

# Základy matematiky a statistiky pro humanitní obory II

Pavel Rychlý Vojtěch Kovář

Fakulta informatiky, Masarykova univerzita  
Botanická 68a, 60200 Brno, Czech Republic  
{pary, xkovar3}@fi.muni.cz

23. 3. 2011

## Obsah přednášky

Náhodná veličina

Rozložení pravděpodobnosti

Distribuční funkce

Náhodný vektor

## Náhodná veličina

- ▶ Též náhodná proměnná, random variable
- ▶ Vlastnost  $X$ , jejíž hodnotu neznáme
  - ▶ protože nemáme dost informací
  - ▶ protože vlastnost ještě nenabyla hodnoty
  - ▶ „teplota vzduchu v Brně zítra v poledne“
  - ▶ „výsledek hodu kostkou“
  - ▶ (kdybychom dokonale znali rychlost kostky, její rotaci, vlastnosti povrchu atd., byli bychom schopni **vypočítat**, co padne)
- ▶ Většinou ale máme nějaké informace o dané vlastnosti
  - ▶ výsledky hodů kostkou v minulosti
  - ▶ měření teploty v minulosti
  - ▶ → pravděpodobnosti jednotlivých hodnot nejsou stejné

## Rozložení pravděpodobnosti

- ▶ Též distribuce pravděpodobnosti, probability distribution
- ▶ Pravděpodobnost, že vlastnost  $X$  nabude hodnotu  $x$ 
  - ▶ určíme např. z dřívějších pozorování
  - ▶ pravděpodobnost, že zítra v poledne bude -100 stupňů
  - ▶ vs. pravděpodobnost, že zítra v poledne bude 10 stupňů
- ▶ Formálněji
  - ▶ funkce  $f$ , která každé možné hodnotě vlastnosti přiřadí číslo (pravděpodobnost) od 0 do 1
  - ▶ součet hodnot funkce pro všechny možné hodnoty vlastnosti je 1
  - ▶  $f_X(x) = P(X = x)$

## Rozložení pravděpodobnosti

- ▶ Rozlišujeme diskrétní a spojitá rozložení
  - ▶ konečně nebo spočetně mnoho hodnot vs. nespočetně mnoho hodnot
  - ▶ omezíme se na diskrétní
- ▶ pravděpodobnostní prostor
  - ▶ množina všech hodnot vlastnosti spolu s distribucí pravděpodobnosti

## Statistika a pravděpodobnost

- ▶ Pravděpodobnostní rozložení je určeno
  - ▶ „ideální“ funkcí
  - ▶ pozorování v minulosti zachyceným ve **statistickém souboru**
  - ▶ → z naměřených dat určujeme pravděpodobnost neznámých dat
  - ▶ např. pravděpodobnostní rozložení hodnot teploty zítra v poledne
  - ▶ → na základě měření poledních teplot v minulých dnech
  - ▶ např. pravděpodobnostní rozložení slov (slovních druhů, ...) v jazyce
  - ▶ → na základě dostatečně velkého vzorku textů

## Statistika a pravděpodobnost

- ▶ Statistický soubor a pravděpodobnostní prostor
  - ▶ relativní četnosti ve statistickém souboru odpovídají hodnotám pravděpodobnostního rozložení v pravděpodobnostním prostoru

## Distribuční funkce

- ▶ Cumulative distribution function
- ▶ Pravděpodobnost, že vlastnost  $X$  nabude hodnotu  $x$  nebo menší
  - ▶  $F_X(x) = P(X \leq x)$
  - ▶ hodnota distribuční funkce odpovídá percentilu
  - ▶ pro medián je hodnota distribuční funkce 0.5

# Náhodný vektor

- ▶ Posloupnost náhodných veličin

- ▶ např. počasí v Brně zítra v poledne
- ▶ (*teplota, tlak, vlhkost*)

- ▶ Rozložení pravděpodobnosti

- ▶ pravděpodobnost vektoru může být jiná než součin pravděpodobností jednotlivých jeho složek

- ▶ Distribuční funkce

- ▶ např.  $F(x, y) = P(X \leq x \wedge Y \leq y)$