

## 3. cvičení

### Bayes

#### Úloha 1 (Spam or ham?)

Petr dostává hodně emailů, ale 80 % z nich jsou spamy. Jeho spamový filtr 90 % spamů správně označí, ale také 5 % řádných emailů označí jako spam.

1. Kolik procent emailů bude označeno jako spamy?
2. Kolik procent řádných emailů je mezi těmi, co jsou označené jako spamy?
3. Kolik procent spamů je mezi emaily, které testem prošly?

#### Řešení

Značíme jevy:  $O$ : „označený jako spam“,  $S$ : „je spam“.

1.  $P(O) = P(O|S) \cdot P(S) + P(O|\neg S) \cdot P(\neg S) = 0.9 \cdot 0.8 + 0.05 \cdot 0.2 = 0.73$
2.  $P(\neg S|O) = \frac{P(O|\neg S) \cdot P(\neg S)}{P(O)} = \frac{0.05 \cdot 0.2}{0.73} = \frac{1}{73}$
3.  $P(S|\neg O) = \frac{P(\neg O|S) \cdot P(S)}{P(\neg O)} = \frac{0.1 \cdot 0.8}{0.27} = \frac{8}{27}$

### Nezávislost jevů

#### Úloha 2 (Nezávislost a doplňky)

Ukažte, že jsou-li jevy  $A, B$  nezávislé, pak jsou nezávislé i  $A, B^c$  a také  $A^c, B^c$ .

#### Řešení

$$P(A \cap B^c) = P(A) - P(A \cap B) = P(A) - P(A) \cdot P(B) = P(A) \cdot (1 - P(B)) = P(A) \cdot P(B^c).$$

Dva doplňky analogicky pro  $A = B^c, B = A$ .

#### Úloha 3 (Nezávislost a disjunktnost)

Mohou být dva jevy nezávislé a zároveň disjunktní?

#### Řešení

Pokud  $A, B$  jsou disjunktní, pak  $P(A \cap B) = 0$ , a tedy pro nezávislost nutně musí platit  $P(A)P(B) = 0$ , což nastane jen tehdy, když  $P(A) = 0 \vee P(B) = 0$ .

Tedy ano, můžou, ale jen tehdy, když alespoň jeden z těchto jevů je nemožný (má nulovou pravděpodobnost).

#### Úloha 4 (Nezávislost a podmnožiny)

Mohou být jevy  $A, B$  nezávislé a zároveň  $A \subseteq B$ ?

#### Řešení

Máme  $P(A) = P(A \cap B) = P(A)P(B)$ , a tedy  $P(A) = P(A)P(B)$ , což může nastat jen tehdy, když  $P(A) = 0$ , nebo  $P(B) = 1$ .

#### Úloha 5 (Na lovu jevů)

Najděte jevy  $A, B, C$  (na libovolném pravděpodobnostním prostoru), které

1. jsou nezávislé.
2. nejsou po dvou nezávislé, ale  $P(A \cap B \cap C) = P(A)P(B)P(C)$ .
3. jsou po dvou nezávislé, ale  $P(A \cap B \cap C) \neq P(A)P(B)P(C)$ .

**Řešení** 1. Házeme tři kostky,  $A, B, C$  jsou na první/druhé/třetí kostce padla jednička

2. Hážeme jednou šestistěnnou kostkou:  $A$ : „padla sedmička“,  $B$ : „padla dvojka“,  $C$ : „padla trojka“  $P(B \cap C) = 0 \neq P(B)P(C) = 1/36$ , a  $P(A)P(B)P(C) = 0 = P(A \cap B \cap C)$ .

Méně líně: hážeme dvěma šestistěnkami, a máme  $A$ : „na první kostce padla 1, 2 nebo 3“,  $B$ : „na první kostce padlo 3, 4, 5“ a  $C$ : „součet na obou kostkách je 9“. Pak  $P(A) = P(B) = 1/2$ ,  $P(C) = 1/9$ ,  $P(A \cap B) = 1/6 \neq 1/4$ ,  $P(A \cap C) = 1/36 \neq 1/18$ ,  $P(B \cap C) = 1/12 \neq 1/18$ , ale  $P(A \cap B \cap C) = 1/36 = P(A)P(B)P(C)$ .

3. Opět dvě kostky, jevy:  $A$ : „součet kostek je 7“,  $B$ : „na první kostce padla 3“,  $C$ : „na druhé kostce padla 4“. Pak  $P(A) = P(B) = P(C) = 1/6$ , a  $P(A \cap B) = P(A \cap C) = P(B \cap C) = 1/36$ , ale  $P(A \cap B \cap C) = 1/36 \neq 1/216 = P(A)P(B)P(C)$ .

## Přípravka na náhodné veličiny

### Úloha 6 (Pokusy)

Prokop hází basketbalovým míčem na koš, v každém pokusu má pravděpodobnost zásahu  $1/10$ , pokusy jsou nezávislé. Skončí po první trefě. Označme  $X$  celkový počet hodů.

- Jaká je  $P[X > k]$ ?
- Jaká je distribuce  $X$ ? Tj. určete pravděpodobnostní funkci  $p_X$ , tj. pro každé  $x$  určete  $P[X = x]$ .
- Jaká je  $P[X \geq 10 | X \geq 5]$ ?

**Řešení** 1.  $P[X > k] = (1 - 1/10)^k$

2.  $P[X = k] = (1 - \frac{1}{10})^{k-1} \frac{1}{10}$

3.  $P[X \geq 10 | X \geq 5] = \frac{(1-1/10)^9}{(1-1/10)^4} = (1 - 1/10)^5$

### Úloha 7 (Pokusy, ale jinak)

Navazujeme na předchozí úlohu: označme  $Y = X \bmod 2$ , tj.  $Y = 0$ , pokud je  $X$  sudé, jinak  $Y = 1$ . Určete distribuci  $Y$  (tj. určete  $P[Y = y]$  pro všechna  $y \in \mathbb{R}$ ).

### Řešení

Jediné dvě zajímavé hodnoty jsou 0 a 1, a pro ty:  $P[Y = 1] = \sum_{i \geq 0} \frac{1}{10} \cdot \left(\frac{9}{10}\right)^{2k} = \frac{1}{10} \sum_{i \geq 0} \left(\frac{81}{100}\right)^k = \frac{1}{10} \cdot \frac{1}{1 - \frac{81}{100}} = \frac{1}{10} \cdot \frac{100}{19} = \frac{10}{19}$ , a tím pádem  $P[Y = 0] = \frac{9}{10}$ .

### Úloha 8 (Pokusy, ale ještě jinak)

Quido také hází míčem na koš, má pravděpodobnost  $p$ , že se trefí a také jsou hody nezávislé. Označme  $Z$  počet zásahů z  $n$  pokusů. Určete distribuci  $Z$  (tj.  $P[Z = k]$  pro všechna  $k$ ).

### Řešení

Pro  $k < 0$ ,  $k > n$ ,  $k \notin \mathbb{N}$ :  $P[Z = k] = 0$ . Jinak  $P[Z = k] = \binom{n}{k} p^k (1 - p)^{n-k}$ .

Název	Značení	Pravděpodobnostní funkce	Rozsah (Im( $X$ ))	Střední hodnota
Bernoulliho	$X \sim \text{Ber}(p)$	$p_X(1) = p, p_X(0) = 1 - p$	$\{0, 1\}$	$p$
Binomické	$X \sim \text{Bin}(n, p)$	$p_X(k) = \binom{n}{k} p^k (1 - p)^{n-k}$	$\{0, 1, \dots, n\}$	$np$
Geometrické	$X \sim \text{Geo}(p)$	$p_X(k) = (1 - p)^{k-1} p$	$\{1, 2, \dots\}$	$1/p$
Poissonovo	$X \sim \text{Poi}(\lambda)$	$p_X(k) = e^{-\lambda} \frac{\lambda^k}{k!}$	$\{0, 1, \dots\}$	$\lambda$
Uniformní	$X \sim \text{Uni}(a, b)$	$p_X(k) = \frac{1}{b-a+1}$	$\{a, a+1, \dots, b\}$	$\frac{a+b}{2}$
Hypergeometrické	$X \sim \text{Hyp}(N, K, n)$	$p_X(k) =$	$\{0, 1, 2, \dots, \min(n, K)\}$	$n \frac{K}{N}$

### Úloha 9 (Klíče)

Na kroužku máme pět klíčů, jeden z nich je správný, ale my nevíme jaký. Zkoušíme otevřít dveře.

- Po každém pokusu se nám kroužek vysmekne, a vybíráme vždy znovu náhodně.
- Vybíráme v náhodném pořadí, ale každý klíč jenom jednou (můžeme si je poznačit).

V obou případech zkoumáme, kolikátým pokusem dveře otevřeme. Jaké je rozdělení této náhodné veličiny? Tj., určete, jaká je pravděpodobnost, že dveře otevřeme  $k$ -tým pokusem.  
Jak by oba předchozí případy dopadly, kdybychom měli 10 klíčů, ale dva byly správné?

**Řešení** 1.  $Geom(1/5)$

- uniformní na  $[5]$
- $Geom(1/5)$
- $p_Y(1) = 2/10, p_Y(2) = \frac{8}{10} \cdot \frac{2}{9}, p_Y(k) = \frac{2}{10-k} \cdot \prod_{j=1}^{k-1} \frac{9-j}{11-j}$

**Úloha 10** (Narozeniny)

Na přednášku je přihlášeno 234 lidí. Jaká je pravděpodobnost, že přesně jeden z nich má dnes narozeniny? Ignorujte přestupné roky, uvažujte, že všechny dny jsou stejně pravděpodobné pro narození.

- Použijte binomické rozdělení.
- Použijte aproximaci pomocí Poissonova rozdělení:  $Bin(n, \lambda/n)$  je přibližně  $Poi(\lambda)$ .
- Co se změní, když budeme uvažovat narozeniny zítra?

**Řešení** 1. Rozdělení je  $X \sim Bin(234, 1/365) \rightsquigarrow p_X(1) = 234 \cdot \frac{364^{233}}{365^{234}} \approx 0.338304648$

2. Vezmeme tedy  $Poi(234/365)$ , z čehož vyjde  $e^{-234/365} \cdot (234/365) \approx 0.337674748$ .

3. Rozdělení a pravděpodobnosti jsou stejné, ale hodnoty  $X$  a  $Y$  se mohou lišit.

**Úloha 11** (Která je tohle distribuce?)

Uvažme  $m + n$  hodů spravedlivou šestistěnnou kostkou. Označme  $X$  počet šestek z prvních  $m$  hodů,  $Y$  počet šestek z posledních  $n$  hodů. Jaká je distribuce  $X$ ,  $Y$  a  $X + Y$ ?

**Řešení**

Po řadě  $Bin(m, 1/6), Bin(n, 1/6), Bin(m + n, 1/6)$ .

**Úloha 12** (Odpověď zní: Jigglypuff, pohled shora)

V pytlíku je  $N$  bonbónů, z nichž  $K$  je dobrých. Náhodně vytáhneme  $n$  z nich, označíme  $X$  počet dobrých vytažených bonbónů.

- Jak se jmenuje rozdělení n.v.  $X$ ?
- Jaká je  $P(X = k)$ ?
- S pomocí tabulky určete  $\mathbb{E}[X]$ , pro  $n = 1$  si rozmyslete, že to je jasné.

**Řešení** 1. Hypergeometrické, mělo zaznít na přednášce

$$2. P(X = k) = \frac{\binom{K}{k} \binom{N-K}{n-k}}{\binom{N}{n}}$$

3. Dle tabulky  $n \frac{K}{N}$ .

### Bonusové úlohy

**Úloha 13** (Sir, this is a Wendy's)

Pokud vidíme bílého pudla, zvyšuje to naši důvěru, že je každá vrána černá?

**Řešení**

Více zde: [https://en.wikipedia.org/wiki/Raven\\_paradox](https://en.wikipedia.org/wiki/Raven_paradox)

### Úloha 14 (Hra v Sankt Petěrburgu)

Házíme opakovaně mincí. Pokud poprvé padla panna v  $n$ -tém hoďu, dostaneme odměnu  $2^n$ . Kolik byste byli ochotní zaplatit za účast v této hře?

#### Řešení

Střední hodnota je nekonečno.

### Úloha 15 (Prosecutor's fallacy)

Paní C. umřely dvě děti krátce po narození. Je obžalovaná za dvojnásobnou vraždu. Žalobce argumentuje takto: Pravděpodobnost syndromu náhlého úmrtí kojenců je  $1/8500$ . Takže pravděpodobnost dvou takových jevů je  $1/8500^2$ . Tudíž pravděpodobnost, že paní C. je nevinná, je  $1/8500^2$ , což je hodně málo. Formulujte argumenty žalobce v řeči pravděpodobnosti a nalezněte v nich dvě chyby.

#### Řešení

Pravděpodobnosti nemusí být nezávislé, a tohle je pst dvou náhlých úmrtí za předpokladu, že je nevinná. Naopak, jaká je pst toho, že je nevinná, za předpokladu dvou úmrtí je těžké říct.

## Užitečné pojmy

**Definice.** Bud'  $I$  libovolná množina indexů. Jevy  $\{A_i : i \in I\}$  nazveme *nezávislé (independent)*, pokud pro každou konečnou množinu  $J \subseteq I$

$$P\left(\bigcap_{j \in J} A_j\right) = \prod_{j \in J} P(A_j).$$

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