10. cvičení z NMSA202
23. 4. 2015 & 27. 4. 2015

Bodový odhad II.

1. Nechť $X_1, \ldots, X_n$ je náhodný výběr z Poissonova rozdělení s parametrem $\lambda > 0$.
   (a) Odvoďte odhad T_n parametru λ metodou maximální věrohodnosti.
   (b) Zjistěte, zda je tento odhad T_n nestranný a konzistentní odhad λ.
   (c) Uvažujme odhad
   $$U_n = \left(1 - \tfrac{1}{n}\right)^{n\overline{X}_n}$$
   pravděpodobnosti $p_0 = \mathsf{P}(X_1 = 0)$. Zjistěte, zda je U_n nestranný odhad p_0.
   (d) Zjistěte, zda je U_n konzistentní odhad p_0.
   (e) Zjistěte, zda je $V_n = e^{-\overline{X}_n}$ nestranný (resp. asymptoticky nestranný) a konzistentní odhad p_0.
   (f) Zjistěte, zda je odhad $W_n = \frac{1}{2}(U_n + V_n)$ nestranný (resp. asymptoticky nestranný) a konzistentní odhad p_0.
2. Sledujeme nehodovost na ulici Sokolovská. Lze předpokládat, že počet nehod v daný den má Poissonovo rozdělení s parametrem $\lambda > 0$ a že nehodovosti v různých dnech jsou nezávislé a stejně rozdělené. Po 30 dnech pečlivého měření jsme obdrželi následující data:

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

   Pomocí výsledků příkladu 1 odhadněte parametr λ a pravděpodobnost, že se v náhodně vybraný den nestane žádná nehoda.
3. Nechť $X_1, \ldots, X_n$ je náhodný výběr z normálního rozdělení s hustotou
   $$f_{\mu,\sigma^2}(x) = \frac{1}{\sqrt{2\,\pi\,\sigma^2}} \exp\left\{-\frac{(x-\mu)^2}{2\,\sigma^2}\right\}, \quad x \in \mathbb{R},$$
   kde oba parametry μ i $\sigma^2 > 0$ jsou neznámé. Najděte maximálně věrohodný odhad obou parametrů (současně) a vyšetřete jeho vlastnosti.
4. Nechť $X_1, \ldots, X_n$ je náhodný výběr z geometrického rozdělení, tj.
   $$\mathsf{P}(X_i = k) = (1-p)^k p, \qquad k = 0, 1, 2, \ldots$$
   Najděte maximálně věrohodný odhad parametru p a vyšetřete jeho konzistenci.
5. Každý stírací los vyhraje s (neznámou) pravděpodobností p, kterou bychom rádi odhadli. Za tímto se nám podařilo přesvědčit deset svých kamarádů, aby si kupovali losy tak dlouho, než narazí na výherní los. Počet nevýherních losů, které předcházely prvnímu výhernímu losu, zachycuje následující tabulka

Číslo kamaráda	1	2	3	4	5	6	7	8	9	10
Počet nevýherních losů	8	1	5	1	6	23	2	7	13	8

   Na základě výsledků předchozího příkladu odhadněte parametr p.

Opakování z přednášky

- Řekneme, že odhad $T_n = T_n(X_1, \dots, X_n)$ je **nestranný** odhad parametrické funkce $g(\theta)$, jestliže
$$\mathsf{E}\, T_n = \mathsf{E}_\theta\, T_n = g(\theta), \qquad \text{pro všechna } \theta \in \Theta. \tag{10.1}$$
Odhad T_n je **asymptoticky nestranný** odhad $g(\theta)$, jestliže $\mathsf{E}_\theta T_n \to g(\theta)$ pro $n \to \infty$ pro všechna $\theta \in \Theta$.
- Řekneme, že odhad $T_n = T_n(X_1, \dots, X_n)$ je **konzistentní** odhad parametrické funkce $g(\theta)$, jestliže $T_n \to g(\theta)$ s.j. pro $n \to \infty$ pro všechna $\theta \in \Theta$, t.j.
$$\mathsf{P}_\theta\Big(\lim_{n\to\infty} T_n = g(\theta)\Big) = 1 \qquad \text{pro všechna } \theta \in \Theta. \tag{10.2}$$

Stručný popis konstrukce odhadu metodou maximální věrohodnosti. Nechť $X_1, \dots, X_n$ 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 Lebesguova míra v případě spojitého rozdělení). Odhad $T_n = T_n(X_1, \dots, X_n)$ je **odhad metodou maximální věrohodnosti**, jestliže T_n maximalizuje tzv. věrohodnost
$$L_n(\theta) = \prod_{i=1}^{n} f(X_i, \theta) \tag{10.3}$$
přes všechna $\theta \in \Theta$. Ekvivalentně, T_n maximalizuje logaritmickou věrohodnost
$$l_n(\theta) = \log\big[L_n(\theta)\big] = \sum_{i=1}^{n} \log\big[f(X_i, \theta)\big]. \tag{10.4}$$
Máme tedy $T_n = \arg\max_{\theta\in\Theta} L(\theta) = \arg\max_{\theta\in\Theta} l(\theta)$.

Za určitých dalších předpokladů lze T_n najít jako řešení věrohodnostní rovnice $\frac{\mathrm{d}}{\mathrm{d}\theta}\, l_n(\theta) = 0$, tj.
$$\frac{\mathrm{d}}{\mathrm{d}\theta} \sum_{i=1}^{n} \log\big[f(X_i, \theta)\big] = 0. \tag{10.5}$$

Výsledky

1. (a) $T_n = \overline{X}_n$,
   (b) T_n je nestranný i konzistentní odhad λ,
   (c) U_n je nestranný odhad $p_0 = \mathrm{e}^{-\lambda}$,
   (d) U_n je konzistentní odhad p_0,
   (e) V_n není nestranný, ale je asymptoticky nestranný a konzistentní odhad p_0.
   (f) W_n není nestranný, ale je asymptoticky nestranný a konzistentní odhad p_0.
2. $\widehat{\lambda} = \overline{x}_n = \frac{1}{n}\sum_{i=1}^n x_i = 62/30 \doteq 2{,}0667$, $\widehat{p}_0 = \left(1 - \frac{1}{n}\right)^{n\,\overline{x}_n} = (29/30)^{62} = 0{,}1222$ nebo $\overline{p}_0 = \mathrm{e}^{-\overline{x}_n} = \mathrm{e}^{-62/30} = 0{,}1266$.
3. $\widehat{\mu}_n = \frac{1}{n}\sum_{i=1}^n X_i$, $\widehat{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^n (X_i - \bar{X}_n)^2$. Odhad $\widehat{\mu}_n$ je nestranný a konzistentní odhad parametru μ. Odhad $\widehat{\sigma}_n^2$ je konzistentní odhad parametru σ^2, ale není nestranný.
4. $\widehat{p}_n = \frac{1}{1+\frac{1}{n}\sum_{i=1}^n X_i}$, odhad je konzistentní.
5. Za předpokladu, že se naše data dají považovat za realizaci náhodného výběru z geometrického rozdělení, pak maximálně věrohodný odhad parametru p je $\widehat{p} = \frac{1}{1+\frac{1}{n}\sum_{i=1}^n x_i} = \frac{1}{1+7{,}4} \doteq 0{,}119$.