7. STATISZTIKA GYAKORLAT

Zh-ra gyakorlás. Anyag:

- Rendezett minták.
- ML-becslés, elégségesség, teljesség, torzítatlanság
- Fisher-info, Cramér–Rao határ.
- Blackwellizálás.

ZH I. mintafeladatok

1. Legyen $X_1, \ldots, X_n$ fae. minta a $(0, \theta)$-n egyenletes eloszlásból.
   (a) $\mathbb{P}_\theta(X_n^* - X_1^* < \frac{\theta}{2}) = ?$
   (b) Adjon X_1^* felhasználásával torzítatlan becslést θ-ra!
   (c) Adjon elégséges statisztikát θ-ra!
   (d) Blackwellizálja a b.-beli statisztikát a c.-belivel!

2. Legyen $X_1, \ldots, X_n$ fae. minta az

$$f_\theta(x) = \frac{1}{\theta} \cdot x^{\frac{1-\theta}{\theta}}, \quad \text{ha} \quad 0 < x < 1$$

   (és 0 különben) sűrűségfüggvénnyel megadott eloszlásból, ahol $\theta > 0$ paraméter.
   (a) Keressen elégséges statisztikát θ-ra a minta alapján!
   (b) Adjon maximum likelihood becslést θ-ra a minta alapján!
   (c) Torzítatlan becslést ad-e a talált maximum likelihood becslés θ-ra?
   (d) Tagja-e a fenti eloszlás az exponenciális eloszlácsaládnak? Miért igen, ill. miért nem? $I_n(\theta) = ?$

3. Legyen $X_1, \ldots, X_n$ fae. minta az

$$f_\theta(x) = \frac{3x^2}{2\theta^3}, \quad \text{ha} \quad -\theta \le x \le \theta$$

   (és 0 különben) sűrűségfüggvénnyel megadott eloszlásból, ahol $\theta > 0$ paraméter.
   (a) Keressen elégséges statisztikát θ-ra a minta alapján!
   (b) Adjon maximum likelihood becslést θ-ra a minta alapján!
   (c) Tagja-e a fenti eloszlás az exponenciális eloszlácsaládnak? Miért igen, ill. miért nem?

4. Korábbi megfigyelések alapján három párt népszerűsége rendre: $\frac{2+\theta}{4}$, $\frac{1-\theta}{2}$, $\frac{\theta}{4}$, ahol $0 < \theta < 1$ paraméter. 100 embert megkérdeztek, és közülük rendre 50,30,20 támogatta az egyes pártokat.
   (a) Adjon maximum likelihood becslést θ-ra a minta alapján!

(b) Határozza meg egy 100 elemű (fenti eloszlásból vett) minta Fisher-információját!

Megoldások

1. Legyen $X_1, \ldots, X_n$ fae. minta a $(0, \theta)$-n egyenletes eloszlásból.

   (a) $\mathbb{P}_\theta(X_n^* - X_1^* < \frac{\theta}{2}) = ?$
   Mo: $\mathbb{P}_\theta(X_n^* - X_1^* < \frac{\theta}{2}) = \mathbb{P}(Y_n^* - Y_1^* < \frac{1}{2})$, ahol Y_1^* és Y_n^* egy (0,1)-en egyenletes minta legkisebb és legnagyobb elemei.
   Y_1^* és Y_n^* együttes sűrűsége:
$$f(x, y) = n(n-1)(y-x)^{n-2}, \qquad 0 \le x \le y \le 1.$$
   Innen
$$\begin{aligned}\mathbb{P}(Y_n^* - Y_1^* < \frac{1}{2}) = 1 - \mathbb{P}(Y_n^* - Y_1^* \ge \frac{1}{2}) = \\ = 1 - n(n-1)\int_{\frac{1}{2}}^{1}\int_0^{y-\frac{1}{2}} (y-x)^{n-2}\,dx\,dy = \\ = 1 - n(n-1)\int_{\frac{1}{2}}^{1}\left[\frac{(y-x)^{n-1}}{n-1}(-1)\right]_{x=0}^{y-\frac{1}{2}} dy = \\ = \left(\frac{1}{2}\right)^{n-1} \cdot \frac{n+1}{2}.\end{aligned}$$

   (b) Adjon X_1^* felhasználásával torzítatlan becslést θ-ra!
   Mo. Tanultuk, hogy $\mathbb{E}(X_1^*) = \frac{1}{n+1}\theta$, így $S = (n+1)X_1^*$ torzítatlan θ-ra.

   (c) Adjon elégséges statisztikát θ-ra!
   Mo. Gyakorlaton láttuk, hogy X_n^* elégséges θ-ra.

   (d) Blackwellizálja a b.-beli statisztikát a c.-belivel!
   Mo.
$$\mathbb{E}((n+1)X_1^* | X_n^* = t) = \int_0^t (n+1)x \frac{n(n-1)(t-x)^{n-2}\frac{1}{\theta^n}}{\frac{nt^{n-1}}{\theta^n}}\,dx = \frac{n+1}{n}t.$$
   Innen
$$\mathbb{E}((n+1)X_1^* | X_n^*) = \frac{n+1}{n}\mathbf{X}_n^*,$$
   ui. láttuk, hogy X_n^* peremsűrűsége a t helyen:
$$\mathbb{P}(X_n^* < t) = \prod_i \mathbb{P}(X_i < t) = \left(\frac{t}{\theta}\right)^n$$
   deriváltja, azaz $\frac{nt^{n-1}}{\theta^n}$.

2. Legyen $X_1, \ldots, X_n$ fae. minta az
$$f_\theta(x) = \frac{1}{\theta} \cdot x^{\frac{1-\theta}{\theta}}, \quad \text{ha} \quad 0 < x < 1$$
   (és 0 különben) sűrűségfüggvénnyel megadott eloszlásból, ahol $\theta > 0$ paraméter.

(a) Keressen elégséges statisztikát θ-ra a minta alapján! Mo.

$$L_\theta(\mathbf{x}) = \frac{1}{\theta^n}(\prod x_i)^{\frac{1}{\theta}-1} = \frac{1}{\theta^n}(\prod x_i)^{\frac{1}{\theta}}(\prod x_i)^{-1},$$

ahonnan $\prod X_i$ elégséges és $\sum \ln X_i$ is az.

(b) Adjon maximum likelihood becslést θ-ra a minta alapján! Mo. mivel

$$\ln L_\theta(\mathbf{x}) = -n\ln\theta + \frac{1}{\theta}\sum \ln x_i + h(\mathbf{x}),$$

a likelihood egyenlet:

$$\frac{\partial \ln L_\theta(\mathbf{x})}{\partial\theta} = -\frac{n}{\theta} - \frac{1}{\theta^2}\sum \ln x_i = 0,$$

ahonnan

$$\hat\theta = -\frac{\sum \ln X_i}{n} > 0$$

(ami globális max. is).

(c) Torzítatlan becslést ad-e a talált maximum likelihood becslés θ-ra? Mo. igen, mert

$$\mathbb{E}_\theta\hat\theta = -\frac{\sum \mathbb{E}_\theta \ln X_i}{n} = -\mathbb{E}_\theta \ln X_1 = -\frac{1}{\theta}\int_0^1 \ln x \cdot x^{\frac{1}{\theta}-1}dx = \theta$$

parciális integrálással.

(d) Tagja-e a fenti eloszlás az exponenciális eloszlácsaládnak? Miért igen, ill. miért nem? Mo. igen, mert

$$f_\theta(x) = \frac{1}{\theta}\cdot e^{\frac{1}{\theta}\cdot\ln x}\cdot x^{-1}.$$

Reg. feltételek teljesülnek, ezért

$$I_1(\theta) = -\mathbb{E}_\theta\left(\frac{\partial^2}{\partial\theta^2}\ln f_\theta(X_1)\right) = -\mathbb{E}_\theta\left(\frac{\partial}{\partial\theta}(-\frac{1}{\theta} - \frac{1}{\theta^2}\ln X_1)\right) =$$
$$= -\mathbb{E}_\theta\left(\frac{1}{\theta^2} + 2\frac{1}{\theta^3}\ln X_1\right) = -\left(\frac{1}{\theta^2} + 2\frac{1}{\theta^3}(-\theta)\right) = \frac{1}{\theta^2}$$

ahol használtam az előző részek eredményeit. Így $I_n(\theta) = \frac{n}{\theta^2}$.

3. Legyen $X_1, \dots, X_n$ fae. minta az

$$f_\theta(x) = \frac{3x^2}{2\theta^3}, \qquad \text{ha} \qquad -\theta \le x \le \theta$$

(és 0 különben) sűrűségfüggvénnyel megadott eloszlásból, ahol $\theta > 0$ paraméter.

(a) Keressen elégséges statisztikát θ-ra a minta alapján! Mo.

$$L_\theta(\mathbf{x}) = (\frac{3}{2})^n \cdot \frac{1}{\theta^{3n}} \cdot I(|x_i| \le \theta, \quad i = 1, \dots, n) \cdot (\prod x_i)^{2n},$$

ahonnan $\max_i |X_i|$ elégséges.

(b) Adjon maximum likelihood becslést θ-ra a minta alapján! Mo. a $\max_i |X_i| \leq \theta$ feltétel mellett, minél kisebb θ, annál nagyobb $\frac{1}{\theta^{3n}}$. Így $\hat{\theta} = \max_i |X_i|$ ML-becslés.

(c) Tagja-e a fenti eloszlás az exponenciális eloszlácsaládnak? Miért igen, ill. miért nem? Mo. nem, mert a tartó függ a paramétertől.

4. Korábbi megfigyelések alapján három párt népszerűsége rendre: $\frac{2+\theta}{4}$, $\frac{1-\theta}{2}$, $\frac{\theta}{4}$, ahol $0 < \theta < 1$ paraméter. 100 embert megkérdeztek, és közülük rendre 50,30,20 támogatta az egyes pártokat.

(a) Adjon maximum likelihood becslést θ-ra a minta alapján! Mo. polinomiális eo. szerint

$$L_\theta(\mathbf{x}) = c \cdot (\frac{2+\theta}{4})^{50} \cdot (\frac{1-\theta}{2})^{30} \cdot (\frac{\theta}{4})^{20},$$

melynek term. alapú logaritmusát deriválva és 0-val egyenlővé téve $\hat{\theta} = \frac{1}{2}$ adódik, ami tényleg max. hely.

(b) Határozza meg egy 100 elemű (fenti eloszlásból vett) minta Fisher-információját! Mo. lehetne második deriváltat venni vagy

$$I_1(\theta) = \sum_x \frac{\left(\frac{\partial}{\partial\theta} p_\theta(x)\right)^2}{p_\theta(x)} = \frac{(\frac{1}{4})^2}{\frac{2+\theta}{4}} + \frac{(-\frac{1}{2})^2}{\frac{1-\theta}{2}} + \frac{(\frac{1}{4})^2}{\frac{\theta}{4}} = \frac{2\theta+1}{2\theta(1-\theta)(2+\theta)},$$

ahonnan

$$I_{100}(\theta) = 100\frac{2\theta+1}{2\theta(1-\theta)(2+\theta)} = 50\frac{2\theta+1}{\theta(1-\theta)(2+\theta)}.$$