Matematikai statisztika

4. előadás

2022. október 3.

Becsléselmélet

- A minta eloszlásának ismeretlen paraméterét közelítjük a minta függvényével
- Def.: becslőfüggvény: $\hat{\vartheta}: \mathcal{X} \to \Theta$
- Def.: becslés: $\hat{\vartheta}(\boldsymbol{\xi})$

- A becslések maguk is statisztikák. Szubjektíven: olyan statisztikák, amik jól közelítik az ismeretlen paramétert.

Likelihood függvény

Def.: A $\xi_1,\ldots,\xi_n$ független, azonos eloszlású minta likelihood függvénye

$$L(\mathbf{x},\theta)=\begin{cases} P_\theta(\boldsymbol{\xi}=\mathbf{x})=\prod_{i=1}^{n} P_\theta(\xi_i=x_i) & \text{diszkrét minta esetén} \\ f_\theta(\mathbf{x})=\prod_{i=1}^{n} f_\theta(x_i) & \text{abszolút folytonos minta esetén} \end{cases}$$

ahol f_θ ξ_i sűrűségfüggvénye.

$l(\mathbf{x},\theta)=\ln L(\mathbf{x},\theta)$ a loglikelihood függvény.

Maximum likelihood becslés

- Definíció heurisztikusan: azt a paraméterértéket keressük, amelyre az adott minta bekövetkezési valószínűsége maximális.

Def.: θ maximum likelihood becslése $\hat{\theta} = T(\boldsymbol{\xi}) \in \Theta$, ha

$$L(\boldsymbol{\xi}, \hat{\theta}) = \max_{\theta \in \Theta} L(\boldsymbol{\xi}, \theta)$$

Likelihood egyenlet

Gyakran a loglikelihood függvény maximumhelyét keresik a

$$\frac{\partial l(\mathbf{x},\theta)}{\partial\theta}=0$$ egyenletet (vagy egyenletrendszert) megoldva.

Ez diszkrét minta esetén a

$$\sum_{i=1}^{n}\frac{\partial \ln P_\theta(\xi_i=x_i)}{\partial\theta}=0$$

egyenletet (vagy egyenletrendszert) jelenti.

Abszolút folytonos minta esetén

$$\sum_{i=1}^{n}\frac{\partial \ln f_\theta(x_i)}{\partial\theta}=0$$

egyenletet (vagy egyenletrendszert) oldjuk meg.

Példa (indikátor)

$$L(\mathbf{x}, p) = p^{\sum_{i=1}^{n} x_i} (1-p)^{n-\sum_{i=1}^{n} x_i},$$

$$l(\mathbf{x}, p) = \ln L(\mathbf{x}, p) = \left(\sum_{i=1}^{n} x_i \right) \ln p + \left(n - \sum_{i=1}^{n} x_i \right) \ln(1-p).$$

Likelihood egyenlet

$$\frac{\partial l(\mathbf{x}, p)}{\partial p} = \left(\sum_{i=1}^{n} x_i \right) \frac{1}{p} - \left(n - \sum_{i=1}^{n} x_i \right) \frac{1}{1-p} = 0$$

Ennek megoldása

$$p = \frac{\sum_{i=1}^{n} x_i}{n}.$$

És ez valóban maximumhely!

Így a ML becslés

$$\hat{p} = \frac{\sum_{i=1}^{n} \xi_i}{n}.$$

Példa. (Poisson)

Tegyük fel, hogy $\eta_1, \eta_2, \ldots, \eta_n \sim Poisson(\lambda)$. Ekkor

$$\mathrm{L}(\underline{k}, \lambda) = \mathrm{P}_\lambda(\eta_1 = k_1, \ldots, \eta_n = k_n) = \prod_{i=1}^{n} \frac{\lambda^{k_i} e^{-\lambda}}{k_i!} = \left(\prod_{i=1}^{n} \frac{1}{k_i!}\right) \lambda^{\sum k_i} e^{-n\lambda}$$

$$\ell(\underline{k}, \lambda) = \ln \mathrm{L}(\underline{k}, \lambda) = \left(\sum_{i=1}^{n} \ln\left(\frac{1}{k_i!}\right)\right) + \left(\sum_{i=1}^{n} k_i\right) \ln \lambda - n\lambda$$

$$\frac{\partial \ell}{\partial \lambda} = \frac{\sum k_i}{\lambda} - n = 0 \iff \lambda = \frac{\sum k_i}{n}$$

így az ML becslés

$$\widehat{\lambda} = \frac{\sum \eta_i}{n}$$

Becslések tulajdonságai

- Def.: *Torzítatlanság:* A paraméter $\hat{\vartheta}(\boldsymbol{\xi})$ becslése torzítatlan, ha
$$E_{\vartheta}\left(\hat{\vartheta}(\boldsymbol{\xi})\right) = \vartheta, \forall\vartheta \in \Theta.$$
- *Konzisztencia*: $\hat{\vartheta}(\boldsymbol{\xi}) \to \vartheta$ sztochasztikusan $(n\to\infty)$ minden paraméterértékre.
- Példák:
  - Valószínűség becslése relatív gyakorisággal.
  - Glivenko tétele: a tapasztalati eloszlásfüggvény egyenletesen is konvergál az elméleti eloszlásfüggvényhez.
  - Várható érték becslése mintaátlaggal

Konzisztencia

- Elégséges feltétel $E_\vartheta\left(\hat{\vartheta}_n(\boldsymbol{\xi})\right) \to \vartheta$ (aszimptotikus torzítatlanság) és ${D_\vartheta}^2\left(\hat{\vartheta}_n(\boldsymbol{\xi})\right) \to 0$.

Példák

- Poisson eloszlás paraméterére: mintaátlag
- Exponenciális eloszlás paraméterére:
  - mintaátlag reciproka: aszimptotikusan torzítatlan, konzisztens
  - $n \cdot \min(X_1, \dots, X_n)$ torzítatlan, de nem konzisztens
- Szórásnégyzetre

Becslések összehasonlítása

- Melyik a jobb becslés?

$$X_i = \begin{cases} 1, & i.\,\text{fiú} \\ 0, & i.\,\text{lány} \end{cases}, P_p(X_i = 1) = p,$$

$$\hat{p}_1 = \frac{\sum_{i=1}^{n} X_i}{n},$$
$$\hat{p}_2 = X_1, \text{vagy}$$

$$\hat{p}_3 = \frac{\sum_{i=1}^{[n/2]} X_i}{[n/2]}?$$

Becslések összehasonlítása (hatásos becslések)

- Torzítatlan becslésekre: T_1 hatásosabb becslése $h(\Theta)$-nak a T_2-nél, ha
$$D_\theta{}^2\left(T_1(\underline{X})\right) \le D_\theta{}^2\left(T_2(\underline{X})\right)$$
- teljesül minden Θ paraméterértékre.

Példa: a mintaátlag hatásosabb becslés a várható értékre minden
$$\sum_{i=1}^{n} c_i X_i$$
alakú becslésnél.

Hatásos becslés

- Def.: A T torzítatlan becslés hatásos, ha minden más torzítatlan becslésnél hatásosabb.
- Miért a torzítatlanokra? Furcsa példa: azonosan 0-val becsüljük az ismeretlen paramétert.
- Ezért érdemes a hatásos becsléseket csak a torzítatlan becslések között keresni.
- Átlagos négyzetes eltérés:

$$E_\theta\left(T(\underline{X}) - \theta\right)^2$$

Hatásos becslés egyértelműsége

Áll.: Amennyiben T_1 és T_2 hatásos becslései $h(\theta)$-nak, akkor 1 valószínűséggel megegyeznek minden lehetséges paraméter esetén.

$E_\theta T_1 = E_\theta T_2 = h(\theta)$, továbbá $D_\theta T_1 = D_\theta T_2$. Ebből

$$\mathrm{D}_\theta^2(T_1) \le \mathrm{D}_\theta^2\left(\frac{T_1+T_2}{2}\right) = \frac{\mathrm{D}_\theta^2(T_1) + 2\operatorname{cov}(T_1,T_2) + \mathrm{D}_\theta^2(T_2)}{4} = \frac{\mathrm{D}_\theta^2(T_1) + \operatorname{cov}(T_1,T_2)}{2} \Rightarrow$$

$$\mathrm{D}_\theta^2(T_1) \le \operatorname{cov}(T_1,T_2) = D_\theta T_1 \cdot D_\theta T_2 \cdot R(T_1,T_2) = \mathrm{D}_\theta^2(T_1) \cdot R(T_1,T_2) \le \mathrm{D}_\theta^2(T_1) \Rightarrow$$

$$\mathrm{D}_\theta^2(T_1) = \mathrm{D}_\theta^2(T_2) = \operatorname{cov}(T_1,T_2) \Rightarrow \mathrm{D}_\theta^2(T_1 - T_2) = \mathrm{D}_\theta^2(T_1) - 2\operatorname{cov}(T_1,T_2) + \mathrm{D}_\theta^2(T_2) = 0.$$

Így $E_\theta(T_1 = T_2) = 1 \ \forall \theta \in \Theta$.

Mit kell tudni a mintáról?

- Benzinkutas példa. Megfigyelések: 78, 89, 167, 90, 85.
- Svájcban 1871 és 1900 között a 2.644.757 megszületetett gyermekből 1.359.671 fiú és 1.285.086 lány volt.

Kár-szám	0	1	2	3	4	5	6	7	> 7	Össze-sen
Veze-tők száma	129524	16267	1966	211	31	5	1	1	0	148006

Mennyi információt hordoz a statisztika?

Példa: $\xi_1, \ldots, \xi_n$ független $N(m,1)$ minta. Ekkor

$$\bar{\xi} = \frac{\sum_{i=1}^{n} \xi_i}{n} \sim N(m, \frac{1}{n})$$ eloszlású (függ m-től!),

miközben

$$s^2 = \frac{\sum_{i=1}^{n} (\xi_i - \bar{\xi})^2}{n}$$ eloszlása nem függ m-től!

Elégséges statisztika

- Minden információt (ugyanannyit mint az eredeti minta) tartalmaz az ismeretlen paraméterre vonatkozóan.
- ”Elég” az ő értékét ismerni.
- Ismeretében már ”nincs bizonytalanság” a mintában (úgy értve, hogy egyértelmű a minta eloszlása, már nem függ az ismeretlen paramétertől).

Elégséges statisztika diszkrét minta esetén

Def.: A diszkrét $\boldsymbol{\xi}$ mintából képzett $T(\boldsymbol{\xi})$ statisztika elégséges $\theta-$ra, ha a $P_\theta(\boldsymbol{\xi} = \mathbf{x} | T(\boldsymbol{\xi}) = t)$ feltételes valószínűség nem függ $\theta-$tól

Feltételes várható érték

Legyenek X és Y diszkrét val. változók.

$E(X|Y)$ az a val. változó, ami az $Y = y_k$ eseményen az $E(X|Y = y_k)$ értéket veszi fel.

Tulajdonságok:

- Ha $X \geq 0$, akkor $E(X|Y) \geq 0$
- $E(E(X|Y)) = EX$ (a teljes várható érték tételének általánosítasa)
- Ha X_1, X_2 várható értéke véges, akkor $E(c_1X_1 + c_2X_2|Y) = c_1E(X_1|Y) + c_2E(X_2|Y)$
- Ha X független Y –tól, akkor $E(X|Y) = E(X)$
- Ha X és $h(Y)$ várható értéke véges, akkor $E(h(Y)X|Y) = h(Y)E(X|Y)$
- Teljes szórásnégyzet tétele:

$$D^2(X) = D^2(E(X|Y)) + E(D^2(X|Y))$$

Példa (indikátor minta)

$$X_i = \begin{cases} 1, & p \text{ valószínűséggel} \\ 0, & 1-p \text{ valószínűséggel} \end{cases} \Rightarrow P_p\left(X_i = x\right) = p^x (1-p)^{1-x}, x = 0 \text{ és} 1.$$

$$P_p\left(\mathbf{X} = \mathbf{x} \,\middle|\, \sum_{i=1}^n X_i = t\right) = P_p\left(X_1 = x_1, \ldots, X_n = x_n \,\middle|\, \sum_{i=1}^n X_i = t\right) =$$

$$\frac{P_p\left(X_1 = x_1, \ldots, X_n = x_n, \sum_{i=1}^n X_i = t\right)}{P_p\left(\sum_{i=1}^n X_i = t\right)} = \begin{cases} 0 & \sum_{i=1}^n x_i \neq t \\ \dfrac{P_p\left(X_1 = x_1, \ldots, X_n = x_n\right)}{P_p\left(\sum_{i=1}^n X_i = t\right)} & \sum_{i=1}^n x_i = t \end{cases} =$$

$$\begin{cases} 0 & \sum_{i=1}^n x_i \neq t \\ \dfrac{p^{\sum_{i=1}^n x_i} (1-p)^{n-\sum_{i=1}^n x_i}}{\binom{n}{t} p^t (1-p)^{n-t}} & \sum_{i=1}^n x_i = t \end{cases} = \begin{cases} 0 & \sum_{i=1}^n x_i \neq t \\ \dfrac{1}{\binom{n}{t}} & \sum_{i=1}^n x_i = t \end{cases}$$

Tétel (Neyman-féle faktorizációs):

A diszkrét ξ mintából képzett $T(\xi)$ statisztika pontosan akkor elégséges Θ-ra, ha $\exists\, g_\theta(t)$ és $h(\mathbf{x})$ úgy, hogy $\forall\,\theta \in \Theta$ és $\mathbf{x} \in \mathcal{X}$-ra

$P_\theta(\boldsymbol{\xi} = \mathbf{x}) = h(\mathbf{x}) g_\theta(T(\mathbf{x}))$.

Biz.:

$\Rightarrow T(\boldsymbol{\xi})$ elégséges, ekkor $P_\theta(\boldsymbol{\xi} = \mathbf{x}) = P_\theta(T(\boldsymbol{\xi}) = T(\mathbf{x})) \dfrac{P_\theta(\boldsymbol{\xi} = \mathbf{x}, T(\boldsymbol{\xi}) = T(\mathbf{x}))}{P_\theta(T(\boldsymbol{\xi}) = T(\mathbf{x}))}$

$= P_\theta(T(\boldsymbol{\xi}) = T(\mathbf{x})) P_\theta(\boldsymbol{\xi} = \mathbf{x} \,|\, T(\boldsymbol{\xi}) = T(\mathbf{x})) = g_\theta(T(\mathbf{x})) h(\mathbf{x})$.

$\Leftarrow P_\theta(\boldsymbol{\xi} = \mathbf{x} \,|\, T(\boldsymbol{\xi}) = t) = 0,$ ha $t \neq T(\mathbf{x})$. Amennyiben ez teljesül:

$$P_\theta(\boldsymbol{\xi} = \mathbf{x} \,|\, T(\boldsymbol{\xi}) = t) = \frac{P_\theta(\boldsymbol{\xi} = \mathbf{x}, T(\boldsymbol{\xi}) = t)}{P_\theta(T(\boldsymbol{\xi}) = t)} = \frac{P_\theta(\boldsymbol{\xi} = \mathbf{x}, T(\boldsymbol{\xi}) = t)}{P_\theta(T(\boldsymbol{\xi}) = t)} = \frac{P_\theta(\boldsymbol{\xi} = \mathbf{x})}{\sum\limits_{\mathbf{y}:T(\mathbf{y})=t} P_\theta(\boldsymbol{\xi} = \mathbf{y})}$$

$$= \frac{h(\mathbf{x}) g_\theta(T(\mathbf{x}))}{\sum\limits_{\mathbf{y}:T(\mathbf{y})=t} h(\mathbf{y}) g_\theta(T(\mathbf{y}))} = \frac{h(\mathbf{x}) g_\theta(t)}{\sum\limits_{\mathbf{y}:T(\mathbf{y})=t} h(\mathbf{y}) g_\theta(t)} = \frac{h(\mathbf{x})}{\sum\limits_{\mathbf{y}:T(\mathbf{y})=t} h(\mathbf{y})}.$$

Ez nem függ θ-tól!

Példa (Poisson minta)

$\eta_i -$k független λ Poissonok. Ekkor

$$P_\lambda\left(\eta_1 = k_1, \ldots, \eta_n = k_n\right) = \prod_{i=1}^{n} \frac{\lambda^{k_i} e^{-\lambda}}{k_i\,!} = \left(\prod_{i=1}^{n} \frac{1}{k_i\,!}\right) \lambda^{\sum_{i=1}^{n} k_i} e^{-n\lambda} =$$

$$= h(\mathbf{k}) g_\lambda\left(\sum_{i=1}^{n} k_i\right),$$

ahol

$$h(\mathbf{k}) = \prod_{i=1}^{n} \frac{1}{k_i\,!}, \quad g_\lambda\left(t\right) = \lambda^t e^{-n\lambda}.$$

Elégséges statisztika általában

Def.: A $\boldsymbol{\xi}$ mintából képzett $T(\boldsymbol{\xi})$ statisztika elégséges

Θ-ra, ha minden $\mathbf{x} \in \mathbf{R}^n$-re a $P_\theta(\boldsymbol{\xi} < \mathbf{x} \,|\, T(\boldsymbol{\xi}) = t) = P_\theta(\xi_1 < x_1, \ldots, \xi_n < x_n \,|\, T(\boldsymbol{\xi}) = t)$

feltételes eloszlásfüggvény nem függ θ-tól.

Probléma: A feltételes valószínűség és várható érték fogalmát nem tanultuk általánosan!

Abszolút folytonos eset

- Definíció a faktorizációval

Def.:

Az abszolút folytonos $\boldsymbol{\xi}$ mintából képzett $T(\boldsymbol{\xi})$ statisztika elégséges Θ−ra, ha $\exists\ g_\theta(t)$ és $h(\mathbf{x})$ úgy, hogy $\forall\theta \in \Theta$ és $\mathbf{x} \in \mathcal{X}$−ra a likelihood függvény felírható a következő alakban:
$L(\mathbf{x}, \theta) = h(\mathbf{x}) g_\theta(T(\mathbf{x}))$.

Példa (normális $N(m,\sigma^2)$ minta)

ξ_i – k független, $N(m,\sigma^2)$ eloszlásúak. Ekkor $\theta=(m,\sigma^2)$

$$L(\mathbf{x},(m,\sigma^2))=\prod_{i=1}^{n}\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{(x_i-m)^2}{2\sigma^2}\right)=$$

$$\left(2\pi\sigma^2\right)^{-n/2}\exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}\left(x_i^2-2x_im+m^2\right)\right)=$$

$$\left(2\pi\sigma^2\right)^{-n/2}\exp\left(-\frac{1}{2\sigma^2}\left(\sum_{i=1}^{n}x_i^2-2nm\overline{x}+nm^2\right)\right)=$$

$$\left(2\pi\sigma^2\right)^{-n/2}\exp\left(-\frac{1}{2\sigma^2}\left(\sum_{i=1}^{n}\left(x_i-\overline{x}\right)^2+n\overline{x}^2-2nm\overline{x}+nm^2\right)\right).$$

Ebből következik, hogy $\left(\sum_{i=1}^{n}x_i^2,\overline{x}\right)$ elégséges statisztika.

Hasonlóan $\left(\sum_{i=1}^{n}\left(x_i-\overline{x}\right)^2/n,\overline{x}\right)$ is.

Példa (egyenletes $E(0, a)$ minta)

$\xi_i -$ k független, $E(0, a)$eloszlásúak. Sűrűségfüggvényük

$$f_a(x) = \begin{cases} 1/a & 0 \leq x \leq a \\ 0 & \text{különben} \end{cases}$$

$$L(\mathbf{x}, a) = \prod_{i=1}^{n} \frac{1}{a} \chi\{x_i \leq a\} = \frac{1}{a^n} \chi\left\{\max_{1 \leq i \leq n} x_i \leq a\right\} \Rightarrow$$

$\max\limits_{1 \leq i \leq n} x_i$ elégséges!

Maximum likelihood becslés

- A maximum likelihood becslés mindig az elégséges statisztika függvénye
- $E(0,a)$ minta esetén a ML becslése

$$\max_{1 \le i \le n} \xi_i$$

Momentum módszer

- Ha az eloszlást k db paraméter határozza meg, akkor k db egyenletből kaphatunk rájuk becslést. Az egyenletek a tapasztalati és az elméleti momentumok egybevetéséből adódnak:

$$m_i(\underline{\theta}) = E_{\underline{\theta}}(X^i)$$

$$m_i(\underline{\theta}) = \frac{\sum_{j=1}^{n} \left(\xi_j\right)^i}{n}$$