Rešitve kolokvija iz verjetnosti in statistike z dne 7. 6. 2012

Matematika – univerzitetni študij

1. Dana porazdelitev ima gostoto:

$$f(x) = \frac{1}{a\sqrt{2\pi}}\, e^{-(x-a)^2/(2a^2)} = \frac{1}{a\sqrt{2\pi}} \exp\left(-\frac{x^2}{2a^2} + \frac{x}{a} - \frac{1}{2}\right),$$

kar nam da verjetje:

$$L = \frac{1}{a^n (2\pi)^{n/2}} \exp\left(-\frac{1}{2a^2}\sum_{i=1}^n X_i^2 + \frac{1}{a}\sum_{i=1}^n X_i - \frac{n}{2}\right).$$

Le-to je ves čas strogo pozitivno in gre proti nič, brž ko gre a proti nič ali proti neskončno. Ugodno je gledati njegov logaritem:

$$\ln L = -\frac{n}{2}\ln(2\pi) - n\ln a - \frac{1}{2a^2}\sum_{i=1}^n X_i^2 + \frac{1}{a}\sum_{i=1}^n X_i - \frac{n}{2},$$

ki ima odvod:

$$\frac{\mathrm{d}\ln L}{\mathrm{d}a} = -\frac{n}{a} + \frac{1}{a^3}\sum_{i=1}^n X_i^2 - \frac{1}{a^2}\sum_{i=1}^n X_i\,.$$

Edina ničla odvoda na intervalu $(0, \infty)$ je:

$$\hat{a} = -\frac{1}{2n}\sum_{n=1}^{\infty} X_i + \sqrt{\left(\frac{1}{2n}\sum_{n=1}^{\infty} X_i\right)^2 + \frac{1}{n}\sum_{i=1}^{\infty} X_i^2}$$

in to je iskana cenilka za a.

2. Iz zapisa gostote:

$$f(x) = \frac{1}{x\,\mathrm{B}(a, 1-a)} \exp\left(a \ln\frac{x}{1-x}\right)$$

(za $0 < x < 1$) razberemo, da gre za eksponentno družino, katere parametrični prostor je odprta množica $(0, 1)$ (za vse a iz te množice ima porazdelitev smisel). Nepristranska cenilka z najmanjšo možno disperzijo bo torej vsaka funkcija pripadajoče zadostne statistike $\ln\frac{X}{1-X}$, ki bo nepristranska cenilka za a. Ker je funkcija $x \mapsto \ln\frac{x}{1-x}$ strogo naraščajoča, je dovolj iskati kar funkcije opažanja X. Iz:

$$\begin{aligned} E(X) &= \frac{1}{\mathrm{B}(a, 1-a)}\int_0^1 x^a (1-x)^{-a}\,\mathrm{d}x = \frac{\mathrm{B}(a+1, 1-a)}{\mathrm{B}(a, 1-a)} = \\ &= \frac{\Gamma(a+1)\,\Gamma(1-a)}{\Gamma(2)}\,\frac{\Gamma(1)}{\Gamma(a)\,\Gamma(1-a)} = \frac{\Gamma(a+1)}{\Gamma(a)} = a \end{aligned}$$

sledi, da je iskana cenilka kar opažanje samo, t. j. X.

3. Označimo dano statistično spremenljivko z X. Najprej opazimo, da je X pri $\theta = 1$ z verjetnostjo 1 enak 0, kar pomeni, da mora interval zaupanja pri opažanju $X = 0$ obvezno vsebovati tudi 1. Zaradi predpisane oblike mora biti torej to kar interval $[0, 1]$. Preostala dva parametra, b_1 in b_2, lahko nastavimo strogo manjša od 1.
Če nastavimo $b_1 > b_2$, velja:

$$P_\theta(\theta \in I) = \begin{cases} 1 & ; \theta < b_2 \\ P_\theta(X = 0) + P_\theta(X = 1) & ; b_2 \le \theta < b_1 \\ P_\theta(X = 0) & ; b_1 \le \theta \le 1 \end{cases}$$

oziroma:

$$P_\theta(\theta \in I) = \begin{cases} 1 & ; \theta < b_2 \\ \dfrac{\theta}{\theta^2 - \theta + 1} & ; b_2 \le \theta < b_1 \\ \dfrac{\theta^2}{\theta^2 - \theta + 1} & ; b_1 \le \theta \le 1 \end{cases} .$$

Ni se težko prepričati, da sta funkciji $\theta \mapsto \frac{\theta}{\theta^2 - \theta + 1}$ in $\theta \mapsto \frac{\theta^2}{\theta^2 - \theta + 1}$ na intervalu $[0, 1]$ obe naraščajoči, torej je:

$$\inf_{0 \le \theta \le 1} P_\theta(\theta \in I) = \min \left\{ \frac{b_2}{b_2^2 - b_2 + 1}, \frac{b_1^2}{b_1^2 - b_1 + 1} \right\} .$$

Pogoj $\inf_{0 \le \theta \le 1} P_\theta(\theta \in I) = \beta$ bo zagotovo izpolnjen, če bo:

$$\frac{b_2}{b_2^2 - b_2 + 1} = \frac{b_1^2}{b_1^2 - b_1 + 1} = \beta .$$

Ker sta funkciji $\theta \mapsto \frac{\theta}{\theta^2 - \theta + 1}$ in $\theta \mapsto \frac{\theta^2}{\theta^2 - \theta + 1}$ na intervalu $[0, 1]$ obe naraščajoči in ker za $\theta \in (0, 1)$ velja $\frac{\theta}{\theta^2 - \theta + 1} > \frac{\theta^2}{\theta^2 - \theta + 1}$, je pri zgornji izbiri tudi $b_1 > b_2$. Velja:

$$b_1 = \frac{-\beta + \sqrt{\beta(5\beta - 4)}}{2(1 - \beta)}, \quad b_2 = \frac{1 + \beta - \sqrt{(1 + 3\beta)(1 - \beta)}}{2\beta} .$$

Pri $\beta = 0{,}95$ dobimo $b_1 \doteq 0{,}9523$ in $b_2 \doteq 0{,}7954$ (obakrat smo zaokrožili navzgor). No, dovolj je, da je velja le ena enakost, drugi parameter pa se lahko poveča (a potem dobimo daljši interval zaupanja za enako stopnjo zaupanja).
Ni pa to edina možnost. Če postavimo $b_1 < b_2$, dobimo:

$$P_\theta(\theta \in I) = \begin{cases} 1 & ; \theta < b_2 \\ P_\theta(X = 0) + P_\theta(X = 2) & ; b_1 \le \theta < b_2 \\ P_\theta(X = 0) & ; b_2 \le \theta \le 1 \end{cases}$$

oziroma:

$$P_\theta(\theta \in I) = \begin{cases} 1 & ; \theta < b_1 \\ \dfrac{2\theta^2 - 2\theta + 1}{\theta^2 - \theta + 1} & ; b_1 \le \theta < b_2 \\ \dfrac{\theta^2}{\theta^2 - \theta + 1} & ; b_2 \le \theta \le 1 \end{cases} .$$

Kot smo že omenili, je funkcija $\theta \mapsto \frac{\theta^2}{\theta^2-\theta+1}$ na intervalu $[0,1]$ naraščajoča. Funkcija $\theta \mapsto \frac{2\theta^2-2\theta+1}{\theta^2-\theta+1}$ pa je na intervalu $\left[0,\frac{1}{2}\right]$ padajoča, na intervalu $\left[\frac{1}{2},1\right]$ pa naraščajoča; v $\frac{1}{2}$ ima minimum $\frac{2}{3}$. Zato mora biti $b_1 \geq \frac{1}{2}$, brž ko je $\beta \geq \frac{1}{2}$. V tem primeru velja:

$$\inf_{0\le\theta\le1} P_\theta(\theta \in I) = \min\left\{\frac{b_2}{b_2^2-b_2+1}, \frac{b_1^2}{b_2^2-b_2+1}\right\}.$$

Pogoj $\inf_{0\le\theta\le1} P_\theta(\theta \in I) = \beta$ bo zagotovo izpolnjen, če bo:

$$\frac{2b_1^2-2b_1+1}{b_1^2-b_1+1} = \frac{b_2^2}{b_2^2-b_2+1} = \beta\,.$$

Spet ker sta funkciji $\theta \mapsto \frac{2\theta^2-2\theta+1}{\theta^2-\theta+1}$ in $\theta \mapsto \frac{\theta^2}{\theta^2-\theta+1}$ na intervalu $\left[\frac{1}{2},1\right]$ obe naraščajoči in ker za $\theta \in \left(\frac{1}{2},1\right)$ velja $\frac{2\theta^2-2\theta+1}{\theta^2-\theta+1} > \frac{\theta^2}{\theta^2-\theta+1}$, je pri zgornji izbiri tudi $b_1 < b_2$. Velja:

$$b_1 = \frac{2-\beta+\sqrt{(3\beta-2)(2-\beta)}}{2(2-\beta)}\,, \quad b_2 = \frac{-\beta+\sqrt{\beta(5\beta-4)}}{2(1-\beta)}\,.$$

Pri $\beta = 0{\cdot}95$ dobimo $b_1 \doteq 0{\cdot}9499$ in $b_2 \doteq 0{\cdot}9523$ (obakrat smo zaokrožili navzgor). Ta izbira je torej slabša glede na seštevek dolžin pri vseh opazovanjih. Spet je podobno kot prej dovolj, da je velja le ena enakost, drugi parameter pa se lahko poveča (a potem dobimo še daljši interval zaupanja za enako stopnjo zaupanja).

Možno pa je izbrati tudi $b_1 = b_2$. V tem primeru dobimo:

$$b_1 = b_2 = \frac{-\beta+\sqrt{\beta(5\beta-4)}}{2(1-\beta)}$$

(torej $0{\cdot}9523$ pri $\beta = 0{\cdot}95$), kar je kombinacija prejšnjih dveh možnosti.

4. Označimo naše opažanje, t. j. število metov, pri katerih pade šestica, z X. Velja $X \sim \mathrm{b}(13,\theta)$, kar je enoparametrična eksponentna družina, katere naravni parameter je naraščajoča funkcija parametra θ, pripadajoča zadostna statistika pa je natančno X. Torej ničelno hipotezo zavrnemo, če je opažanje X preveliko. Ker je $P_{1/6}(X \geq 5) \doteq 0{\cdot}0512$, hipoteze ne moremo zavrniti – vsaj če testa ne randomiziramo.

Če pa dopustimo randomizacijo, nam verjetnost $P_{1/6}(X > 5) \doteq 0{\cdot}0127$ pove, da lahko hipotezo zavrnemo z verjetnostjo:

$$\frac{0{\cdot}05 - P_{1/6}(X>5)}{P_{1/6}(X\geq5) - P_{1/6}(X>5)} = \frac{0{\cdot}05 - P_{1/6}(X>5)}{P_{1/6}(X=5)} \doteq 0{\cdot}97\,.$$