Matematikai statisztika

Informatika alapszak, "A" spec.

Zempléni András

Valószínűségelméleti és Statisztika Tanszék
Matematikai Intézet
Természettudományi Kar
Eötvös Loránd Tudományegyetem

Honlap: `zempleni.elte.hu`

E-mail: `andras.zempleni@ttk.elte.hu`

Szoba: D 3-310

3. előadás

- **Maximum likelihood módszer** (ML-módszer): Azt a paraméterértéket keressük, ahol a likelihood függvény a legnagyobb értéket veszi fel: $\max_{\vartheta} L(\vartheta; \mathrm{x})$.
- Ez nyilván megegyezik azzal a paraméterértékkel, ahol a log-likelihood függvény veszi fel a legnagyobb értéket, azaz: $\max_{\vartheta} \ell(\vartheta; \mathrm{x})$.
- Amennyiben a függvény deriválható ϑ szerint, akkor a maximumot kereshetjük a deriváltak segítségével
- Célszerűbb a likelihood függvény helyett a log-likelihood függvény maximumhelyét keresni.
- Ha ϑ 1 dimenziós, akkor $\partial_\vartheta \ell(\vartheta, \mathrm{x}) = 0$, míg ha $\vartheta = (\vartheta_1, \ldots, \vartheta_p)$ p dimenziós, akkor $\partial_{\vartheta_i} \ell(\vartheta, \mathrm{x}) = 0$ megoldásából kapjuk a becslést.

Tétel. [ML-becslés invariáns tulajdonsága] Ha ϑ ML-becslése $\hat{\vartheta}$, akkor tetszőleges g függvény esetén $g(\vartheta)$ ML-becslése $g(\hat{\vartheta})$.

Becslési módszerek II

Momentum módszer:

- A mintából számítható tapasztalati momentumokat ($m_i := \frac{1}{n}\sum_j x_j^i$) egyenlővé tesszük az elméleti momentumokkal ($M_i(\vartheta) := E_\vartheta X^i$),
- mégpedig annyit, amennyiből a paramétereket meg tudjuk határozni. p darab ismeretlen paraméter esetén p ismeretlenes egyenletrendszert oldunk meg ϑ-ra: $M_1(\vartheta) = m_1, \ldots, M_p(\vartheta) = m_p$ (megjegyzés: $m_1 = \overline{x}$)
- Ha valamelyik egyenlet nem ad információt a keresett paraméterre, akkor magasabb hatványokat nézünk (míg megoldható nem lesz az egyenletrendszer)

Példák: Határozzuk meg a momentum és a maximum likelihood becslést a(z)

- Poisson
- exponenciális
- normális

eloszlás paramétere(i)re!

Feladatok

E1.) El szeretnénk dönteni egy érméről, hogy az szabályos-e, avagy cinkelt. Írjuk fel a problémát leíró statisztikai mezőt!

E2.) Legyen $X_1, \ldots, X_n$ i.i.d. $\mathrm{Bin}(4; p)$ eloszlású valószínűségi változó, ahol $p \in (0; 1)$ ismeretlen valós paraméter.

a.) Adjuk meg a mintateret és a paraméterteret!

b.) Határozzuk meg az ismeretlen paraméter ML-becslését!

c.) Határozzuk meg az ismeretlen paraméter momentum becslését!

d.) Torzítatlan, illetve konzisztens az ML-becslés? Amennyiben nem torzítatlan, tegyük azzá!

e.) Adjunk torzítatlan becslést $g(p) = p^2(1-p)^2$-re!

E3.) Legyen X_1 $\mathrm{Bin}(2; p)$ eloszlású (egyelemű) minta, ahol $p \in (0; 1)$ ismeretlen valós paraméter. Adjunk X_1 segítségével torzítatlan becslést $g(p) = \frac{1}{p}$-re!

Feladatok

E4.) Tegyük fel, hogy egy hét munkanapjain az alábbi várakozási időket mértük a 4-es 6-os villamosra (percben): 1,2 2 1,5 3 2,1
A várakozási időről tegyük fel, hogy exponenciális eloszlású.

a.) Adjuk meg a mintateret és a paraméterteret!
b.) Határozzuk meg az ismeretlen paraméter ML-becslését!
c.) Határozzuk meg az ismeretlen paraméter momentum-becslését!
d.) Szimulációval vizsgáljuk meg, hogy 10, 20, 50 és 100 elemű exponenciális mintából számolt ML-becslés torzítatlanul becsüli-e az ismeretlen paramétert!
e.) Torzítatlan, illetve konzisztens az ML-becslés? Amennyiben nem torzítatlan, tegyük azzá!
f.) Mutassuk meg, hogy az $S(\mathrm{X}) = n \cdot X_1^*$ statisztika torzítatlan, de nem konzisztens becslése $g(\vartheta) = \frac{1}{\lambda}$-nak!

Elégséges statisztika

- Legyen az $X_1, \ldots, X_n$ minta diszkrét eloszlású. $S(\mathrm{X})$ minden információt tartalmaz a ϑ paraméterre (azaz elégséges ϑ-ra), ha a

$$P_\vartheta(\mathrm{X} = \mathrm{x} | S(\mathrm{X}) = s)$$

feltételes valószínűség nem függ ϑ-tól.

- Faktorizáció: S pontosan akkor elégséges, ha

$$f_\vartheta(\mathrm{x}) = h(\mathrm{x}) g_\vartheta(S(\mathrm{x}))$$

E**5.)** Határozzuk meg az elégséges statisztikát a Poisson eloszlás paraméterére!

- Az abszolút folytonos esetre a faktorizáció a definíció.
E**6.)** Határozzuk meg az elégséges statisztikát az exponenciális és az egyenletes eloszlás paramétere(i)re!

Tulajdonságok

- Az elégséges statisztika nem egyértelmű: szigorúan monoton függvénybe behelyettesíthetjük
- Sőt: ha hozzávesszünk újabb statisztikákat, akkor is elégséges marad.
- Ezért a célunk minél egyszerűbb elégséges statisztika keresése.
- A maximum likelihood becslés az elégséges statisztika függvénye.
- Rao-Blackwell tétel: Ha T torzítatlan becslés a ϑ paraméterre és S elégséges statisztika ϑ-ra, akkor megadható olyan becslés, mely
  - S függvénye
  - torzítatlan ϑ-ra
  - kisebb szórású, mint T
- $E(T|S)$ lesz a megoldás.