Intervalové odhady

8.1.2020

1. V hospodě jsme zakoupili n piv a zaznamenali jsme objem natočeného piva s cílem zjistit, zda je skutečná střední hodnota natočeného piva rovna 0,5 l. Označme naměřené hodnoty jako $X_1, \dots, X_n$. Lze předpokládat, že se jedná o náhodný výběr z rozdělení se střední hodnotou μ a rozptylem $\sigma^2 \in (0, \infty)$.
   (a) Nechť $\alpha \in (0,1)$ je malé a nechť σ^2 je známé. Zdůvodněte, že pak pro $n \to \infty$ platí
   $$\mathsf{P}\left(-u_{1-\alpha/2} \le \sqrt{n}\frac{\overline{X}_n - \mu}{\sigma} \le u_{1-\alpha/2}\right) = \mathsf{P}\left(\sqrt{n}\frac{|\overline{X}_n - \mu|}{\sigma} \le u_{1-\alpha/2}\right) \to 1-\alpha,$$
   kde $u_{1-\alpha/2} = \Phi^{-1}(1-\alpha/2)$ je kvantil $\mathsf{N}(0,1)$ na hladině $1-\alpha/2$. Na základě tohoto vztahu zkonstruujte intervalový odhad pro μ s asymptotickou spolehlivostí $1-\alpha$.
   (b) Prozkoumejte, na čem a jak závisí délka intervalového odhadu z (a).
   (c) Pomocí Sluckého věty odvoďte intervalový odhad pro μ, je-li σ^2 neznámé.
   (d) Pro $n = 10$ jsme dostali $\overline{X}_{10} = 0,480$, $S_{10} = 0,019$. Zkonstruujte intervalový odhad se spolehlivostí 95 % pro střední hodnotu natočeného piva. Překrývá tento interval hodnotu $0,5$?

2. Počet vstřelených gólů v jednom fotbalovém zápase se řídí Poissonovým rozdělením s neznámým parametrem $\lambda > 0$. Pro data z roku 2008 máme k dispozici 306 zápasů, kde bylo vstřeleno v průměru $2,92$ gólů a výběrový rozptyl vyšel $(1,64)^2$. Zkonstruujte intervalový odhad se spolehlivostí 95% pro λ. Rozhodněte na základě tohoto intervalu, zda jsou naše data v souladu s tvrzením, že je střední počet vstřelených gólů v jednom zápase roven 3.

Momentová vytvořující funkce, meze pro pravděpodobnost

3. Nechť X má binomické rozdělení $\mathsf{Bi}(n,p)$, kde $p \in (0,1)$.
   (a) Spočtěte momentovou vytvořující funkci X.
   (b) Pomocí (a) určete střední hodnotu a rozptyl X.

   Uvažujte dále $p = 1/2$.

   (c) Zapište, jakou mez nám pro $\mathsf{P}(X \ge \frac{3}{4}n)$ dává Markovova nerovnost pro $k = 1$ a $k = 2$.
   (d) Pomocí Chernoffovy meze ukažte, že platí
   $$\mathsf{P}(X \ge \frac{3}{4}n) \le e^{-n/4[3\log(3/2)-1]}.$$
   Návod: Využijte, že pro $x \ge 0$ platí $\log(1+x) \le x$.
   (e) Vyjádřete $\mathsf{P}(X \ge \frac{3}{4}n)$ přibližně pomocí centrální limitní věty.

4. Nechť X má Poissonovo rozdělení s parametrem $\lambda > 0$.
   (a) Spočtěte momentovou vytvořující funkci.
   (b) Pomocí (a) určete střední hodnotu a rozptyl X.

Opakování z přednášky

Intervalový odhad Nechť $\alpha \in (0,1)$ je dané (malé) a nechť $X_1, \ldots, X_n$ je náhodný výběr z rozdělení, které závisí na neznámém parametru $\theta \in\subset \mathbb{R}$. Intervalový odhad θ se spolehlivostí $1-\alpha$ je interval s náhodnými mezemi $[D, H]$, kde $D = D(X_1, \ldots, X_n)$ a $H = H(X_1, \ldots, X_n)$ jsou funkce náhodného výběru, jejichž předpis nezávisí na θ (a ani na jiném neznámém parametru) a které splňují

$$\mathsf{P}_\theta(D \leq \theta \leq H) \geq 1-\alpha \quad \text{pro všechna } \theta \in \Theta.$$

Říkáme, že interval $[D, H]$ pokryje θ s pravděpodobností alespoň $1-\alpha$.

Někdy je obtížné nalézt přesný intervalový odhad a spokojíme se s asymptotickým intervalovým odhadem, který uvedenou podmínku splňuje pro $n \to \infty$.

Konstrukce intervalových odhadů: Neexistuje jednoznačný návod, jak sestrojit intervalový odhad, ale ve většině případů vycházíme z následujících tvrzení:

– *Centrální limitní věta:* Nechť $\{X_n\}$ je posloupnost nezávislých stejně rozdělených náhodných veličin s $0 < \mathsf{Var}\, X_1 < \infty$. Pak

$$\sqrt{n} \cdot \frac{\overline{X}_n - \mathsf{E}X_1}{\sqrt{\mathsf{Var}\, X_1}} \xrightarrow{D} \mathsf{N}(0,1).$$

– *Sluckého věta:* Nechť $\{X_n\}, \{Y_n\}, \{Z_n\}$ jsou posloupnosti náhodných veličin takových, že $X_n \xrightarrow{D} \mathsf{N}(0,1)$, $Y_n \xrightarrow{P} c$ a $Z_n \xrightarrow{P} d$, kde $c, d \in \mathbb{R}$ jsou reálné konstanty. Pak $Y_n \cdot X_n + Z_n \xrightarrow{D} \mathsf{N}(d, c^2)$. Speciálně, jestliže $\{T_n\}$ pro nějaké $a \neq 0$ splňuje $\sqrt{n}\frac{(T_n - \theta)}{a} \xrightarrow{D} \mathsf{N}(0,1)$ a $V_n \xrightarrow{P} a$, pak

$$\frac{\sqrt{n}(T_n - \theta)}{V_n} \xrightarrow{D} \mathsf{N}(0,1).$$

Momentová vytvořující funkce náhodné veličiny X je definována jako

$$\psi_X(t) = \mathsf{E}e^{tX}$$

pro všechna $t \in \mathbb{R}$, pro které má výraz na pravé straně smysl. Jestliže je ψ_X konečná na nějakém okolí nuly, pak platí $\mathsf{E}X^k = \psi_X^{(k)}(0)$.

Markovova nerovnost. Nechť $X \geq 0$ je náhodná veličina, $a > 0$ je libovolné a $k \in \mathbb{N}$. Pak

$$\mathsf{P}(X > a) \leq \frac{\mathsf{E}X^k}{a^k}.$$

Chernoffova mez. Nechť X je náhodná veličina a $a \in \mathbb{R}$, pak platí

$$\mathsf{P}(X \geq a) \leq \frac{\psi_X(t)}{e^{ta}}, \quad \text{pro všechna } t > 0$$

a tedy

$$\mathsf{P}(X \geq a) \leq \inf_{t>0} \frac{\psi_X(t)}{e^{ta}}.$$