NMAI059 Pravděpodobnost a statistika 1
2. přednáška

Robert Šámal

Co už víme

- definice pravděpodobnostního prostoru $(\Omega, \mathcal{F}, P)$: dva axiomy
- **naivní** pravděpodobnostní prostor: Ω konečná, $\mathcal{F} = \mathcal{P}(\Omega)$, $P(A) := |A|/|\Omega|$
- **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, $P(A) := V_d(A)/V_d(\Omega)$
- 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)

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

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

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

Přehled

Podmíněná pravděpodobnost

Diskrétní náhodné veličiny

Příklady diskrétních n.v.

Střední hodnota

Bonusy

Zřetězené podmiňování

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

Věta

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

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

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

- Příklad: vytáhneme 3 karty z balíčku 52 karet. Jaká je P(žádné srdce)?

Věta o úplné pravd. = 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

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

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

Věta o úplné pravd. = Rozbor všech možností

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

Věta o úplné pravd. = Rozbor všech možností

- 2. aplikace. Gambler's ruin – zbankrotování hazardního hráče.
  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?

Bayesova věta

Věta

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

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

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

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 $P(B) > 0$.

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"

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).*

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.

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).

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*

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

Pravděpodobnostní funkce

Definice

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

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

- $\sum_{x \in Im(X)} p_X(x) =$?
- $S := Im(X) \qquad Q(A) := \sum_{x \in A} p_X(x)$
  $(S, \mathcal{P}(S), Q)$ 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

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 $k \neq 0, 1$

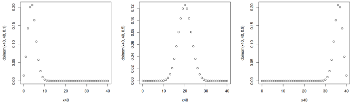
- Pro libovolný jev $A \in \mathcal{F}$ definujeme *indikátorovou n.v.* I_A:
- $I_A(\omega) = 1$ pokud $\omega \in A$, $I_A(\omega) = 0$ jinak.
- $I_A \sim Bern(P(A))$

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



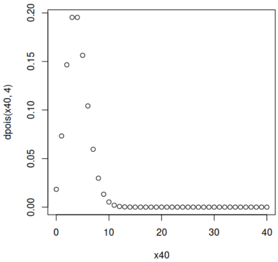
Vygenerováno následujícím kódem v R

```
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.

Přehled

Střední hodnota

Definice

Pokud X je diskrétní n.v., tak její střední hodnota (expectation) je označována $\mathbb{E}(X)$ a definována

$$\mathbb{E}(X) = \sum_{x \in Im(X)} x \cdot P(X = x),$$

pokud součet má smysl.

LOTUS

- Pro reálnou funkci g a diskrétní n.v. X je $Y = g(X)$ také diskrétní n.v.

Věta (LOTUS)

Pokud X je diskrétní n.v. a g reálná funkce, tak

$$\mathbb{E}(g(X)) = \sum_{x \in Im(X)} g(x) P(X = x)$$

pokud součet má smysl.

Vlastnosti $\mathbb{E}$

Věta

Nechť X, Y jsou diskrétní n.v. a $a, b \in \mathbb{R}$.

1. *Pokud $P(X \geq 0) = 1$ a $\mathbb{E}(X) = 0$, tak $P(X = 0) = 1$.*
2. *Pokud $\mathbb{E}(X) \geq 0$ tak $P(X \geq 0) > 0$.*
3. $\mathbb{E}(a \cdot X + b) = a \cdot \mathbb{E}(X) + b$.
4. $\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y)$.

Podmíněná střední hodnota

Definice

Pokud X je diskrétní n.v. a $P(B) > 0$, tak podmíněná střední hodnota X za předpokladu B (conditional expectation of X given B) je

$$\mathbb{E}(X \mid B) = \sum_{x \in Im(X)} x \cdot P(X = x \mid B),$$

pokud součet má smysl.

Rozbor všech možností

Věta

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

$$\mathbb{E}(X) = \sum_i \mathbb{E}(X \mid B_i) P(B_i),$$

kdykoliv má součet smysl. (Sčítance s $P(B_i) = 0$ považujeme za 0.)

Rozbor všech možností

Přehled

Bertrand's paradox

Simpsons's paradox