Bodové odhady

2. 5. a 3. 5. 2019

1. Lze předpokládat, že počet dopravních nehod na ulici Sokolovská se řídí Poissonovým rozdělením s parametrem $\lambda > 0$ a že počty nehod v jednotlivých dnech jsou nezávislé a stejně rozdělené náhodné veličiny. Plánujeme zaznamenávat počty nehod po následujících n dní a tím dostaneme náhodný výběr $X_1, \ldots, X_n$ z $\mathsf{Po}(\lambda)$.

   (a) Nalezněte odhad parametru λ momentovou metodou a zjistěte, zda je tento odhad nestranný a konzistentní.
   (b) Odvoďte odhad parametru λ metodou maximální věrohodnosti. Vyšetřete jeho vlastnosti.
   (c) Uvažujte odhady $\check{\lambda}_n = (X_1 + X_n)/2$ a $\widetilde{\lambda}_n = \frac{1}{n-1}\sum_{i=1}^n X_i$. Zjistěte, zda jsou tyto odhady nestranné a konzistentní odhady λ.

   Dále nás bude zajímat odhad pravděpodobnosti, že se v daný den nestane žádná nehoda, tj. odhad $p_0 = \mathsf{P}(X_1 = 0) = \mathrm{e}^{-\lambda}$. Uvažujme následující tři odhady p_0:

   $$W_n = \frac{\sum_{i=1}^n I[X_i = 0]}{n}, \quad V_n = e^{-\overline{X}_n}, \quad U_n = \left(1 - \frac{1}{n}\right)^{n\overline{X}_n}.$$

   (d) Vyšetřete vlastnosti odhadu W_n.
   (e) Zjistěte, zda je V_n nestranný (resp. asymptoticky nestranný) a konzistentní odhad p_0.
   (f) Zjistěte, zda je U_n nestranný a konzistentní odhad p_0.
   (g) Porovnejte vlastnosti odhadů W_n, U_n a V_n.
   (h) Po 30 dnech pečlivého měření jsme nakonec obdrželi následující data:

Počet nehod	0	1	2	3	4	5	6
Počet dní	5	8	8	3	2	3	1

   Pomocí předchozích výsledků odhadněte parametr λ a pravděpodobnost, že se v náhodně vybraný den nestane žádná nehoda.

2. Nechť $X_1, \ldots, X_n$ je náhodný výběr z alternativního rozdělení s parametrem $p \in (0, 1)$, tj.

   $$\mathsf{P}(X_i = k) = p^k(1-p)^{1-k}, \quad k = 0, 1.$$

   (a) Nalezněte odhady p momentovou metodou a metodou maximální věrohodnosti a vyšetřete jejich vlastnosti.
   (b) Uvažujte $T_n = \overline{X}_n(1-\overline{X}_n)$ odhad rozptylu $\mathsf{var}\, X_1 = p(1-p)$. Zjistěte, zda je tento odhad nestranný a konzistentní. Pokud není nestranný, navrhněte, jak jej modifikovat, abychom dostali nestranný odhad.

3. Nechť $X_1, \ldots, X_n$ je náhodný výběr ze spojitého rozdělení s hustotou $f(x) = \dfrac{1}{\sqrt{\pi b}}\mathrm{e}^{-x^2/b}$ pro $x \in \mathbb{R}$, kde $b > 0$ je neznámý parametr. Odhadněte b metodou maximální věrohodnosti a vyšetřete vlastnosti tohoto odhadu.

Bodový odhad. Nechť $X_1, \ldots, X_n$ je náhodný výběr (nezávislé stejně rozdělené veličiny) z rozdělení, které závisí na neznámém parametru $\theta \in \Theta$. Odhadem parametru θ (nebo obecněji parametrické funkce $g(\theta)$) je libovolná borelovská funkce náhodného výběru $X_1, \ldots, X_n$, jejíž funkční předpis nezávisí na θ. **Odhad** $T_n = T_n(X_1, \ldots, X_n)$ je tedy **náhodná veličina**.

V praxi pak pracujeme pouze s odhady, které mají „pěkné“ vlastnosti:

- Řekneme, že odhad $T_n = T_n(X_1, \ldots, X_n)$ je **(silně) konzistentní** odhad parametrické funkce $g(\theta)$, jestliže $T_n \to g(\theta)$ s.j. pro $n \to \infty$ pro všechna $\theta \in \Theta$, tj.

$$\mathsf{P}_\theta\big(\lim_{n\to\infty} T_n = g(\theta)\big) = 1 \qquad \text{pro všechna } \theta \in \Theta.$$

- Řekneme, že odhad $T_n = T_n(X_1, \ldots, X_n)$ je **nestranný** odhad parametrické funkce $g(\theta)$, jestliže

$$\mathsf{E}\, T_n = \mathsf{E}_\theta\, T_n = g(\theta), \qquad \text{pro všechna } \theta \in \Theta.$$

  Odhad T_n je **asymptoticky nestranný** odhad $g(\theta)$, jestliže $\mathsf{E}_\theta\, T_n \to g(\theta)$ při $n \to \infty$ pro všechna $\theta \in \Theta$.

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

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

kde $\overline{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$ je výběrový průměr.

Konstrukce odhadu metodou maximální věrohodnosti. Nechť $X_1, \ldots, X_n$ je náhodný výběr z rozdělení, které závisí na neznámém parametru $\theta \in \Theta$. Předpokládejme, že toto rozdělení má hustotu $f(x;\theta)$ vzhledem k nějaké σ-konečné míře ν (ν bude čítací míra v případě diskrétního rozdělení a Lebesgueova míra v případě spojitého rozdělení). Odhad $T_n = T_n(X_1, \ldots, X_n)$ je **odhad metodou maximální věrohodnosti**, jestliže T_n maximalizuje tzv. věrohodnost

$$L_n(\theta) = \prod_{i=1}^n f(X_i;\theta)$$

přes všechna $\theta \in \Theta$. Ekvivalentně, T_n maximalizuje logaritmickou věrohodnost

$$l_n(\theta) = \log[L_n(\theta)] = \sum_{i=1}^n \log[f(X_i;\theta)].$$

Máme tedy $T_n = \arg\max_{\theta\in\Theta} L_n(\theta) = \arg\max_{\theta\in\Theta} l_n(\theta)$.

Za určitých dalších předpokladů lze T_n najít jako řešení věrohodnostní rovnice $\frac{\mathrm{d}}{\mathrm{d}\theta} l_n(\theta) = 0$, tj.

$$\sum_{i=1}^n \frac{\mathrm{d}}{\mathrm{d}\theta} \log[f(X_i;\theta)] = 0.$$

Věta o spojité transformaci. Nechť $T_n \xrightarrow[n\to\infty]{s.j.} \theta$ a g je spojitá funkce v bodě θ. Pak $g(T_n) \xrightarrow[n\to\infty]{s.j.} g(\theta)$.