

Pravděpodobnost a statistika - cvičení 2

Pravděpodobnost

Simona Domesová

<http://home1.vsb.cz/~dom0015>

Náhodný pokus = konečný děj, jehož výsledek není předem jednoznačně určen podmínkami, za kterých probíhá

- při opakování za stejných podmínek dává různé výsledky
- není předem znám výsledek, pouze množina možných výsledků

(např. hod kostkou)

Základní prostor Ω = množina všech možných výsledků

(zde $\Omega = \{1, 2, 3, 4, 5, 6\}$)

Elementární jev = prvek základního prostoru Ω (jednoprvková podmnožina Ω)

(zde $\{1\}, \{2\}, \{3\}, \{4\}, \{5\}$ a $\{6\}$)

Jev A = libovolná podmnožina základního prostoru Ω

(např. $A = \{2, 4, 6\}$ - slovy „Padne sudé číslo.“,
 $A = \{6\}$ - slovy „Padne šestka.“,

$A = \emptyset$ - slovy např. „Padne osmička“ = tzv. jev nemožný,
 $A = \{1, 2, 3, 4, 5, 6\}$ - slovy např. „Padne kladné číslo.“ = tzv. jev jistý)

Úplná množina vzájemně disjunktích jevů = množina disjunktích jevů, jejichž sjednocením je základní prostor

$$\text{(např. } \{2, 4, 6\} \cup \{1, 2, 3\} = \Omega,$$

$$\{1\} \cup \{2, 5\} \cup \{3, 4, 6\} = \Omega,$$

$$\{1\} \cup \{2\} \cup \{3\} \cup \{4\} \cup \{5\} \cup \{6\} = \Omega)$$

Jevové pole \mathcal{A} = neprázdný systém podmnožin základního prostoru uzavřený vůči doplňku a vůči sjednocení

(např. systém všech podmnožin základního prostoru Ω ,

systém tvořený pouze množinami \emptyset a Ω ,

systém tvořený množinami \emptyset , $\{1, 5\}$, $\{2, 3, 4, 6\}$, Ω)

Podmíněná pravděpodobnost

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{\text{pravděpodobnost, že nastanou oba jevy současně}}{\text{pravěpodobnost podmínky}}$$

Jaká je pravděpodobnost, že padlo sudé číslo, jestliže vím, že padlo číslo menší než 6?

$A = \{2, 4, 6\}$ - slovy „Padlo sudé číslo.“

$B = \{1, 2, 3, 4, 5\}$ - slovy „Padlo číslo menší než 6.“ $\rightarrow P(B) = \frac{5}{6}$

$A \cap B = \{2, 4\}$ - nastaly oba jevy současně $\rightarrow P(A \cap B) = \frac{2}{6}$

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{\frac{2}{6}}{\frac{5}{6}} = \frac{2}{5}$$

Nezávislé jevy

Jestliže $P(A \cap B) = P(A) \cdot P(B)$ nebo $P(B) = 0$, říkáme, že jevy A a B jsou nezávislé.

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

Věta o úplné pravděpodobnosti

$$P(A) = \sum_{i=1}^n P(A|B_i) \cdot P(B_i),$$

kde $\{B_1, B_2, \dots, B_n\}$ je úplná množina vzájemně disjunktních jevů.

Např. jev $B = \{1, 2, 3, 4, 5\}$ spolu s jevem $B_2 = \{6\}$ tvoří úplnou množinu vzájemně disjunktních jevů.

$$\Rightarrow P(A) = P(A|B) \cdot P(B) + P(A|B_2) \cdot P(B_2)$$

Bayesova věta

$$P(B_k|A) = \frac{P(A|B_k) \cdot P(B_k)}{\sum_{i=1}^n P(A|B_i) \cdot P(B_i)},$$

kde $\{B_1, B_2, \dots, B_n\}$ je úplná množina vzájemně disjunktních jevů.

Definice pravděpodobnosti

Klasická definice pravděpodobnosti (Laplaceova) předpokládá, že všechny elementární jevy jsou stejně pravděpodobné

$$P(A) = \frac{\text{počet elementárních jevů, ze kterých je složen jev } A}{\text{počet všech elementárních jevů, které mohou nastat}}$$

- jaká je pravděpodobnost, že z klasického balíčku karet vytáhnu srdcovou?

Geometrická definice pravděpodobnosti zobecňuje klasickou definici pro případ, kdy je počet všech možných výsledků náhodného pokusu nespočetný

$$P(A) = \frac{\text{"velikost" jevu } A}{\text{"velikost" celého základního prostoru } \Omega}$$

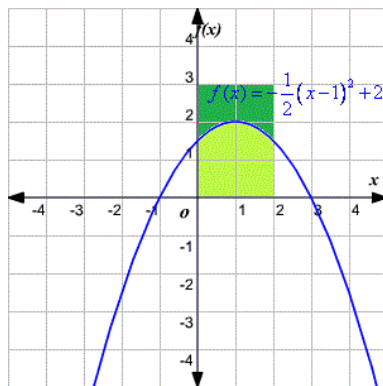
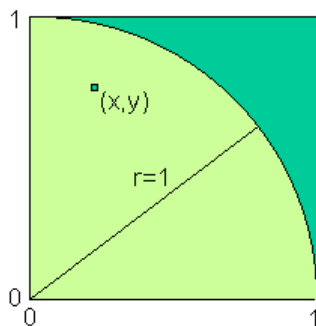
- např. jaká je pravděpodobnost, že na semaforu svítí zelená?

Statistická definice pravděpodobnosti vychází z opakování téhož náhodného pokusu

$$P(A) = \lim_{n \rightarrow \infty} \frac{\text{počet pokusů, v nichž nastal jev } A}{\text{počet všech provedených pokusů}}$$

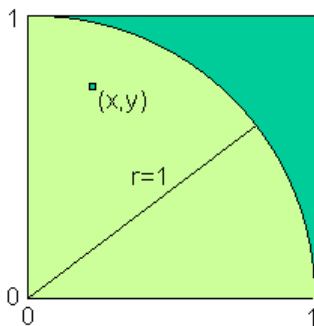
Příklad 1

Jaká je pravděpodobnost, že náhodně vybraný bod množiny Ω bude patřit do vyznačené (světle zelené) oblasti?

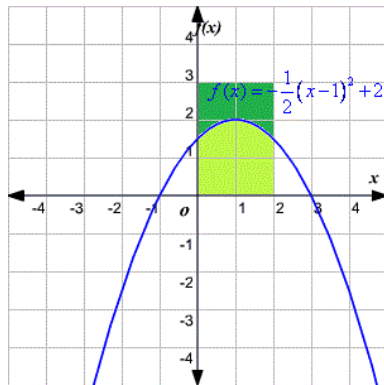


Příklad 1

Jaká je pravděpodobnost, že náhodně vybraný bod množiny Ω bude patřit do vyznačené (světle zelené) oblasti?



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

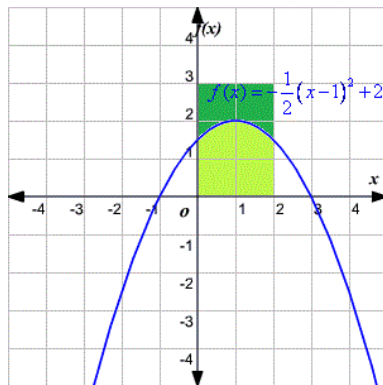
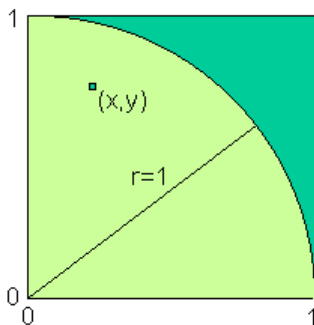


$$P(A) = \frac{\int_0^2 -\frac{1}{2}(x-1)^2 + 2 \, dx}{2 \cdot 3} = \frac{\frac{11}{6}}{6} = \frac{11}{18}$$

Jak bychom naopak odhadli obsah světle zelené oblasti?

Příklad 1

Jaká je pravděpodobnost, že náhodně vybraný bod množiny Ω bude patřit do vyznačené (světle zelené) oblasti?



Jak bychom naopak odhadli obsah S světle zelené oblasti?

$$S = P(A) \cdot (\text{obsah oblasti } \Omega)$$

Příklad 2 (Narozeninový paradox)

Uvažujeme skupinu 25 náhodně vybraných lidí. Jaká je pravděpodobnost, že někteří dva lidé mají narozeniny ve stejný den?

Příklad 2 (Narozeninový paradox)

Uvažujeme skupinu 25 náhodně vybraných lidí. Jaká je pravděpodobnost, že někteří dva lidé mají narozeniny ve stejný den?

Určíme nejdříve pravděpodobnost, že se má každý narozeniny v jiný den:

$$\bar{p} = 1 \cdot \frac{364}{365} \cdot \frac{363}{365} \cdots \frac{341}{365} \doteq 0.431$$

Pravděpodobnost, že se alespoň dva narodili ve stejný den:

$$p = 1 - \bar{p} \doteq 0.569$$

Jak by tomu bylo ve skupině padesátí lidí?

Příklad 2 (Narozeninový paradox)

Uvažujeme skupinu 25 náhodně vybraných lidí. Jaká je pravděpodobnost, že někteří dva lidé mají narozeniny ve stejný den?

Určíme nejdříve pravděpodobnost, že se má každý narozeniny v jiný den:

$$\bar{p}(25) = 1 \cdot \frac{364}{365} \cdot \frac{363}{365} \cdots \frac{341}{365} \doteq 0.431$$

Pravděpodobnost, že se alespoň dva narodili ve stejný den:

$$p(25) = 1 - \bar{p}(25) \doteq 0.569$$

Jak by tomu bylo ve skupině padesátí lidí?

$$\bar{p}(50) \doteq 0.0296$$

$$p(50) = 1 - \bar{p}(50) \doteq 0.9704$$

Příklad 3 (Bayesova věta)

Bylo zjištěno, že 60% všech e-mailů tvoří spam. Pravděpodobnost, že filtr spam rozpozná, je 80%. Pravděpodobnost, že filtr chybně označí za spam e-mail, který spamem není, je 5%.

Příchozí e-mail byl označen za spam, jaká je pravděpodobnost, že se skutečně jedná o spam?

Příklad 3 (Bayesova věta)

Bylo zjištěno, že 60% všech e-mailů tvoří spam. Pravděpodobnost, že filtr spam rozpozná, je 80%. Pravděpodobnost, že filtr chybně označí za spam e-mail, který spamem není, je 5%.

Příchozí e-mail byl označen za spam, jaká je pravděpodobnost, že se skutečně jedná o spam?

S ... jedná se o spam

H ... nejedná se o spam

FS ... filtrem označeno za spam

FH ... neoznačeno za spam

$$\begin{aligned} P(S|FS) &= \frac{P(FS|S) \cdot P(S)}{P(FS|S) \cdot P(S) + P(FS|H) \cdot P(H)} = \\ &= \frac{0.8 \cdot 0.6}{0.8 \cdot 0.6 + 0.05 \cdot 0.4} = 96\% \end{aligned}$$