = P(B_j) \cdot P(A \mid B_j)$$

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

→ dosadíme za $P(A)$ z věty o úplné pr. VÚP

Bayesova věta

N = nemocný

T = test +

specif. $\cancel{P(N/T)}$ $P(T^c/N^c)$

senzit. $P(T/N)$

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

false positive rate

$p = 0.001$ 7%

$p = 0.01$ ... 56%

$p = 0.05$ ... 80%

Nezávislost jevů

$\frac{1}{2} \cdot \frac{1}{2}$ $\frac{1}{4}$

A, B nezávislé

„z definice pravd. prostoru“

Definice

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

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

$\frac{P(A \cap B)}{P(B)}$

$\Omega = \{OO, OP, PO, PP\}$

Příklad: Hodíme dvakrát mincí. Označme

- $A = \{\omega \in \Omega : \omega_1 = P\}$ = „poprvé padla panna“
- $B = \{\omega \in \Omega : \omega_2 = P\}$ = „podruhé padla panna“
- $C = \{\omega \in \Omega : \omega_1 \neq \omega_2\}$ = „padla právě jedna panna“

$P(A) = \frac{2}{4} = \frac{1}{2}$

$P(B) = P(C)$

$P(\{\omega \in \Omega, \ldots\}) = P(\text{poprvé panna})$

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

$= P(A \cap C) = P(B \cap C)$

Nezávislost více jevů

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

Definice

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

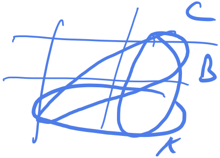
nestačí

$P($

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

$J = I$

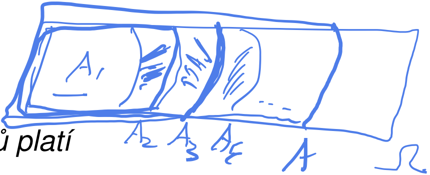
Pokud podmínka platí jen pro dvouprvkové množiny J, nazýváme jevy $\{A_i\}$ po dvou nezávislé (pairwise independent).



jsou po dvou nezávislé

$P(A\cap B\cap C) = 0 \neq P(A)\,P(B).P(C)$

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*

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

Dk $A = A_1 \cup (A_2 \setminus A_1) \cup (A_3 \setminus A_2) \cup \dots$ disj. nezol!

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

$= \lim_{i\to\infty} \left(P(A_1) + \dots + P(A_i \setminus A_{i-1})\right) = \lim_{i\to\infty} P(A_i)$ ∎

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

$P(A) = P(\geq 1$ orel v ∞ hodech$) = \lim_{n\to\infty} P(A_n) = \lim_{n\to\infty} \left(1 - \frac{1}{2^n}\right) = 1$

Přehled

Náhodná veličina/proměnná

Často nás zajímá číslo dané výsledkem náhodného pokusu.

- Hodíme na terč a změříme vzdálenost od středu.
- Házíme kostkou, dokud nepadne šestka, ale pak si všimneme jenom toho, kolik hodů to trvalo.
- U quicksortu (algoritmus na třídění) měříme počet kroků (v závislosti na náhodné volbě pivotu).

libovolný

Definice

Mějme pravděpodobnostní prostor $(\Omega, \mathcal{F}, P)$*. Funkci* $X : \Omega \to \mathbb{R}$ *nazveme diskrétní náhodná veličina (discrete random variable), pokud* $Im(X)$ *(obor hodnot* X*) je spočetná množina a pokud pro všechna reálná* x *platí*

$$X^{-1}(\omega) = \underbrace{\{\omega \in \Omega : X(\omega) = x\}}_{\text{chceme měřit}} \in \mathcal{F}.$$

Pravděpodobnostní funkce

Definice

Pravděpodobnostní funkce (probability mass function, pmf) diskrétní náhodné veličiny X je funkce $p_X : \mathbb{R} \to [0,1]$ taková, že

$$p_X(x) = \underbrace{P(X = x)}_{\text{zkratka}} = P(\underbrace{\{\omega \in \Omega : X(\omega) = x\}}_{\text{přesný zápis}})$$

- $\sum_{x \in \underbrace{Im(X)}_{\text{spoč.}}} p_X(x) = 1$

  $= \sum_{x} P(\{\omega\in\Omega : X(\omega)=x\}) = P\left(\bigcup_{x\in Im X}\{\omega\in\Omega : X(\omega)=x\}\right) = P(\Omega) = 1$
- $S := Im(X) \qquad Q(A) := \sum_{x \in A} p_X(x) = P(X \in A)$
  $(S, \mathcal{P}(S), Q)$ je diskrétní pravděpodobnostní prostor.
- Pro $S = \{s_i : i \in I\}$ spočetnou množinu reálných čísel a $c_i \in [0,1]$ splňující $\sum_{i \in I} c_i = 1$ existuje pravděpodobnostní prostor a diskrétní n.v. X na něm taková, že $p_X(s_i) = c_i$ pro $i \in I$.

Přehled

Podmíněná pravděpodobnost

Diskrétní náhodné veličiny

Příklady diskrétních n.v.

Střední hodnota

Bonusy

Bernoulliho/alternativní rozdělení

- $X =$ počet orlů při jednom hodu nespravedlivou mincí.
- Značíme $X \sim Bern(p)$. (Někdy se značí $Alt(p)$.)

- Dáno $p \in [0, 1]$.
- $p_X(1) = p$
- $p_X(0) = 1 - p$
- $p_X(k) = 0$ pro $k \neq 0, 1$

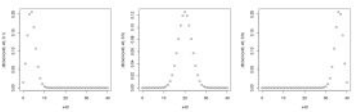
- Pro libovolný jev $A \in \mathcal{F}$ definujeme *indikátorovou n.v.* I_A:
- $I_A(\omega) = 1$ pokud $\omega \in A$, $I_A(\omega) = 0$ jinak.
- $I_A \sim Bern(P(A))$

Binomiální rozdělení

- $X =$ počet orlů při n hodech nespravedlivou mincí.
- Značíme $X \sim Bin(n, p)$.

- Dáno $p \in [0, 1]$.
- $p_X(k) = \binom{n}{k} p^k (1-p)^{n-k}$ pro $k \in \{0, 1, \ldots, n\}$

Binomiální rozdělení: pravděpodobnostní funkce



Vygenerováno následujícím kódem v R

```
x40 <- 0:40
plot(x40, dbinom(x40, 40, 0.1))
plot(x40, dbinom(x40, 40, 0.5))
plot(x40, dbinom(x40, 40, 0.9))
```

Poissonovo rozdělení

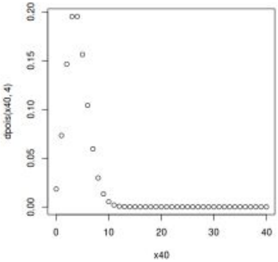
- Značíme $X \sim Pois(\lambda)$.

- Dáno reálné $\lambda > 0$.
- $p_X(k) = \frac{\lambda^k}{k!} e^{-\lambda}$

- $Pois(\lambda)$ je limitou $Bin(n, \lambda/n)$

- X popisuje např. počet emailů, které dostaneme za jednu hodinu.

Poissonovo rozdělení: pravděpodobnostní funkce



Vygenerováno následujícím kódem v R

```
x40 <- seq(0, 40,by=1)
plot(x40, dpois(x40, 4))
```

Poissonovo paradigma

- $A_1, \ldots, A_n$ jsou (skoro-)nezávislé jevy s $P(A_i) = p_i$, $\lambda = \sum_i p_i$. Nechť n je velké, každé z p_i malé. Pak přibližně platí
$$\sum_{i=1}^{n} I_{A_i} \sim Pois(\lambda).$$

Geometrické rozdělení

- $X =$ kolikátým hodem mincí padl první orel.
- Značíme $X \sim Geom(p)$.

- Dáno $p \in [0, 1]$.
- $p_X(k) = (1-p)^{k-1}p$

- Někdy se tomuto rozdělení říká posunuté geometrické, a za normální geometrické se považuje rozdělení $X - 1$, tj. počet neúspěšných hodů.