Valószínűségszámítás és statisztika
9. gyakorlat
2015. 11. 20.

Elmélet

[Minta] $X_1, ..., X_n$ valószínűségi változó sorozat. (Jel. $\mathbf{X} = X_1, ..., X_n$)
A továbbiakban feltesszük, hogy függetlenek és azonos eloszlásúak (FAE minta, *i.i.d. minta* (independent, identically distributed).
Az elméleti értékeket nagy, a konkrét, realizált mintából számolt értékeket mindig kis betű fogja jelölni, azaz minta esetén $x_1, ..., x_n$.

[Statisztika] A minta valamely függvénye: $T : \mathbf{X} \to ...$

[Becslés] A minta eloszlásának ismeretlen paraméterét közelíti a minta segítségével. Megj.: Minden becslés statisztika.

Néhány lényeges statisztika:

- *Rendezett minta*: $X_1^* \leq ... \leq X_n^*$ nem csökkenő sorrendbe tesszük a mintaelemeket.
- *Mintaátlag*: $\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n}$
- *Tapasztalati szórás*: $S_n = \sqrt{\frac{\sum_{i=1}^{n} (X_i - \overline{X})^2}{n}}$
  Értelmezése: az átlagtól való átlagos eltérés abszolút mértékegységben.
- *Korrigált tapasztalati szórás*: $S_n^* = \sqrt{\frac{\sum_{i=1}^{n} (X_i - \overline{X})^2}{n-1}}$
- *Szórási együttható*: $V = \frac{S_n}{\overline{X}}$
  Értelmezése: az átlagtól való átlagos eltérés százalékban. Megj.: relatív szórásnak is hívják.
- *Tapasztalati eloszlásfüggvény*: $F_n(x) = \frac{\sum_{i=1}^{n} I(X_i < x)}{n}$
  ahol $I(X_i < x) = \begin{cases} 1 & \text{ha } X_i < x \\ 0 & \text{ha } X_i \geq x \end{cases} \rightsquigarrow$ indikátor függvény

[Torzítatlan becslés:] T($\mathbf{X}$) statisztika torzítatlan becslése θ-nak, ha $E_\theta T(\mathbf{X}) = \theta \quad \forall\theta$-ra.

[Tapasztalati eloszlásfüggvény konvergenciája (Glivenko-Cantelli):] Az $F_n(x) = \sum_{i=1}^{n} \frac{\mathrm{I}(X_i < x)}{n}$ tapasztalati eloszlásfüggvény 1 valószínűséggel tart a (valódi) $F(x)$ eloszlásfüggvényhez.

[Likelihood függvény:] Legyen $\mathbf{X} = (X_1, ..., X_n)$ i.i.d. minta

- $L(\theta, \mathbf{x}) = f_\theta(\mathbf{x}) = \prod_{i=1}^{n} f_\theta(x_i)$, ha az eloszlás folytonos
- $L(\theta, \mathbf{x}) = P_\theta(\mathbf{X} = \mathbf{x}) = \prod_{i=1}^{n} P_\theta(X_i = x_i)$, ha az eloszlás diszkrét.

[Log-likelihood függvény:] $l(\theta, \mathbf{x}) = \log(L(\theta, \mathbf{x}))$.

Paraméterbecslési módszerek

- *Maximum likelihood módszer (ML-módszer):* Azt a paraméterértéket keressük, ahol a likelihood függvény a legnagyobb értéket veszi fel: $\max_\theta L(\theta, \mathbf{x})$
  Amennyiben a függvény deriválható θ szerint, akkor a maximumot kereshetjük a szokásos módon, az első és második deriváltak segítségével, azonban a feladatunkat jelentősen megnehezíti, hogy olyan n-szeres szorzatot

Személyes adatok:
Név: Bondici László
E-mail: bondici@cs.elte.hu
Honlap: www.cs.elte.hu/∼bondici

kellene deriválni, amelyiknek minden tagjában ott van az a változó, ami szerint deriválnunk kellene. Ezért a likelihood függvény helyett a log-likelihood függvény maximumhelyét keressük.
Ha θ 1 dimenziós, akkor az

- elsőrendű feltétel: $\partial_\theta l(\theta, \mathbf{x}) = 0 \quad \rightsquigarrow \quad \hat{\theta}$
- másodrendű feltétel: $\partial_\theta^2 l(\hat{\theta}, \mathbf{x}) < 0$

Ha θ p dimenziós, akkor $\theta = (\theta_1, ..., \theta_p)$, az

- elsőrendű feltétel:
  $\partial_{\theta_i} l(\theta, \mathbf{x}) = 0 \quad \rightsquigarrow \quad \hat{\theta}_i \ \ (i = 1, ..., p) \quad \rightsquigarrow \quad \hat{\theta} = (\hat{\theta}_1, ..., \hat{\theta}_p)$
- másodrendű feltétel: $H(\theta_1, ..., \theta_p) = \left(\partial_{\theta_i}\partial_{\theta_j} l(\theta, \mathbf{x})\right)_{i,j=1,...,p}$ Hesse-mátrix negatív definit a $\theta = \hat{\theta}$ helyen

- *Momentum módszer (MM):* A mintából számítható tapasztalati momentumokat ($m_i := \frac{\sum_j x_j^i}{n}$) egyenlővé tesszük az elméleti momentumokkal ($M_i := E_\theta X^i$), az elsőtől kezdve, mégpedig annyit, amennyi paraméter van. Tehát p darab ismeretlen paraméter esetén a következő p ismeretlenes egyenletrendszert oldjuk meg:
  $$M_1 = m_1$$
  $$\vdots$$
  $$M_p = m_p$$
  Megjegyzés: $m_1 = \overline{x}$

Fisher-tétel: Ha θ ML-becslése $\hat{\theta}$, akkor tetszőleges g függvény esetén $g(\theta)$ ML-becslése $g(\hat{\theta})$.

Feladatok

1. Egy szabályos dobókockával 7-szer dobtunk. A következő eredmények születtek: 3,4,1,6,2,3,2. Határozzuk meg a következőket:
   a) mintatér;
   b) mintaátlag;
   c) tapasztalati szórásnégyzet és szórás;
   d) korrigált tapasztalati szórásnégyzet és szórás;
   e) második és harmadik tapasztalati momentum;
   f) szórási együttható;
   g) tapasztalati eloszlásfüggvény.
2. Határozzuk meg a tapasztalati közép és a tapasztalati szórásnégyzet várható értékét.
3. Mutassuk meg, hogy a konvex (lineáris) becslések közül az átlag a legkisebb szórásnégyzetű.
4. Legyen Z eloszlása: $P(Z = 1) = c, P(Z = 2) = 3c, P(Z = 3) = 1 - 4c$. ($c$ az ismeretlen paraméter).
   a) Határozzuk meg Z várható értékét és szórását.
   b) Adjunk torzítatlan becslést c-re Z_1 , illetve a $Z_1, Z_2, ...Z_n$ minta segítségével!
5. Határozzuk meg az ismeretlen paraméter maximum likelihood (ML) és MM becslését, ha az n elemű minta
   (a) Pascal (p az ismeretlen paraméter);
   (b) binomiális (p az ismeretlen paraméter)
   (c) $E(a, 1)$ (a az ismeretlen paraméter)
   (d) exponenciális (λ az ismeretlen paraméter)

   eloszlású.
6. Legyen a $Z_1, Z_2, \ldots, Z_n$ minta $N(2m + 5, (1/d)^2)$ eloszlású. Határozzuk meg az ismeretlen paraméterek MM becslését.

Házi feladat

1. Az első ZH eredményeit felhasználva határozzuk meg az 1. feladatban szereplő objektumokat, mennyiségeket.