Prostý náhodný výběr, Bodový odhad

10. – 14. 12. 2012

Rozbor zápočtové písemky

Prostý náhodný výběr

1. Navrhněte algoritmus pro provedení prostého náhodného výběru o rozsahu n ze souboru jednotek $\{1, \ldots, N\}$ v následujících situacích:

   (a) nesekvenční výběr,
   (b) sekvenční výběr, N známé,
   (c) sekvenční výběr, N neznámé.

   Dokažte, že Váš algoritmus skutečně vede na prostý náhodný výběr.

Bodový odhad

2. Nechť $X_1, \ldots, X_n$ je náhodný výběr z Poissonova rozdělení s parametrem $\lambda > 0$.

   (a) Najděte $\widehat{\lambda}_n$ odhad λ momentovou metodou.
   (b) Zjistěte, zda je odhad $\widehat{\lambda}_n$ nestranný.
   (c) Zjistěte, zda je odhad $\widehat{\lambda}_n$ konzistentní.
   (d) Nalezněte $\tilde{\lambda}_n$ odhad λ metodou maximální věrohodnosti.
   (e) Uvažujte odhad $T_n = (X_1 + X_2)/2$. Je tento odhad nestranný a konzistentní?
   (f) Uvažujte odhad $U_n = \frac{1}{n+1} \sum_{i=1}^{n} X_i$. Je tento odhad nestranný a konzistentní?

Opakování z přednášky

Prostý náhodný výběr :

- Ze základního souboru $U = \{1, \dots, N\}$ jednotek chceme náhodně vybrat výběr $s \subset U$ o velikosti n tak, že pro libovolné $s \subset U$

$$P(s) = \begin{cases} \frac{1}{\binom{N}{n}}, & \text{pro } |s| = n, \\ 0 & \text{jinak.} \end{cases}$$

Teorie odhadu:

- Veličiny $X_1, \dots, X_n$ tvoří **náhodný výběr**, jestliže jsou nezávislé a mají všechny stejné rozdělení. Chceme-li specifikovat jejich rozdělení, mluvíme o náhodném výběru z rozdělení F (např. náhodný výběr z $\mathsf{N}(\mu, \sigma^2)$, náhodný výběr z $\mathsf{Alt}(p)$ atd.)
- Nechť $X_1, \dots, X_n$ je náhodný výběr z rozdělení, které závisí na **neznámém** parametru $\theta \in \Theta$. **Odhadem** θ je libovolná (měřitelná) funkce náhodného výběru $X_1, \dots, X_n$, jejíž funkční předpis nezávisí na θ.
  Odhad $T_n = T_n(X_1, \dots, X_n)$ je tedy **náhodná veličina**.

Vlastnosti odhadů.

- V praxi pracujeme pouze s odhady, které mají „pěkné“ vlastnosti.
- Řekneme, že odhad $T_n = T_n(X_1, \dots, X_n)$ je **nestranný** odhad θ, jestliže

$$\mathsf{E}T_n = \mathsf{E}_\theta T_n = \theta, \qquad \text{pro všechna } \theta \in \Theta.$$

- Řekneme, že odhad $T_n = T_n(X_1, \dots, X_n)$ je **konzistentní** odhad θ, jestliže $T_n \to \theta$ v pravděpodobnosti pro $n \to \infty$ pro všechna $\theta \in \Theta$, tedy $\mathsf{P}(|T_n - \theta| > \epsilon) \to 0$ pro všechna $\epsilon > 0$ a $\theta \in \Theta$.

Konstrukce bodových odhadů

- Metoda maximální věrohodnosti (MLE): Hledáme

$$\mathsf{argmax}_\theta \prod_{i=1}^{n} f(X_i, \theta) = \mathsf{argmax}_\theta \sum_{i=1}^{n} \log f(X_i, \theta),$$

  kde $f(x, \theta)$ je „hustota“ veličin X (hustota pro spojité rozdělení a $f(x, \theta) = \mathsf{P}(X = x)$ pro diskrétní rozdělení).
- Momentová metoda: Za odhad θ vezmeme $\widehat{\theta}$, pro které

$$\mathsf{E}_{\widehat{\theta}} X^k = \frac{1}{n} \sum_{i=1}^{n} X_i^k.$$

- Odhadovacích metod existuje ještě mnohem více (metoda nejmenších čtverců v regresi, ...).