Náhodný výběr

21.11.2018

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.
   (a) Nejprve bychom rádi odhadli střední dobu tikání bomby. Navrhněte vhodný postup a vhodný odhad.
   (b) Spočtěte střední hodnotu a rozptyl odhadu z (a).
   Jak se budou tyto hodnoty měnit, budeme-li zvyšovat počet pozorování?
   (c) Zřejmě lze předpokládat, že náhodná veličina X má spojité rozdělení s hustotou f na intervalu $[a, b]$, kde $0 \leq a < b$ jsou nějaká neznámá reálná čísla. Navrhněte, jak bychom mohli odhadnout b (resp. a). Vyjádřete rozdělení (distribuční funkci a hustotu) těchto odhadů pomocí F a f.
   (d) Kdyby rozdělení X bylo rovnoměrné na $[a, b]$, jaká by byla střední hodnota odhadů z (c)?
   (e) Navrhněte postup, jak bychom mohli odhadnout pravděpodobnost, že bomba bude tikat méně než 10 sekund. Spočtěte střední hodnotu a rozptyl takového odhadu.
   (f) Pomocí čeho bychom mohli (alespoň přibližně) ověřit, zda je předpoklad rovnoměrného rozdělení z (d) vhodný pro naši veličinu X?

2. Náhodná veličina X je indikátor jevu, zda náhodně vybraná osoba píše primárně levou rukou.
   (a) Jaké rozdělení má X? Co vyjadřuje parametr tohoto rozdělení? Jaká je střední hodnota a rozptyl X?
   (b) Navrhněte, jak bychom na našem cvičení mohli odhadnout pravděpodobnost, že je náhodně vybraná osoba levák.
   (c) Jaká je střední hodnota a rozptyl odhadu z (b)?
   (d) Experiment proveďte. Porovnejte obdržený výsledek s obecně uváděných faktem, že leváků je v populaci 10 %.

3. Porovnejte příklady 1. a 2.: jak se liší náš přístup k „ modelování“ rozdělení X?

4. Uvažujte odhady z 1.(d). Vyšetřete chování jejich rozdělení (distribuční funkce) pro $n \to \infty$.

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:

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

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

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

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

Každému $x \in \mathbb{R}$ se tedy přiřadí náhodná veličina $\widehat{F_n}(x)$.
- $\widehat{F_n}(x)$ slouží jako odhad hodnoty distribuční funkce $F(x)$.
- Pro pevné x má $n\widehat{F_n}(x)$ binomické rozdělení $\mathsf{Bi}(n, F(x))$.

Vlastnosti odhadů. Nechť $X_1, \dots, X_n$ je náhodný výběr z rozdělení, které závisí na neznámém parametru $\theta \in \Theta$ (např. $\mathsf{Alt}(\theta)$, $\mathsf{Po}(\theta)$, $\mathsf{Exp}(\theta)$ aj.). 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 $\varepsilon > 0$

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