Intervalové odhady

9.1.2019

1. 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
   $$\mathsf{P}\left(-u_{1-\alpha/2} \le \sqrt{n}\frac{\overline{X}_n - \mu}{\sigma} \le u_{1-\alpha/2}\right) = \mathsf{P}\left(\sqrt{n}\frac{|\overline{X}_n - \mu|}{\sigma} \le u_{1-\alpha/2}\right) \to 1 - \alpha, \quad n \to \infty,$$
   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) Prozkoumejte, na čem a jak závisí délka intervalového odhadu z (a).
   (c) Odvoďte intervalový odhad pro μ, je-li σ^2 neznámé.
   (d) V hospodě jsme zakoupili 10 piv a zaznamenali jsme objem natočeného piva s cílem zjistit, zda je skutečná střední hodnota natočeného piva rovna 0.5 l. Dostali jsme $\overline{X}_{10} = 0,480$, $S_{10} = 0,019$. 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.

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 (výb. směr. odch. 13,69) a pro chlapce byl průměr 107,66 (výb. 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?

Podmíněné rozdělení

5. Uvažujme hod dvěma kostkami. Označme X počet ok, které padnou na první kostce, Y počet ok, které padnou na druhé kostce, a $S = X + Y$ jejich součet. Určete podmínené rozdělení a podmíněnou střední hodnotu X za podmínky, že $S = 6$.

6. V předchozím příkladě určete podmíněné rozdělení a podmíněnou střední hodnotu $V = XY$ za podmínky $S = 6$.

7. Na zkoušku dorazí N studentů, kde N je náhodná veličina s Poissonovým rozdělením s parametrem $\lambda > 0$. Každý student (nezávisle na ostatních) neuspěje u zkoušky s pravděpodobností $p \in (0, 1)$. Nechť Z je počet studentů, kteří i zkoušky neuspějí.
   (a) Jaké je podmíněné rozdělení Z při daném $N = n$?
   (b) Jaké je nepodmíněné rozdělení Z?
   (c) Jaké je podmíněné rozdělení N za podmínky $Z = 5$?

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 $1-\alpha$.

Podmíněné rozdělení. Nechť $(X, Y)'$ je náhodný vektor s diskrétním rozdělením s hodnotami v $\mathbb{N}^2$ a nechť $n \in \mathbb{N}$ je takové, že $\mathsf{P}(Y = n) > 0$. Pak definujeme podmíněné rozdělení X za podmínky $Y = n$ jako

$$\mathsf{P}(X = k | Y = n) = \frac{\mathsf{P}(X = k, Y = n)}{\mathsf{P}(Y = n)}.$$

Je-li $\mathsf{E}X < \infty$, pak definujeme podmíněnou střední hodnotu X za podmínky $Y = n$ jako $\mathsf{E}[X | Y = n] = \sum_{k \in \mathbb{N}} k \mathsf{P}(X = k | Y = n)$.