Leíró és matematikai statisztika

Matematika alapszak, matematikai elemző szakirány

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

6. előadás

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:

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

Standard hiba

Def.: **Standard hiba.** A becslés standard hibája a becslés szórása (jele: s.e., az angol standard error megnevezésből).

- Példa: $s.e.(\overline{X}) = \frac{\sigma}{\sqrt{n}}$
- Ha σ ismeretlen, akkor becsüljük: $\widehat{s.e.}(\overline{X}) = \frac{\hat{\sigma}}{\sqrt{n}}$
- Megj.: A becslés nem torzítatlan! (Csak aszimptotikusan)
- Alkalmazás: a becslési hiba behatárolható a segítségével (ld. Csebisev-egyenlőtlenség)
- Általában ennél pontosabb közelítés is adható

Intervallumbecslések, példák normális elo. mintára

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

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

- m-re
  - ha σ ismert, akkor $\boxed{\overline{X} \pm u_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}}$
  - ha σ ismeretlen, akkor $\overline{X} \pm t_{n-1,1-\frac{\alpha}{2}} \frac{S_n^*}{\sqrt{n}}$
- σ^2-re: $\left[\frac{(n-1)\cdot(S_n^*)^2}{\chi^2_{n-1,1-\frac{\alpha}{2}}} \ , \ \frac{(n-1)\cdot(S_n^*)^2}{\chi^2_{n-1,\frac{\alpha}{2}}} \right]$

Egyoldali (alsó) $(1-\alpha)$-megbízhatóságú konfidenciaintervallumok:

- m-re
  - ha σ ismert, akkor $() - \infty \ , \ \overline{X} + u_{1-\alpha} \frac{\sigma}{\sqrt{n}}$
  - ha σ ismeretlen, akkor $\left(-\infty \ , \ \overline{X} + t_{n-1,1-\alpha} \frac{S_n^*}{\sqrt{n}} \right)$
- σ^2-re: $\left(0 \ , \ \frac{(n-1)\cdot(S_n^*)^2}{\chi^2_{n-1,\alpha}} \right)$

Intervallumbecslések lefedési valószínűsége

E**29.)** Generáljunk $n = 5, 10, 20, 50, 100$ elemű mintát

- a.) $N(1, 2^2)$;
- b.) Exp(2);
- c.) $E(1; 5)$

eloszlásból 10^4 alkalommal, majd becsüljük meg a várható értékre adott $\overline{x} \pm \frac{S_n^*}{\sqrt{n}} t_{n-1,1-\alpha/2}$ intervallum lefedési valószínűségét $\alpha = 0,01$, $0,05$ és $0,1$ esetén, ahol $t_{n-1,1-\alpha/2}$ az $n-1$ szabadságfokú t-eloszlás $1-\alpha/2$ kvantilise!

E**30.)** Generáljunk $n = 5, 10, 20, 50, 100, 200, 500, 1000$ elemű mintát Exp(λ) eloszlásból 10^4 alkalommal, majd adjunk az ismeretlen λ paraméter ML-becslése alapján intervallumbecslést annak aszimptotikus eloszlása segítségével! Becsüljük meg az intervallumok lefedési valószínűségét $\alpha = 0,01$, $0,05$ és $0,1$ esetén!

Kifejtés után: $\hat{\lambda}_n \pm D(\hat{\lambda}_n) \cdot u_{\alpha/2} \approx \hat{\lambda}_n \pm \frac{\hat{\lambda}_n \cdot u_{\alpha/2}}{\sqrt{n}} = \hat{\lambda}_n \cdot \left(1 \pm \frac{u_{\alpha/2}}{\sqrt{n}}\right)$

További példa

E**31.)** Adjunk konfidencia intervallumot a $[0; b]$ intervallumon egyenletes eloszlás paraméterére n elemű minta alapján!
Megoldás:

- az intervallumot érdemes $\max(X_i)$ függvényében keresni, mert ez az elégséges staisztika b-re
- $\frac{\max(X_i)}{b}$ eloszlásfüggvénye z^n a $(0; 1)$ intervallumon (mert X/b E(0;1) eloszlású)
- Mivel a sűrűségfüggvény ($n > 1$-re) szigorúan monoton, ezért a legrövidebb, $1-\alpha$ vszgű tartomány $(\alpha^{1/n}; 1)$
- az $1-\alpha = P(\alpha^{1/n} < \frac{\max(X_i)}{b} < 1)$ összefüggés átrendezéséből $1-\alpha = P(\max(X_i) < b < \frac{\max(X_i)}{\alpha^{1/n}})$ kapjuk a

$$(\max(X_i); \frac{\max(X_i)}{\alpha^{1/n}})$$

konfidencia intervallumot

Mintaelemszám-meghatározás, még egy feladat, megjegyzés

Gyakran az a cél, hogy az m-re kapott konfidencia intervallum adott d szélességű legyen. Ehhez szükséges elemszám $N(m, \sigma^2)$ esetén:
$n \geq \frac{4\sigma^2 u_{1-\frac{\alpha}{2}}}{d^2}$

E**32.)**

a.) Konstruáljunk pontos konfidencia intervallumot az exponenciális eloszlás paraméterére n elemű minta alapján!

b.) Hasonlítsuk össze a maximum likelihood becslés aszimptotikája alapján adódó intervallummal!

- A gyakorlati alkalmazáshoz a Fisher-féle információt becsülni kell, ha nem ismert a zárt alakja: $\hat{I}_n = -\sum_{i=1}^{n} \left(\frac{\partial^2}{\vartheta^2} \ell(\hat{\vartheta}, x_i) \right)$
- Több paraméter esetére is átvihető

Öt perc szünet

- "Rick Thoman klímaszakértő, a fairbanksi Alaszkai Klímakutató Központ munkatársa szerint az északi partvidéken fekvő Barrow városban a hőmérséklet az elmúlt héten mínusz 1 Celsius-fokra emelkedett, holott ilyenkor mínusz 20 Celsius-fok a megszokott." (Forrás: Index)
- Mikor számít ez különlegesnek? Csak akkor, ha az adott időszakban a hőmérséklet szórása elég kicsi ahhoz, hogy a megfigyelt érték kívül essen pl. a 99,9%-os konfidencia intervallumon
- További apróságok: nem derül ki, hogy vajon a -1 fok napi maximum volt-e? A -20 pedig esetleg átlag?
- Ez már szinte semmiség: a cím szerint "20 fokkal van melegebb Alaszkában, mint szokott"

Bayes becslések

- Példa: egy szabályosnak tűnő érmét 10-szer feldobtunk és egy fej sem jött ki, azaz $\hat{p} = 0$. De azért nem gondoljuk, hogy $p = 0$.
- Tegyük fel, hogy van valami előzetes információnk az ismeretlen paraméterről: az 1/2-hez közeli értékeket valószínűbbnek gondoljuk, mint a 0-hoz közelieket.
- Matematikailag: egy eloszlást tételezünk fel a paramétertéren ($(0;1)$ intervallum)
- Elnevezés: apriori eloszlás (pl. béta eloszlás a vszg. esetén), jelölése: π.
- A minta prediktív (feltétel nélküli) eloszlása:

$$f_\pi(x) = \int_\Theta f_\vartheta(x)\pi(\vartheta)d\vartheta$$

- Ez diszkrét esetben $P(\mathbf{X} = \mathbf{x})$, a π indexet gyakran ki sem írjuk

Aposzteriori eloszlás

- A minta megfigyelése után más lesz az eloszlás a paramétertéren (aposzteriori eloszlás):

$$\pi^*(\vartheta) = \frac{P(\mathbf{X} = \mathbf{x}|\vartheta)\pi(\vartheta)}{P(\mathbf{X} = \mathbf{x})}$$

- Példa: ha m érmedobásnál ℓ fej jön ki, a fej valószínűségének apriori eloszlása Béta(k, n), akkor a fej valószínűségének aposzteriori eloszlása Béta$(k + \ell, n + m - \ell)$
- Ha a mintaelemszám nő, akkor egyre inkább a minta határozza meg az aposzteriori eloszlást
- Ha az apriori eloszlás szórása kicsi (prekoncepció), akkor ezt nehéz kompenzálni – csak nagy minta esetén lehetséges

Bayes-becslések: veszteségfüggvények

- Ha már tudjuk a paraméter eloszlását (π^*), akkor már csak a számítási módot kell meghatározzuk a becsléshez
- Veszteségfüggvény: $R(\hat{\vartheta}, \vartheta)$: az a veszteség, ami akkor ér bennünket, ha a ϑ paramétert $\hat{\vartheta}$-al becslünk
- Tulajdonságok:
  - nemnegatív,
  - tipikusan $\hat{\vartheta} - \vartheta$ függvénye,
  - R is valószínűségi változó.
- Példák:

$$R(\hat{\vartheta}, \vartheta) = (\hat{\vartheta} - \vartheta)^2,$$

$$R(\hat{\vartheta}, \vartheta) = |\hat{\vartheta} - \vartheta|$$

- Optimális döntés speciális veszteségfüggvényekre: a veszteségfüggvény várható értékét minimalizáljuk
  - A négyzetes veszteségfüggvényre $E(\pi^*)$
  - Az abszolút eltérésre $medin(\pi^*)$