Odhady

27.11.2019

1. Uvažujme společenskou hru "Tik Tak Bum", ve které se využívá elektronická bomba, která náhodný čas tiká (a během tohoto času hráči střídavě vymýšlejí slova a bombu si předávají) a následně bomba zvukově vybuchne. Princip hry je založen na tom, že nikdo dopředu neví, jak dlouho bude bomba tikat, než vybuchne. Označme jako X náhodnou veličinu udávající čas od spuštění do výbuchu bomby. Tato veličina má tedy neznámé rozdělení s neznámou distribuční funkcí F. Dále lze předpokládat, že náhodná veličina X má spojité rozdělení s hustotou f, která je nulová mimo interval $[0, b]$, kde $b > 0$ je neznámé číslo.
   (a) Nejprve bychom rádi odhadli střední dobu tikání bomby. Navrhněte vhodný postup a vhodný odhad. Je Váš navržený odhad nestranný?
   (b) Pomocí čeho bychom mohli (alespoň přibližně) ověřit, zda lze předpokládat, že má X rovnoměrné rozdělení na $[0, b]$?
   (c) Navrhněte, jak bychom mohli odhadnout b. Jedná se o nestranný odhad?
   (d) Navrhněte postup, jak bychom mohli odhadnout pravděpodobnost, že bomba bude tikat méně než 10 sekund. Spočtěte střední hodnotu a rozptyl takového odhadu.

2. V rybníce plave 100 rybiček, z nichž a je zlatých, kde $a > 0$ je neznámé číslo. Rádi bychom odhadli a. Budeme uvažovat následující postupy a odhady.
   (a) Vylovíme postupně s vracením 10 rybiček a pro každou zaznamenáme, zda je zlatá. Odhadněte na základě tohoto pokusu neznámé a. Dále vyšetřete střední hodnotu a rozptyl tohoto odhadu.
   (b) Nyní vylovíme najednou 10 rybiček (tj. bez vracení). Odhadněte na základě tohoto pokusu neznámé a a opět vyšetřete střední hodnotu a rozptyl tohoto odhadu.
   (c) Nyní budeme postupně lovit rybičky s vracením, ale pouze dokud nevylovíme první zlatou rybičku. Zaznamenáme si počet pokusů. Navrhněte odhad a z těchto dat.
   (d) Provedeme podobný pokus jako v předchozím bodě, ale nyní lovíme, dokud nevylovíme popáté zlatou rybičku (do rybníka vracíme i zlaté rybičky). Navrhněte odhad a z těchto dat.

3. Na základě dat z našeho cvičení odhadněte následující kvantity. Zároveň diskutujte, zda je skutečně vhodné brát výsledky z našeho cvičení jako odhady daných kvantit.
   - Kolik procent studentů informatiky MFF píše levou rukou.
   - Kolik procent studentů informatiky MFF tvoří dívky.
   - Kolik procent studentů informatiky MFF tvoří zahraniční studenti.

Opakování

Náhodný výběr: $X_1, \dots, X_n$ jsou nezávislé a stejně rozdělené náhodné veličiny s rozdělením s distribuční funkcí F.

Funkce náhodného výběru. Nechť $g_n : \mathbb{R}^n \to \mathbb{R}$ je (měřitelná) funkce. Pak $T_n = g_n(X_1, \dots, X_n)$ je náhodná veličina. Jestliže navíc předpis g_n nezávisí na neznámých parametrech rozdělení F, pak můžeme T_n brát jako tzv. **bodový odhad**.

Příklady různých funkcí náhodného výběru:

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i, \quad \min_{1\leq i\leq n} X_i, \quad \max_{1\leq i\leq n} X_i, \quad \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}[X_i \leq 1], \quad S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2 \dots,$$

kde $\mathbb{I}[A]$ je identifikátor jevu A, tj. $\mathbb{I}[A] = 1$, pokud A nastal, a $\mathbb{I}[A] = 0$ jinak.

Vlastnosti odhadů. Nechť $X_1, \dots, X_n$ je náhodný výběr z rozdělení, které závisí na neznámém parametru $\theta \in \Theta$. Nechť $T_n = T_n(X_1, \dots, X_n)$ je odhad θ. V praxi chceme pracovat pouze s odhady, které mají „pěkné“ vlastnosti:

- Řekneme, že odhad T_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 je **konzistentní** odhad θ, jestliže $T_n \xrightarrow{P} \theta$ (konvergence v pravděpodobnosti) pro $n \to \infty$ pro všechna $\theta \in \Theta$, tj. platí, že pro všechna $\varepsilon > 0$

$$\mathsf{P}_\theta(|T_n - \theta| > \varepsilon) \to 0 \qquad \text{pro } n \to \infty \qquad \text{pro všechna } \theta \in \Theta.$$

Empirická distribuční funkce: Je-li $X_1, \dots, X_n$ náhodný výběr, pak empirická distribuční funkce $\widehat{F_n}$ je pro $x \in \mathbb{R}$ dána jako

$$\widehat{F_n}(x) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}[X_i \leq x].$$

Každému $x \in \mathbb{R}$ se tedy přiřadí náhodná veličina $\widehat{F_n}(x)$. $\widehat{F_n}(x)$ je nestranný a konzistentní odhad hodnoty distribuční funkce $F(x)$.

Konstrukce odhadu momentovou metodou: Předpokládejme, že $\mathsf{E}X_i = g(\theta)$ pro nějakou známou spojitou funkci g. Pak momentový odhad $\widehat{\theta}_n$ parametru θ spočítáme tak, že

$$\overline{X}_n = g(\widehat{\theta}_n),$$

kde $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ je výběrový průměr.
Odhadujeme-li více parametrů, nebo pokud $\mathsf{E}X_i$ na θ nezávisí, zapojíme do výběrové momenty vyšších řádů, tj. $\frac{1}{n}\sum_{i=1}^{n} X_i^k$.