Valószínűségszámítás és statisztika
9. gyakorlat
2014. 11. 20.

Elmélet

[Sűrűségfüggvény becslése magfüggvény segítségével n elemű mintából:]

Parzen-Rosenblatt becslés: $f(x) \approx f_n(x) = \frac{1}{nh_n} \sum_{i=1}^{n} k\left(\frac{x-X_i}{h_n}\right)$

[Tapasztalati eloszlásfüggvény konvergenciája:] 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 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: $\partial_\theta^2 l(\theta, \mathbf{x}) < 0$

  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:* 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 $g(\hat{\theta})$.

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

Feladatok

1. (8-as feladatsor 6-os feladata) Legyen X exponenciális eloszlású $\lambda = 1$ paraméterrel. Adjuk meg $1-e^{-X}$ sűrűségfüggvényét és várható értékét.
2. (8-as feladatsor 8-as feladata) Tegyük fel, hogy egy tábla csokoládé tömege normális eloszlású 100g várható értékkel és 3g szórással, valamint, hogy az egyes táblák tömege egymástól független. Legalább hány csokoládét csomagoljunk egy dobozba, hogy a dobozban levő táblák átlagos tömege legalább 0.9 valószínűséggel nagyobb legyen 99.5 g-nál?
3. Legyen X egyenletes eloszlású az $(a, 5)$ intervallumon, és $P(X < E(X+1)) = \frac{5}{6}$. Határozzuk meg a értékét és írjuk fel X sűrűségfüggvényét.
4. Adjunk becslést az alábbi, éves maximum vízállások alapján az eloszlás 99%-os kvantilisére
   a) a tapasztalati eloszlásból;
   b) normális közelítésből;
   c) $500 + Y$-ból, ahol Y exponenciális.

Év	Szint	Év	Szint	Év	Szint	Év	Szint	Év	Szint
2000	690	2001	709	2002	876	2003	544	2004	843
2005	586	2006	546	2007	923	2008	830	2009	873

5. Határozzuk meg az ismeretlen paraméter maximum likelihood (ML) becslését, ha a minta
   (a) Pascal (p az ismeretlen paraméter);
   (b) $E(a, 1)$ (a az ismeretlen paraméter)

   eloszlású.
6. Legyen a $Z_1, Z_2, \ldots, Z_n$ minta a következő diszkrét eloszlásból: $P(Z_i = 1) = c, P(Z_i = 2) = 3c, P(Z_i = 3) = 1 - 4c$ (c az ismeretlen paraméter). Határozzuk meg c momentum módszerrel adódó becslését.
7. Tegyük fel, hogy december 7-e középhőmérséklete Budapesten az elmúlt 10 évben az alábbiak szerint alakult: 2.0, 8.5, 1.6, -4.5, 3.3, 7.9, 0.2, -1.6, -2.2, 5.6. Számítsuk ki és ábrázoljuk a középhőmérséklet sűrűségfüggvényének Parzen-Rosenblatt becslését, ha $h = 0.25$ és a magfüggvényünk $\frac{1}{2}$, ha $|x| < 1$, és 0 különben.

Házi feladatok

1. a) A 4-es feladat adatai alapján adjunk becslést az eloszlás 99%-os kvantilisére a tapasztalati eloszlásból.
   b) Határozzuk meg az ismeretlen paraméter maximum likelihood (ML) becslését, ha a minta Binomiális (p az ismeretlen paraméter, n ismert) eloszlású.