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: `zempleni@caesar.elte.hu`

Szoba: D 3-310

4. előadás

Tulajdonságok

- Az elégséges statisztika nem egyértelmű: szigorúan monoton függvénybe behelyettesíthetjük
- Sőt: ha hozzáveszünk 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.

A Fisher-információ, becslések hatásossága

- A minta (elméleti) információja:

$$I_n(\vartheta) = E_\vartheta \left(\partial_\vartheta \ell(\vartheta, \mathbf{X})\right)^2$$

- Tulajdonságok:
  - A reguláris esetben $I_n(\vartheta) = nI_1(\vartheta)$.
  - Ha másodrendben is reguláris a függvény, akkor

$$I_n(\vartheta) = -E_\vartheta \left(\frac{\partial^2}{\vartheta^2} \ell(\vartheta, \mathbf{X})\right)$$

Definíció. **[** hatásosság] Egy T torzítatlan becslés hatásos, ha minden más T^* torzítatlan becslésre $D^2_\vartheta T(\mathbf{X}) \leq D^2_\vartheta T^*(\mathbf{X})$ minden $\vartheta \in \Theta$ paraméterre

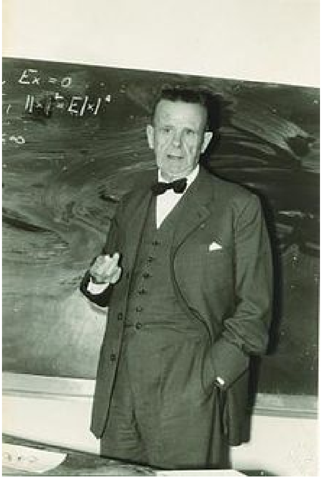
- Megjegyzés: a hatásos becslés nem mindig létezik, de ha van ilyen, akkor egyértelmű.
- Az általános (nem feltétlenül torzítatlan) esetre $E(T(\mathbf{X}) - \vartheta)^2)$ minimalitása definiálja a hatásosságot

A Fisher-információ alkalmazásai

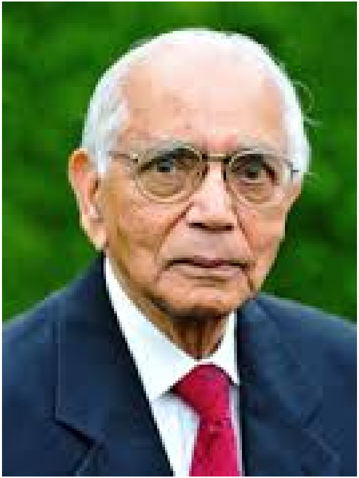
- hatásos becslés keresése az információs határ segítségével
  Cramér-Rao egyenlőtlenség: A reguláris esetben
  $$D_\vartheta^2(T(\mathbf{X})) \geq \underbrace{\frac{(g'(\vartheta))^2}{I_n(\vartheta)}}_{\text{információs határ}}$$
  Ha egy $g(\vartheta)$-ra nézve torzítatlan T statisztika esetén egyenlőség teljesül, akkor az a statisztika hatásos becslése $g(\vartheta)$-nak.
- az ML-becslés eloszlásban egy olyan normális eloszláshoz tart, aminek a szórásnégyzete a Fisher-információ inverze
- intervallumbecslés az ML-becslésre
- kísérlettervezés
- Bayes-i statisztika – Jeffrey-féle apriori eloszlás számításához
- neurális hálók, machine learning
- számítógépes agykutatás

Az információs határ felfedezői/névadói

Harald Cramér (1893 – 1985)



C. R. Rao (1920 –)



Hatásos becslés keresése az információs határral

E**5.)** Legyen $X_1, \ldots, X_n$ i.i.d. Ind(p) eloszlású.

a.) Határozzuk meg a mintában lévő Fisher-információ értékét!

b.) Mutassuk meg, hogy a relatív gyakoriság hatásos becslése a valószínűségnek!

E**6.)** (Gyakorlatra) Legyen $X_1, \ldots, X_n$ i.i.d. Exp(λ) eloszlású.

a.) Határozzuk meg a mintában lévő Fisher-információ értékét!

b.) Mutassuk meg, hogy a mintaátlag hatásos becslése $\frac{1}{\lambda}$-nak!

Az ML becslés aszimptotikus tulajdonságai

- Ha a likelihood függvény teljesít bizonyos regularitási feltételeket, akkor a maximum likelihood becslés $\hat{\vartheta}_n$
  - létezik
  - aszimptotikusan torzítatlan
  - aszimptotikusan hatásos:

$$\sqrt{nI_1(\vartheta)}D(\hat{\vartheta}_n) \to 1 \ (n \to \infty)$$

  - Az alábbi függvénye aszimptotikusan standard normális eloszlású:

$$\sqrt{nI_1(\vartheta)}(\hat{\vartheta}_n - \vartheta) \ (n \to \infty)$$

Def.: **Konfidenciaintervallum.** Olyan intervallum, mely legalább $1-\alpha$ valószínűséggel tartalmazza a paramétert minden ϑ értékre.
Legyen $X_1, \ldots, X_n \sim N(m, \sigma^2)$ i.i.d. minta, $\alpha > 0$ "kicsi" valós szám.

Kétoldali $(1-\alpha)$-megbízhatóságú konfidenciaintervallumok:

- m-re
  - ha σ ismert, akkor $\boxed{\overline{X} \pm u_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}}$