

NMAI059 Pravděpodobnost a statistika 1

9. přednáška

Robert Šámal

Přehled

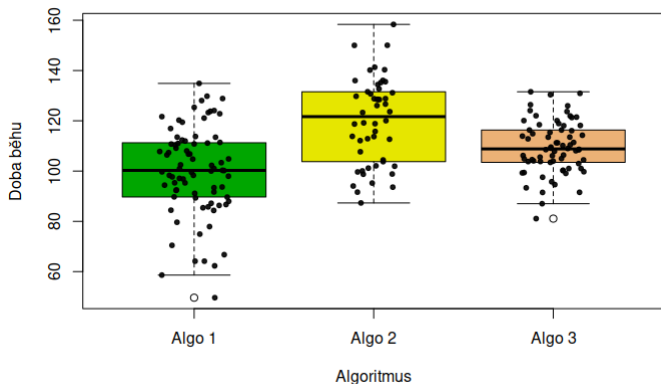
Statistika – úvod

Statistika – bodové odhady

Statistika – intervalové odhady

Intro – explorační analýza dat (exploratory data analysis)

- ▶ posbíráme data (a dáme pozor na systémové chyby – nezávislost, nezaujatost, . . .)
- ▶ různé tabulky (třeba v Excelu a spol.)
- ▶ vhodné obrázky: histogram, krabicový diagram (boxplot), atd.



Náhodný výběr

- ▶ s vracením
- ▶ bez vracení

$\Omega = \{\text{všechny } n\text{-tice obyvatel ČR}\}$

Pro $\omega = (\omega_1, \dots, \omega_n)$ zvolíme $X_i = I(\omega_i \text{ je levák})$.

Statistika – přehled

- ▶ nezávislá měření – hodnoty n.n.v. $X_1, \dots, X_n \sim F$
náhodný výběr s distribuční funkcí F s rozsahem n
- ▶ neparametrické modely: povolujeme velkou třídu F
- ▶ parametrické modely: $F \in \{F_\vartheta : \vartheta \in \Theta\}$
- ▶ příklady:
 - ▶ $Pois(\lambda)$ (parametr $\vartheta = \lambda, \Theta = \mathbb{R}^+$)
 - ▶ $U(a, b)$ (parametr $\vartheta = (a, b), \Theta = \mathbb{R}^2$)
 - ▶ $N(\mu, \sigma^2)$ (parametr $\vartheta = (\mu, \sigma), \Theta = \mathbb{R} \times \mathbb{R}^+$)
- ▶ Všechny modely jsou špatné, ale některé jsou užitečné.

Zkoumané úlohy – cíle konfirmační analýzy (confirmatory data analysis)

- ▶ bodové odhady
 - ▶ intervalové odhady
 - ▶ testování hypotéz
 - ▶ (lineární) regrese
-
- ▶ *statistika* – libovolná funkce náhodného výběru, tj. např. aritmetický průměr, medián, maximum, atd.
Tj. $T = T(X_1, \dots, X_n)$.

Přehled

Statistika – úvod

Statistika – bodové odhady

Statistika – intervalové odhady

Výběrový průměr a rozptyl

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$$

$$\bar{S}_n = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2$$

$$\hat{S}_n = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$$

Cíle

Definice

Odhad $\hat{\Theta}_n = \hat{\Theta}_n(X_1, \dots, X_n)$ parametru ϑ je

- ▶ *neustranný (unbiased)* – pokud $\vartheta = \mathbb{E}(\hat{\Theta}_n)$
- ▶ *asymptoticky neustranný (asymptotically unbiased)* – pokud $\vartheta = \lim_{n \rightarrow \infty} \mathbb{E}(\hat{\Theta}_n)$
- ▶ *vychýlení (bias)* $bias_{\vartheta}(\hat{\Theta}_n) = \mathbb{E}(\hat{\Theta}_n) - \vartheta$
- ▶ *střední kvadratická chyba (mean squared error, MSE)* je $\mathbb{E}((\hat{\Theta} - \vartheta)^2)$

Věta

$$MSE = bias_{\vartheta}(\hat{\Theta}_n)^2 + var_{\vartheta}(\hat{\Theta}_n)$$

Parametry výběrového momentu a rozptylu

Věta

1. \bar{X}_n je konzistentní nestranný odhad μ
2. \bar{S}_n je konzistentní asymptoticky nestranný odhad μ
3. \hat{S}_n je konzistentní nestranný odhad μ

Metoda momentů

- ▶ $m_r(\vartheta) := \mathbb{E}(X^r)$ pro $X \sim F_\vartheta$... r -tý moment
- ▶ $\widehat{m}_r(\vartheta) := \frac{1}{n} \sum_{i=1}^n X_i^r$ pro náhodný výběr X_1, \dots, X_n z F_ϑ
... r -tý výběrový moment

Věta

$\widehat{m}_r(\vartheta)$ je nestranný konzistentní odhad pro $m_r(\vartheta)$

- ▶ Odhad metodou momentů je řešení soustavy rovnic

$$m_r(\vartheta) = \widehat{m}_r(\vartheta) \quad r = 1, \dots, k.$$

Metoda momentů – příklady

Metoda maximální věrohodnosti (maximal likelihood, ML)

- ▶ náh. výběr $X = (X_1, \dots, X_n)$ z modelu s parametrem ϑ
- ▶ možný výsledek $x = (x_1, \dots, x_n)$
- ▶ ... sdružená pravděpodobnostní funkce $p_X(x; \vartheta)$
- ▶ ... sdružená hustota $f_X(x; \vartheta)$
- ▶ *věrohodnost (likelihood)* $L(x; \vartheta)$ značí p_X nebo f_X
- ▶ normálně: máme pevné ϑ , a $L(x; \vartheta)$ je funkce x
- ▶ teď: máme pevné x a $L(x; \vartheta)$ je funkce ϑ

Metoda MV (ML):

volíme takové ϑ , pro které je $L(x; \vartheta)$ maximální

Metoda maximální věrohodnosti (maximal likelihood, ML)

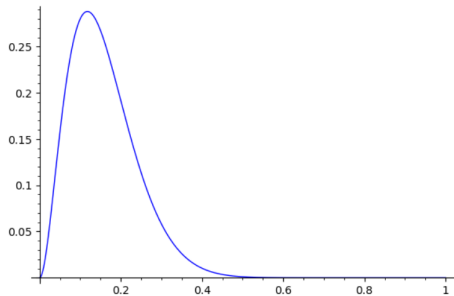
- ▶ **Metoda MV (ML):**
volíme takové ϑ , pro které je $L(x; \vartheta)$ maximální
- ▶ definujeme také $\ell(x; \vartheta) = \log(L(x; \vartheta))$
- ▶ díky nezávislosti je

$$L(x; \vartheta) =$$

$$\ell(x; \vartheta) =$$

ML – leváci

```
plot(binomial(17,2)*p^2*(1-p)^15, [0,1])
```



Přehled

Statistika – úvod

Statistika – bodové odhady

Statistika – intervalové odhady

Intervalové odhady

- ▶ místo jednoho čísla s nejistým významem vypočítáme z dat interval $[\hat{\Theta}^-, \hat{\Theta}^+]$

Definice

Nechť $\hat{\Theta}^-$, $\hat{\Theta}^+$ jsou n.v. které závisí na náhodném výběru $X = (X_1, \dots, X_n)$. Tyto n.v. určují intervalový odhad, též konfidenční interval o spolehlivosti $1 - \alpha$ ($1 - \alpha$ confidence interval), pokud

$$P(\hat{\Theta}^- \leq \vartheta \leq \hat{\Theta}^+) \geq 1 - \alpha.$$

Intervalové odhady normální náhodné veličiny