Základy stochastiky

Statistické charakteristiky a odhady

Prof. RNDr. Gejza Dohnal, CSc.

Statistické charakteristiky

Statistické charakteristiky: jsou spočteny na základě pozorování $x_1, x_2, \ldots, x_n$ výběru $X_1, X_2, \ldots, X_n$.

Pravděpodobnostní charakteristiky	Výběrové charakteristiky
Střední hodnota $E(X) = \int_{-\infty}^{\infty} x f(x) dx$	Výběrový průměr $\bar{X} = \frac{1}{n} \sum_{i=1}^{n} x_i$
Momenty $\mu_k(X) = E(X - E(X))^k$	Výběrové momenty $m_k(X) = \sum_{i=1}^{n} (x_i - \bar{X})^k$
Rozptyl $var(X) = E(X - E(X))^2$	Výběrový rozptyl $s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{X})^2$
Kvantily $\tilde{x}_{100\alpha}$	Výběrové kvantily $X_{([np]+1)}$

Statistické charakteristiky

Statistické charakteristiky: jsou spočteny na základě pozorování $x_1, x_2, \ldots, x_n$ výběru $X_1, X_2, \ldots, X_n$.

Pravděpodobnostní charakteristiky	Výběrové charakteristiky
Střední hodnota $E(X) = \int_{-\infty}^{\infty} x f(x) dx$	Výběrový průměr $\bar{X} = \frac{1}{n} \sum_{i=1}^{n} x_i$
Momenty $\mu_k(X) = E(X - E(X))^k$	Výběrové momenty $m_k(X) = \sum_{i=1}^{n} (x_i - \bar{X})^k$
Rozptyl $var(X) = E(X - E(X))^2$	Výběrový rozptyl $s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{X})^2$
Kvantily $\tilde{x}_{100\alpha}$	Výběrové kvantily $X_{([np]+1)}$

Pořadové statistiky

Statistické charakteristiky: jsou spočteny na základě pozorování $x_1, x_2, \ldots, x_n$ výběru $X_1, X_2, \ldots, X_n$.

- *Uspořádaný výběr*: $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$ vznikne z původního výběru $X_1, X_2, \ldots, X_n$ upořádáním podle velikosti pozorovaných hodnot $x_1, x_2, \ldots, x_n$.
- *Pořadová statistika*: $X_{(k)}$ je náhodná veličina X_m, která je k-tá v po-řadí podle velikosti pozorovaných hodnot $x_1, x_2, \ldots, x_n$.
  Index k nazýváme *pořadím veličiny* X_m a zapisujeme to $R_m = k$.
- Statistika $X_{(1)}$ se nazývá minimum, $X_{(n)}$ je maximum
- medián $\tilde{x}_{50}$: je-li n liché, je roven $X_{([n/2]+1)}$
  pro n sudé je roven $(X_{(n/2)} + X_{(n/2+1)})/2$
- dolní kvartil $\tilde{x}_{25}$: $X_{([n/4]+1)}$ resp. $(X_{(n/4)} + X_{(n/4+1)})/2$
- horní kvartil $\tilde{x}_{75}$: $X_{([3n/4]+1)}$ resp. $(X_{(3n/4-1)} + X_{(3n/4)})/2$

Pořadové statistiky

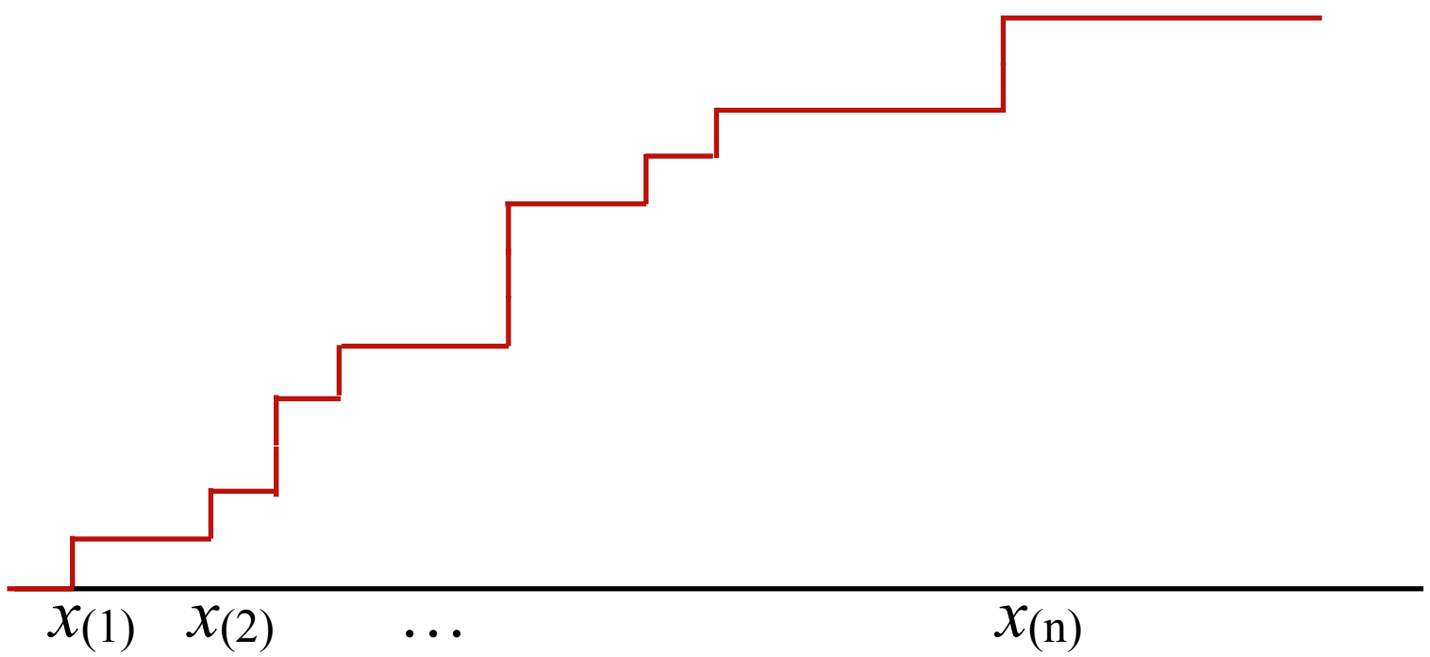
Statistické charakteristiky: jsou spočteny na základě pozorování $x_1, x_2, \ldots, x_n$ výběru $X_1, X_2, \ldots, X_n$.

- *Uspořádaný výběr*: $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$ vznikne z původního výběru $X_1, X_2, \ldots, X_n$ upořádáním podle velikosti pozorovaných hodnot $x_1, x_2, \ldots, x_n$.
- *Pořadová statistika*: $X_{(k)}$ je náhodná veličina X_m, která je k-tá v po-řadí podle velikosti pozorovaných hodnot $x_1, x_2, \ldots, x_n$.
  Index k nazýváme *pořadím veličiny* X_m a zapisujeme to $R_m = k$.
- Statistika $X_{(1)}$ se nazývá minimum, $X_{(n)}$ je maximum
- medián $\tilde{x}_{50}$: je-li n liché, je roven $X_{([n/2]+1)}$
  pro n sudé je roven $(X_{(n/2)} + X_{(n/2+1)})/2$
- dolní kvartil $\tilde{x}_{25}$: $X_{([n/4]+1)}$ resp. $(X_{(n/4)} + X_{(n/4+1)})/2$
- horní kvartil $\tilde{x}_{75}$: $X_{([3n/4]+1)}$ resp. $(X_{(3n/4-1)} + X_{(3n/4)})/2$

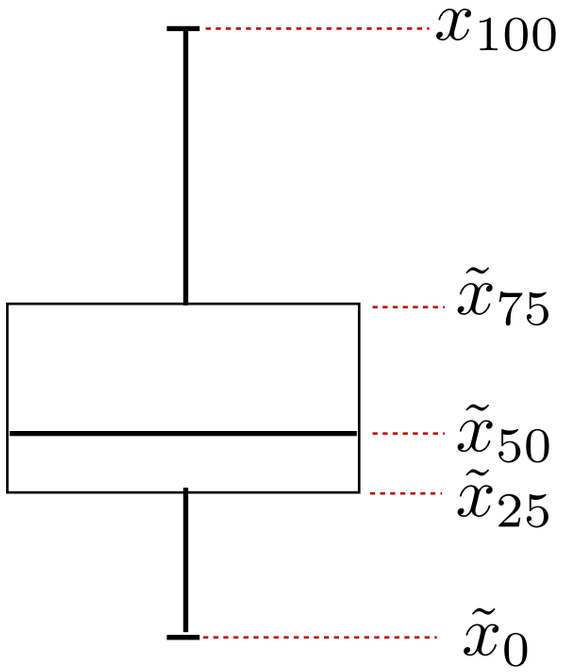
Pořadové statistiky

Statistické charakteristiky: jsou spočteny na základě pozorování $x_1, x_2, \ldots, x_n$ výběru $X_1, X_2, \ldots, X_n$.

empirická distribuční funkce



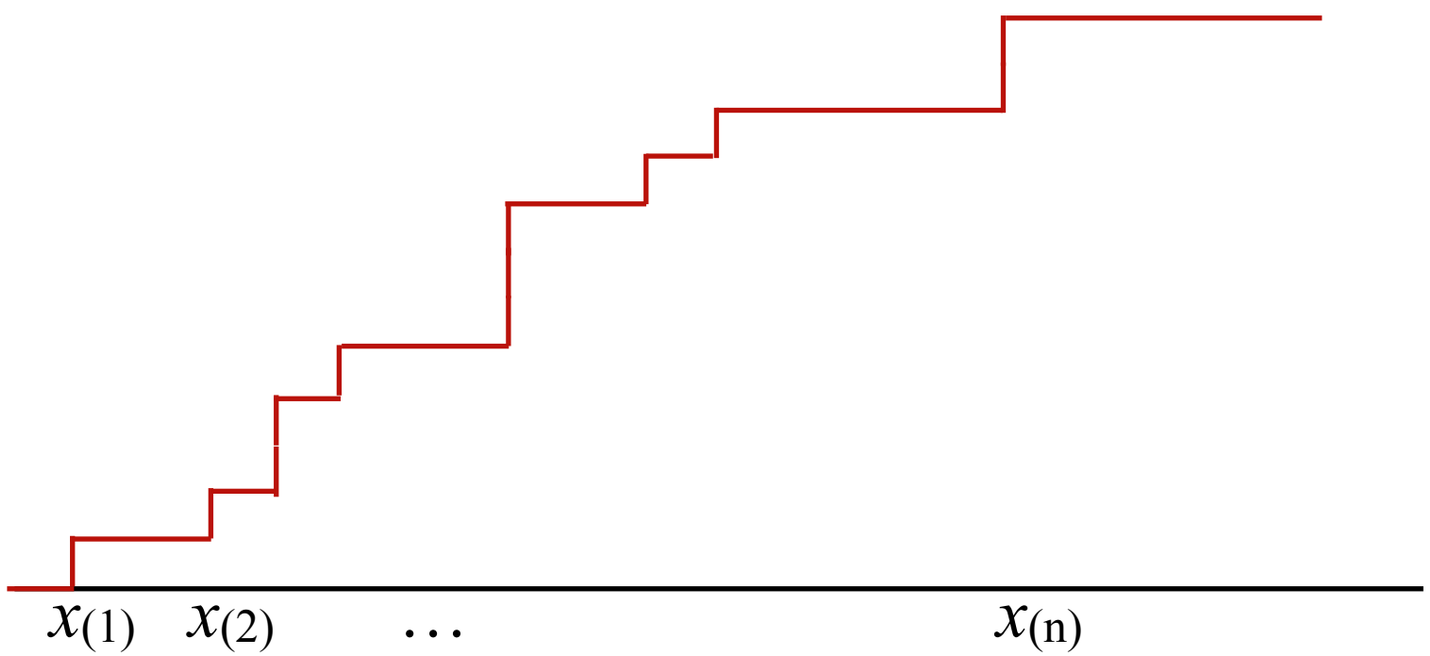
krabicový graf



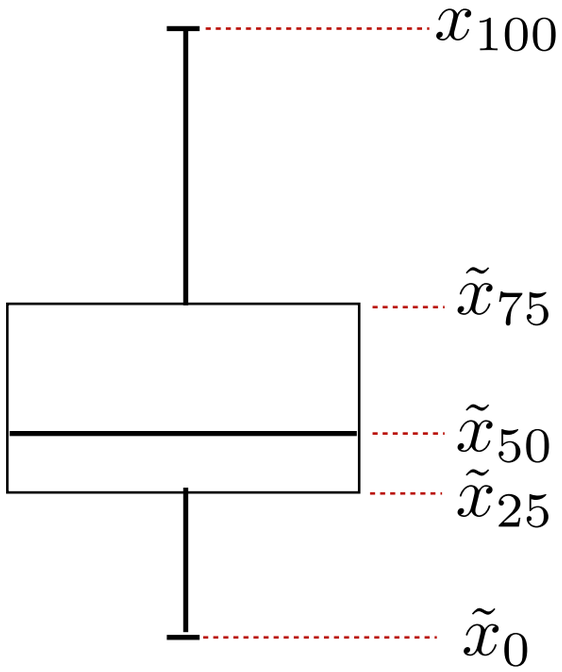
Pořadové statistiky

Statistické charakteristiky: jsou spočteny na základě pozorování $x_1, x_2, \ldots, x_n$ výběru $X_1, X_2, \ldots, X_n$.

empirická distribuční funkce



krabicový graf



Statistické charakteristiky

Za předpokladu, že náhodný výběr je nezávislý a je z normálního rozdělení, t.j. $X_1, X_2, \ldots, X_n$ jsou i.i.d. a $X_k \sim N(\mu,\sigma^2)$, $k = 1, 2, \ldots, n$, lze určit rozdělení pravděpodobnosti některých charakteristik:

- Pokud je μ a σ^2 známé, má výběrový průměr rozdělení $N(\mu,\sigma^2/n)$
- Pokud μ a σ^2 neznáme, má veličina $(X - \bar{X})/s$ tzv. Studentovo t-rozdělení $t(n-1)$
- Veličina $(n-1).s^2/\sigma^2$ má $\chi^2(n-1)$ rozdělení (o $n-1$ stupních volnosti)

Statistické charakteristiky

Za předpokladu, že náhodný výběr je nezávislý a je z normálního rozdělení, t.j. $X_1, X_2, \ldots, X_n$ jsou i.i.d. a $X_k \sim N(\mu,\sigma^2)$, $k = 1, 2, \ldots, n$, lze určit rozdělení pravděpodobnosti některých charakteristik:

- Pokud je μ a σ^2 známé, má výběrový průměr rozdělení $N(\mu,\sigma^2/n)$
- Pokud μ a σ^2 neznáme, má veličina $(X - \bar{X})/s$ tzv. Studentovo t-rozdělení $t(n-1)$
- Veličina $(n-1).s^2/\sigma^2$ má $\chi^2(n-1)$ rozdělení (o $n-1$ stupních volnosti)

Statistické charakteristiky

Další důležité výběrové charakteristiky:

- Výběrová šikmost (skewness): $$Skew(X) = \frac{m_3(X)}{m_2^{3/2}(X)}$$

  pro $X \sim \mathrm{N}(\mu,\sigma^2)$ je

  $$E(Skew(X)) = 0 \qquad var(Skew(X)) = \frac{6(n-2)}{(n+1)(n+3)}$$

- Výběrová špičatost (kurtosis): $$Kurt(X) = \frac{m_4(X)}{m_2^2(X)} - 3$$

  pro $X \sim \mathrm{N}(\mu,\sigma^2)$ je

  $$E(Kurt(X)) = -\frac{6}{n+1} \qquad var(Kurt(X)) = \frac{24n(n-2)(n-3)}{(n+1)^2(n+3)(n+5)}$$

Statistické charakteristiky

Další důležité výběrové charakteristiky:

- Výběrová šikmost (skewness): $$Skew(X) = \frac{m_3(X)}{m_2^{3/2}(X)}$$

  pro $X \sim \mathrm{N}(\mu,\sigma^2)$ je

  $$E(Skew(X)) = 0 \qquad var(Skew(X)) = \frac{6(n-2)}{(n+1)(n+3)}$$

- Výběrová špičatost (kurtosis): $$Kurt(X) = \frac{m_4(X)}{m_2^2(X)} - 3$$

  pro $X \sim \mathrm{N}(\mu,\sigma^2)$ je

  $$E(Kurt(X)) = -\frac{6}{n+1} \qquad var(Kurt(X)) = \frac{24n(n-2)(n-3)}{(n+1)^2(n+3)(n+5)}$$