Bodový a intervalový odhad

17. – 21. 12. 2012

Bodový odhad

1. Nechť $X_1, \dots, X_n$ je náhodný výběr z normálního rozdělení, tj. z hustoty

$$f_{\mu,\sigma^2}(x) = \frac{1}{\sqrt{2\,\pi\,\sigma^2}} \exp\{-\frac{(x-\mu)^2}{2\,\sigma^2}\}, \quad x \in \mathbb{R},$$

kde oba parametry μ i σ^2 jsou neznámé. Najděte maximálně věrohodný odhad obou parametrů (současně).

Intervalový odhad

2. Nechť $X_1, \dots, X_n$ značí hodnoty IQ náhodně vybraných žáků 8. třídy. Předpokládejme, že $X_1, \dots, X_n$ tvoří náhodný výběr z normálního rozdělení s neznámou střední hodnotou a rozptylem 9.
   - (a) Odhadněte bodově typické IQ žáků 8.třídy. Jaké jsou vlastnosti a rozdělení tohoto odhadu?
   - (b) Sestrojte intervalový odhad o spolehlivosti $1 - \alpha$ pro typickou hodnotu IQ žáků 8.třídy. Interpretujte.
   - (c) Jak se bude tento interval měnit, budeme-li zvyšovat spolehlivost $1 - \alpha$?
     Jak se bude tento interval měnit, budeme-li zvyšovat počet dětí zahrnutých do experimentu (a ostatní hodnoty ve vzorcích zůstanou stejné)?
   - (d) Po provedení měření jsme obdrželi následující hodnoty:

     111, 116, 105, 111, 110, 114, 108, 106, 112, 108, 112, 111, 105, 111, 108, 110.

     Jak byste na základě těchto konkrétních hodnot odhadli typické IQ žáků 8.třídy? Uveďte bodový i 95% intervalový odhad.
   - (e) Napište 95% dolní intervalový odhad pro typické IQ žáků 8.třídy a interpretujte jej.

3. Průzkum veřejného mínění měl za úkol zjistit názor občanů ČR na přímou volbu prezidenta republiky. Do studie bylo zahrnuto $n = 400$ občanů, z nichž 240 uvedlo, že půjde volit.
   - (a) Odhadněte bodově podíl občanů ČR, půjdou volit. Jaký model používáte? Jaké jsou vlastnosti, rozdělení a asymptotické rozdělení tohoto bodového odhadu?
   - (b) Určete asymptotický 95% interval spolehlivosti pro podíl občanů, kteří půjdou volit.
   - (c) Jaký by musel být rozsah výběru n, aby intervalový odhad s asymptotickou spolehlivostí 95% pro tento podíl měl šířku nejvýše 0,03 (uvažujeme-li, že podíl občanů zahrnutých ve studii, kteří půjdou volit, se nezmění).

Opakování z přednášky

Model: $X_1, \dots, X_n$ je náhodný výběr (nezávislé stejně rozdělené veličiny) z rozdělení F_θ, které závisí na neznámém parametru $\theta \in \Theta$.

Minulé cvičení: bodový odhad parametrické funkce $g(\theta) \leftrightsquigarrow$ náhodná veličina
+ jeho „pěkné" vlastnosti $\leftrightsquigarrow$ nestrannost, konzistence

Intervalový odhad parametrické funkce $g(\theta) \leftrightsquigarrow$ interval s náhodnými mezemi, který obsahuje skutečnou (neznámou) hodnotu $g(\theta)$ s předepsanou pravděpodobností $1-\alpha$

– Intervalový odhad parametrické funkce $g(\theta)$ o spolehlivosti $1-\alpha$ je dvojice náhodných veličin (L_n, U_n) takových, že

$$P_\theta(L_n < g(\theta) < U_n) = 1 - \alpha \quad \text{pro všechna } \theta \in \Theta.$$

– Veličina $D_n = D_n(X_1, \dots, X_n)$ se nazývá dolní intervalový odhad $g(\theta)$ o spolehlivosti $1-\alpha$, jestliže

$$P_\theta(D_n < g(\theta)) = 1 - \alpha \quad \text{pro všechna } \theta \in \Theta.$$

– Veličina $H_n = H_n(X_1, \dots, X_n)$ se nazývá horní intervalový odhad $g(\theta)$ o spolehlivosti $1-\alpha$, jestliže

$$P_\theta(g(\theta) < H_n) = 1 - \alpha \quad \text{pro všechna } \theta \in \Theta.$$

Podobně jako jsme měli u bodového odhadu, $L_n = L_n(X_1, \dots, X_n)$, $U_n = U_n(X_1, \dots, X_n)$, $D_n = D_n(X_1, \dots, X_n)$ a $H_n = H_n(X_1, \dots, X_n)$ jsou (měřitelné) funkce náhodného výběru $X_1, \dots, X_n$, jejichž funkční předpisy nezávisí na θ.

Číslo $\alpha \in (0,1)$ volíme, v praxi se většinou pracuje s $\alpha = 0{,}05$.

Obecná konstrukce intervalu spolehlivosti pro parametrickou funkci $g(\theta)$. Najdeme funkci náhodného výběru a parametrické funkce $g(\theta)$, tj. náhodnou veličinu $h(X_1, \dots, X_n; g(\theta))$, jejíž rozdělení nezávisí na θ. Nechť $h_{\alpha/2}$ a $h_{1-\alpha/2}$ jsou kvantily tohoto rozdělení. Pak jistě platí

$$P_\theta\left(h_{\alpha/2} < h(X_1, \dots, X_n; g(\theta)) < h_{1-\alpha/2}\right) = 1 - \alpha \quad \text{pro všechna } \theta \in \Theta.$$

Je-li možné nerovnosti v závorce převést ekvivalentními úpravami na takový tvar, že uprostřed stojí $g(\theta)$ a vlevo i vpravo je něco, co na θ nezávisí, sestrojili jsme intervalový odhad.