Intervalové odhady

9.1.2018

0. Nechť $X_1, \dots, X_n$ je náhodný výběr se střední hodnotou μ a rozptylem $\sigma^2 \in (0, \infty)$.
   (a) Nechť σ^2 je známé. Zdůvodněte, že pak
   $$\mathsf{P}\left(-u_{1-\alpha/2} \le \sqrt{n}\frac{\overline{X}_n - \mu}{\sigma} \le u_{1-\alpha/2}\right) \to 1 - \alpha, \quad n \to \infty,$$
   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) Na čem závisí délka intervalového odhadu z (a)?
   (c) Analogicky odvoďte intervalový odhad pro σ^2 neznámé.

1. V hospodě jsme zakoupili 10 piv a zaznamenali jsme objem natočeného piva (viz jedno z minulých cvičení). Dostali jsme $\overline{X}_{10} = 0,4893$, $S_{10} = 0,0197$.
   (a) Zkonstruujte intervalový odhad se spolehlivostí 95 % pro střední hodnotu natočeného piva. Překrývá tento interval hodnotu 0.5?
   (b) Proveďte totéž pro spolehlivost 99% a porovnejte s (a).

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 směrodatná odchylka vyšla 1.64. Zkonstruujte intervalový odhad se spolehlivostí 95% pro počet vstřelených gólů v jednom zápase.

3. Bylo provedeno měření IQ náhodně vybraných žáků ZŠ, a to u 50 děvčat a 50 chlapců. Pro děvčata bylo naměřeno průměrné IQ 110,82 (směr. odch. 13,69) a pro chlapce byl průměr 107,66 (směr. odch. 15,71). Zkonstruujte interval spolehlivosti pro rozdíl středních hodnot IQ dívek a chlapců. Uvažujte spolehlivost 95 %.
   Pokrývá tento interval nulu? Co by to znamenalo, kdyby jej nepokrýval?

4. Ve fiktivní anketě o volebních preferencích ze 100 náhodně dotázaných 35 odpovědělo, že bude v blížících se volbách volit kandidáta X.
   (a) Zkonstruujte intervalový odhad se spolehlivostí 95 % pro pravděpodobnost p, s jakou občan ČR volí kandidáta X. Je pravděpodobné, že bude kandidát X zvolen v prvním kole?
   (b) Jak se změní intervalový odhad, pokud se zeptáme 1000 osob a 350 z nich budou voliči kandidáta X?
   (c) Jestliže podíl voličů zůstane stejný jako v (a) a (b), kolik osob se musíme zeptat, aby šířka intervalu byla maximálně 2%?

Tabulka kvantilové funkce $\mathsf{N}(0,1)$

α	0.5	0.8	0.9	0.95	0.975	0.99	0.995
$u_\alpha = \Phi^{-1}(\alpha)$	0	0.842	1.282	1.645	1.960	2.326	2.576

Opakování z přednášky

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

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

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 $\dfrac{\sqrt{n}(T_n - \theta)}{a} \xrightarrow{D} \mathsf{N}(0,1)$ a $Y_n \xrightarrow{P} a$, pak

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

– Pokud je $\theta = \mathsf{E} X$, pak z CLV $\sqrt{n}(\overline{X}_n - \theta)/\sqrt{\mathsf{Var}\, X_1} \xrightarrow{D} \mathsf{N}(0,1)$. Intervalový odhad θ pak
  – získáme snadno, pokud je $\mathsf{Var}\, X_1$ známé číslo,
  – jestliže je $\mathsf{Var}\, X_1$ neznámé, pak můžeme použít Sluckého větu, kde využijeme, že výběrový rozptyl $S_n^2 \xrightarrow{P} \mathsf{Var}\, X_1$,
  – jestliže je $\mathsf{Var}\, X_1$ funkcí θ (např. Alt, Po, Exp), pak můžeme intervalový odhad získat také přímým výpočtem nebo využít, že $\overline{X}_n$ je konzistentním odhadem θ. V tomto případě tedy můžeme nalézt tři obecně různé intervalové odhady s asymptotickou spolehlivostí $1 - \alpha$.