Matematikai statisztika
Témakör 4: Becslési módszerek

Elek Péter

1. Momentumok módszere

Eloszlások momentumai

- $m_k = E\left(X^k\right)$ az eloszlás k-dik *momentuma*
  - Várható érték: $E\left(X\right)$
    - első momentum
- $E\left[(X - E(X))^k\right]$ az eloszlás k-dik *centrális momentuma*
  - Variancia: $\sigma^2 = E\left[(X - E(X))^2\right]$
    - második centrális momentum
  - Ferdeség (skewness) = $E\left[(X - E(X))^3\right]/\sigma^3$
    - standardizált harmadik centrális momentum
  - Csúcsosság (kurtosis) = $E\left[(X - E(X))^4\right]/\sigma^4$
    - standardizált negyedik centráis momentum

Tapasztalati momentumok

- $\widehat{m}_k = \frac{\sum_{i=1}^n X_i^k}{n}$ a k-dik *tapasztalati momentum*
- Ez torzítatlanul és konzisztensen becsüli m_k-t.

Momentumok módszere (Method of moments, MM)

- Legyenek a becsülendő paraméterek $(\theta_1, \ldots, \theta_k)$.
- Tegyük fel, hogy ezek az eloszlási momentumok függvényei:

$$\theta_j = g_j\left(m_1, m_2, \ldots, m_k\right) \ (j = 1, \ldots, k).$$

- *Momentumok módszere*: a becslőfüggvényt a tapasztalati momentumok ugyanezen függvényeiként definiáljuk:

$$\widehat{\theta}_j = g_j\left(\widehat{m}_1, \widehat{m}_2, \ldots, \widehat{m}_k\right) \;\; (j = 1, \ldots, k)$$

- Általánosabban az *analógia elve*: az eloszlás elméleti jellemzőit a tapasztalati jellemzőknek megfeleltetve konstruáljuk meg a becslőfüggvényt
  - Pl.: a sokasági medián becslése a tapasztalati mediánnal

MM-becslőfüggvények konstruálása a gyakorlatban

- 1. lépés: néhány eloszlási momentumot előállítunk a paraméterek függvényeként:

$$\begin{aligned} m_1 &= h_1(\theta_1, \ldots, \theta_k) \\ m_2 &= h_2(\theta_1, \ldots, \theta_k) \\ m_k &= h_k(\theta_1, \ldots, \theta_k). \end{aligned}$$

- 2. lépés: a tapasztalati momentumokat egyenlővé tesszük ezekkel a függvényekkel, és a kapott egyenletrendszert megoldjuk:

$$\begin{aligned} \widehat{m}_1 &= h_1(\widehat{\theta}_1, \ldots, \widehat{\theta}_k) \\ \widehat{m}_2 &= h_2(\widehat{\theta}_1, \ldots, \widehat{\theta}_k) \\ \widehat{m}_k &= h_k(\widehat{\theta}_1, \ldots, \widehat{\theta}_k). \end{aligned}$$

Várható érték és variancia MM-becslőfüggvénye

- $\overline{X}$ a μ MM-becslőfüggvénye.
  - Definíció szerint igaz.
- s^2 a sokasági variancia MM-becslőfüggvénye.
  - "Majdnem" definíció szerint igaz.
  - Formálisan: tudjuk, hogy $m_1 = \mu$ and $m_2 = \mu^2 + \sigma^2$. Ezért

$$\widehat{\sigma^2}_{MM} = \widehat{m}_2 - \left(\overline{X}\right)^2 = \frac{\sum_{i=1}^n X_i^2}{n} - \left(\frac{\sum_{i=1}^n X_i}{n}\right)^2 = \frac{\sum_{i=1}^n \left(X_i - \overline{X}\right)^2}{n}.$$

- Példák:
  - Ha $X \sim N(\mu, \sigma^2)$, akkor $\widehat{\mu}_{MM} = \overline{X}$ és $\widehat{\sigma^2}_{MM} = s^2$.
  - Ha $X \sim Bernoulli(p)$, akkor $\widehat{p}_{MM} = \overline{X}$.

Kovariancia MM-becslőfüggvénye

- (Elméleti) kovariancia:

$$\sigma_{XY} = E\left[(X - E(X))\,(Y - E(Y))\right]$$

- Ennek MM-becslőfüggvénye:

$$\widehat{\sigma}_{XY} = \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{n}$$

  - Torzított és konzisztens.

- Korrigált tapasztalati kovariancia:

$$\widehat{\sigma}_{XY} = \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{n-1}$$

  - Torzítatlan és konzisztens.

Korrelációs együttható MM-becslőfüggvénye

- (Elméleti) korrelációs együttható:

$$\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y} = \frac{E\left[(X - E(X))\,(Y - E(Y))\right]}{\sqrt{E\left[(X - E(X))^2\right]}\sqrt{E\left[(Y - E(Y))^2\right]}}$$

- MM-becslőfüggvény:

$$\hat{\rho}_{XY} = \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{\sqrt{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2}\sqrt{\sum_{i=1}^{n}\left(Y_i - \overline{Y}\right)^2}}$$

- ami torzított és konzisztens.
- Nem létezik általános torzítatlan becslőfüggvény ρ_{XY}-ra.

Példa: θ MM-becslőfüggvénye exponenciális mintában

- Legyen $X \sim Exp(\theta)$, és $X_1, X_2, \ldots, X_n$ f.a.e. minta ebből az eloszlásból. Adjuk meg θ MM-becslőfüggvényét!
- Megoldás 1: ekkor $E(X) = \theta$ (integrálással belátható), ezért $\widehat{\theta}_{MM} = \overline{X}$.
- Megoldás 2: azt is tudjuk, hogy $\sqrt{Var(X)} = \theta$, tehát $\widehat{\theta}_{MM} = s$ is megfelelő (a második momentum alapján).
- Általában az alacsonyabb momentumok alapján számított MM-becslőfüggvényeket preferáljuk.

MM-becslőfüggvények általános tulajdonságai

- Az MM-becslőfüggvények konzisztensek.
  - Ha $g(\cdot)$ folytonos függvény, akkor $\widehat{\theta} = g(\widehat{m}_1, \ldots, \widehat{m}_k)$ konzisztensen becsüli a $\theta = g(m_1, \ldots, m_k)$ paramétert.
- Az MM-becslőfüggvények nem feltétlenül torzítatlanok.
  - Speciális esetekben igen: ha $g(\cdot)$ lineáris függvény, akkor például $g\left(\overline{X}\right)$ torzítatlanul becsüli a $g(\mu)$ paramétert.
- Az MM-becslőfüggvények csak az eloszlás néhány momentumában rejlő információt hasznosítják, ezért nem mindig hatásosak (aszimptotikusan).
- Ugyanakkor sokszor könnyű őket kiszámítani és robusztusak a modell félrespecifkálására.

2. Maximum Likelihood (ML) becslés

Likelihood függvény

- Legyen a tényleges sokasági sűrűségfüggvény (vagy diszkrét esetben az eloszlás) $f(x;\theta)$, ahol θ az ismeretlen paraméter.
- Egy konkrét $x_1, x_2, \ldots, x_n$ minta bekövetkezésének együttes "valószínűsége": $f(x_1;\theta) \cdot f(x_2;\theta) \cdot \ldots \cdot f(x_n;\theta)$.
- Ez a *likelihood függvény*:

$$\mathcal{L}(x_1, x_2, \ldots, x_n; \theta) = \prod_{i=1}^{n} f(x_i;\theta).$$

- Ezt most θ függvényeként vizsgáljuk (nem pedig az x_i mintaelemek függvényeként).

Maximum likelihood (ML) becslőfüggvény

- A θ paraméter *ML-becslőfüggvény*e a likelihood-függvény maximumhelye:

$$\widehat{\theta}_{ML} = \arg\max_{\theta} \mathcal{L}(X_1, X_2, \ldots, X_n; \theta).$$

- Intuíció: azt a θ-t választjuk, amelyre a konkrét minta bekövetkezési "valószínűsége" maximális.

A ML-becslés meghatározása

- F.a.e. minta *log-likelihood* függvénye:

$$\log \mathcal{L}\left(x_1, x_2, \ldots, x_n; \theta\right) = \sum_{i=1}^{n} \log f\left(x_i; \theta\right).$$

- Sok esetben ez *analitikusan* maximalizálható a deriváltja zérushelyének megkeresésével és annak bizonyításával, hogy tényleg maximumhelyhez jutottunk:

$$\partial \log \mathcal{L}\left(x_1, x_2, \ldots, x_n; \theta\right) / \partial \theta = \sum_{i=1}^{n} \partial \log f\left(x_i; \theta\right) / \partial \theta = 0.$$

- Más esetekben *numerikus* maximalizálás szükséges (pl. Newton-Raphson módszer).

Példa: a p sokasági arány ML-becslése

- Legyen $X \sim Bernoulli(p)$ és $X_1, X_2, \ldots, X_n$ f.a.e. minta ebből az eloszlásból.
- Bizonyítsuk be, hogy p ML-becslőfüggvénye $\widehat{p}_{ML} = \overline{X}$.

Példa: normális eloszlás paramétereinek ML-becslése

- Legyen $X \sim N\left(\mu, \sigma^2\right)$ és $X_1, X_2, \ldots, X_n$ f.a.e. minta ebből az eloszlásból. Határozzuk meg μ and σ^2 ML-becslését!

Megoldás

- $f\left(x_i; \mu, \sigma^2\right) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}$, tehát

$$\begin{aligned} \mathcal{L}\left(x_1, x_2, \ldots, x_n, \mu, \sigma^2\right) &= f\left(x_1\right) \cdot f\left(x_2\right) \cdot \ldots \cdot f\left(x_n\right) \\ &= \frac{1}{\left(\sqrt{2\pi\sigma^2}\right)^n} e^{-\sum_{i=1}^{n} \frac{(x_i-\mu)^2}{2\sigma^2}}, \end{aligned}$$

- és ezt kell maximalizálni μ és σ^2 szerint.
- Ez ekvivalens a log-likelihood maximalizálásával:

$$\begin{aligned} \log \mathcal{L}\left(x_1, x_2, \ldots, x_n, \mu, \sigma^2\right) &= \sum_{i=1}^{n} \log f\left(x_i\right) \\ &= -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log\sigma^2 - \sum_{i=1}^{n} \frac{\left(x_i - \mu\right)^2}{2\sigma^2}. \end{aligned}$$

Megoldás (folyt.)

- Maximalizálva μ and σ^2 szerint:

$$\begin{aligned}\frac{\partial \log \mathcal{L}}{\partial \mu} &= \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu)=0\\ \frac{\partial \log \mathcal{L}}{\partial (\sigma^2)} &= -\frac{n}{2\sigma^2}+\frac{1}{2\sigma^4}\sum_{i=1}^{n}(x_i-\mu)^2=0,\end{aligned}$$

- amiből kapjuk, hogy $\widehat{\mu}_{ML}=\overline{X}$ és $\widehat{\sigma^2}_{ML}=s^2$.
- Belátható, hogy ez tényleg maximumhely.

ML-becslőfüggvények konzisztenciája

- A ML-becslőfüggvények *nem feltétlenül torzítatlanok*.
  - Példa: $\widehat{\sigma^2}_{ML}=s^2$ a normális eloszlás esetén
- A ML-becslőfüggvények általános "regularitási" feltételek mellett *konzisztensek*.
  - Vannak ellenpéldák (pl. bonyolultabb adatstruktúrában: dummy változós regresszió nemlineáris panelmodellekben)
  - (A fentiek akkor érvényesek, ha az eloszlás jól specifikált. Ha nem, akkor is konzisztens lehet a ML-becslőfüggvény, ld. QML (quasi maximum likelihood) elmélet.)

ML-becslőfüggvények aszimptotikus normalitása

- A ML-becslőfüggvények általános "regularitási" feltételek mellett *aszimptotikusan normális eloszlásúak* a következő varianciával:

$$\begin{aligned}\widehat{\theta}_{ML} &\overset{A}{\sim} N\left(\theta,-\left[E\frac{\partial^2 \log \mathcal{L}}{\partial \theta^2}\right]^{-1}\right)\\ &= N\left(\theta,-\left[nE\frac{\partial^2 \log f(X;\theta)}{\partial \theta^2}\right]^{-1}\right).\end{aligned}$$

- Tehát a becslőfüggvény ilyenkor $\sqrt{n}$-konzisztens.
- (Ellenpélda a feltételek nem teljesülése esetén: amikor az eloszlás tartója függ θ-tól, pl. egyenletes eloszlás $[0,\theta]$-n)

Cramer-Rao-egyenlőtlenség

- *Cramer-Rao tétel*: Ha $\widehat{\theta}$ egy *torzítatlan* becslőfüggvény, akkor általános "regularitási" feltételek esetén

$$Var\left(\widehat{\theta}\right) \geq -\left[E\frac{\partial^2 \log \mathcal{L}}{\partial \theta^2}\right]^{-1}.$$

- (A jobb oldalon levő mennyiség a Cramer-Rao határ.)
- Tehát az alsó határt elérő torzítatlan becslőfüggvény szükségképpen hatásos.

ML-becslőfüggvények aszimptotikus hatásossága

- Egy becslőfüggvény *aszimptotikusan hatásos*, ha aszimptotikus eloszlása a következő:

$$\widehat{\theta} \overset{A}{\sim} N\left(\theta, -\left[E\frac{\partial^2 \log \mathcal{L}}{\partial \theta^2}\right]^{-1}\right).$$

  - azaz ha konzisztens, aszimptotikusan normális és aszimptotikusan eléri a Cramer-Rao határt.
  - (Más tankönyvek kissé különböző definíciót használnak.)

- Az előző tétel alapján a ML-becslőfüggvény általános regularitási feltételek mellett aszimptotikusan hatásos.
- Gyakran több is igaz: gyakran a ML-becslőfüggvény hatásos, mert torzítatlan és véges mintában is eléri a Cramer-Rao határt.

Példa: $\overline{X}$ hatásos becslése μ-nek normális minta esetén

- Ha $X \sim N(\mu, \sigma^2)$, akkor $\overline{X}$ hatásosan becsüli μ-t.
- Megjegyzés: azt már eddig is tudtuk általánosan, hogy $\overline{X}$ a μ legjobb *lineáris* torzítatlan becslőfüggvénye.
- (A hatásosság több más eloszlás esetén is igaz, például ha $X \sim Bernoulli(p)$, akkor $\overline{X}$ szintén hatásosan becsüli p-t.)

Megoldás

- Az egyszerűség kedvéért tegyük fel, hogy σ^2 ismert (bár az állítás e nélkül is igaz).

- A log-likelihood második deriváltja:

$$\frac{\partial^2 \log \mathcal{L}}{\partial \mu^2} = -\frac{n}{\sigma^2},$$

- tehát a Cramer-Rao határ:

$$-\left[E\frac{\partial^2 \log \mathcal{L}}{\partial \mu^2}\right]^{-1} = \frac{\sigma^2}{n}.$$

- Mivel $\overline{X}$ torzítatlan és eléri a Cramer-Rao határt, a Cramer-Rao-tétel miatt hatásos is.

3. Legkisebb négyzetek

Legkisebb négyzetek módszere (Least squares, LS)

- *Legkisebb négyzetek módszere*: $\widehat{\theta}$-t úgy választjuk, hogy a mintaelemek és azok várható értékei különbségének négyzetösszege minimális legyen.

- Példa: legyen $X_1, X_2, \ldots, X_n$ egy n elemű f.a.e. minta. Az $E(X) = \mu$ legkisebb négyzetek módszerén alapuló becslése a $\sum_{i=1}^{n} (X_i - \widehat{\mu})^2$ kifejezést minimalizálja, a minimumhely $\widehat{\mu}_{LS} = \overline{X}$. (Korábban szerepelt.)
  - Normális eloszlás esetén $\widehat{\mu}_{LS} = \overline{X}$.
  - Bernoulli-eloszlás esetén $\widehat{p}_{LS} = \overline{X}$.
  - Exponenciális eloszlás esetén $\widehat{\theta}_{LS} = \overline{X}$.

Tananyag

Tananyag

- slide-ok, feladatok és házi feladatok
- Wooldridge Appendix C.4
- Amemiya 7.1.1, 7.3, 7.4 (bizonyítások nélkül)