Matematika A4 - Valószínűségszámítás 10. gyakorlat
ML-becslés, konfidenciaintervallum

Paraméterbecslés

Legyen $\mathcal{X}_n = (X_1, X_2, \ldots, X_n)$ egy n elemű minta! Ez alapján akarjuk az X valószínűségi változó θ paraméterét becsülni.

- Átlag: $\overline{\mathcal{X}}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ torzítatlan becslés $\mathbb{E}(X)$-re
- Tapasztalati szórás: $\hat{s} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{\mathcal{X}}_n)^2}$ torzított becslés $\mathbb{D}(X)$-re
- Torzítatlan becslés szórásra: $\hat{s}^* = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{\mathcal{X}}_n)^2}$ (1)

Maximum Likelihood becslés

Azt a θ értéket akarjuk meghatározni, amire a látott $\mathcal{X}_n$ realizáció a legvalószínűbb. Ehhez maximalizálni kell θ-ban az $l(\mathcal{X}_n, \theta)$ **likelihood függvény**t:

- Diszkrét eset: $l(\mathcal{X}_n, \theta) = \prod_{i=1}^{n} \mathbb{P}(X_i \,|\, \theta)$, ahol $\mathbb{P}(X_i \,|\, \theta)$ annak a valószínűsége, hogy ha az igazi paraméter θ, akkor X_i-t látunk
- Folytonos eset: $l(\mathcal{X}_n, \theta) = \prod_{i=1}^{n} f(X_i \,|\, \theta)$, ahol $f(X_i \,|\, \theta)$ az X változó sűrűségfüggvénye X_i-ben, ha az igazi paraméter θ

Ehhez elég az $L(\mathcal{X}_n, \theta) = \ln(l(\mathcal{X}_n, \theta))$ **log likelihood** függvényt maximalizálni. Vagyis keressük azt a $\hat{\theta}$ paramétert, amire $\frac{\partial}{\partial\theta} L\left(\mathcal{X}_n, \hat{\theta}\right) = 0$. Azaz

- Diszkrét: $\sum_{i=1}^{n} \frac{\partial}{\partial\theta} \ln\left(\mathbb{P}\left(X_i \,|\, \hat{\theta}\right)\right) = 0$
- Folytonos: $\sum_{i=1}^{n} \frac{\partial}{\partial\theta} \ln\left(f\left(X_i \,|\, \hat{\theta}\right)\right) = 0$

1. feladat Tegyük fel, hogy X súlyfüggvénye az ismeretlen θ paraméter függvényében az alábbi:

$$\mathbb{P}(X = 0 \,|\, \theta) = \frac{5\theta}{4}, \quad \mathbb{P}(X = 2 \,|\, \theta) = \frac{2 - 2\theta}{4}, \quad \mathbb{P}(X = 4 \,|\, \theta) = \frac{1 - 2\theta}{4}, \quad \mathbb{P}(X = 5 \,|\, \theta) = \frac{1 - \theta}{4}.$$

Az alábbi mintát látjuk: $\mathcal{X}_8 = (0, 2, 4, 4, 5, 2, 5, 5)$.

(a) Mennyi a minta átlaga, tapasztalati szórása? Adjunk a szórásra torzítatlan becslést!

(b) Adjunk ML becslést θ paraméterre!

2. feladat Legyen $f(x) = a^2 x \mathrm{e}^{-ax}$, ha $x \geq 0$. Adjunk ML becslést az a paraméterre tetszőleges $\mathcal{X}_n = (X_1, \ldots, X_n)$ minta esetén!

3. feladat Egy berendezés élettartamára a következő mintánk van (években):

$$\mathcal{X}_{16} = (2, 3, 4, 2.5, 3, 3, 3.5, 4, 2, 4, 3, 3, 3.5, 2, 3, 4).$$

(a) Mennyi a minta átlaga?

(b) Adjunk ML becslést λ-ra, ha feltesszük, hogy a berendezés élettartamának eloszlása $\mathrm{EXP}(\lambda)$!

(c) Adjunk ML becslést μ-re, ha feltesszük, hogy a berendezés élettartamának eloszlása $\mathcal{N}(\mu, 1)$!

4. feladat Egy cinkelt érmét dobálunk, ahol a fej valószínűsége p.

(a) A következő adatsorunk van az első fej megjelenésére:

$$\mathcal{X}_{16} = (2,\ 3,\ 1,\ 2,\ 3,\ 3,\ 2,\ 1,\ 4,\ 4,\ 4,\ 2,\ 3,\ 2,\ 1,\ 4).$$

Adjunk ML becslést p-re!

(b) Mi az ML becslés p-re tetszőleges $\mathcal{X}_n = (a_1, a_2, \ldots, a_n)$ minta esetén?

Konfidenciaintervallum

Ismeretlen θ paraméter lehetséges értékére egy intervallumot adunk $\mathcal{X}_n$ minta alapján, amiben az elvárt valószínűséggel beleesik az igazi értéke a paraméternek:

$$\mathbb{P}(l(\mathcal{X}_n) \leq \theta \leq u(\mathcal{X}_n)) = 1 - \alpha.$$

$[l(\mathcal{X}_n),\ u(\mathcal{X}_n)]$ a **konfidenciaintervallum**, $1-\alpha$ a **konfidenciaszint**.

Kétoldali Z-próba

$X \sim \mathcal{N}(\mu,\ \sigma^2)$, ahol σ ismert, μ ismeretlen. Ekkor $\mathcal{X}_n$ n elemű minta alapján μ-re az $1-\alpha$ szinthez tartozó konfidenciaintervallum:

$$\left(\overline{\mathcal{X}}_n - \frac{z_{\alpha/2} \cdot \sigma}{\sqrt{n}},\ \overline{\mathcal{X}}_n + \frac{z_{\alpha/2} \cdot \sigma}{\sqrt{n}}\right),$$

ahol $z_{\alpha/2}$ az az érték, aminél csak $\alpha/2$ valószínűséggel nagyobb egy standard normális eloszlású változó, azaz $\Phi\left(z_{\alpha/2}\right) = 1 - \frac{\alpha}{2}$.

Kétoldali T-próba

Ugyanaz a feladat, mint kétoldali Z-próbánál, de ismeretlen σ-val. Ekkor μ-re az $1-\alpha$ szinthez tartozó konfidenciaintervallum:

$$\left(\overline{\mathcal{X}}_n - \frac{t_{\alpha/2,n-1} \cdot \hat{s}^*}{\sqrt{n}},\ \overline{\mathcal{X}}_n + \frac{t_{\alpha/2,n-1} \cdot \hat{s}^*}{\sqrt{n}}\right),$$

ahol $\hat{s}^*$ torzítatlan becslés a szórásra (lásd fent (1) képlet), míg $t_{\alpha/2,n-1}$ az az érték, aminél csak $\alpha/2$ valószínűséggel nagyobb egy $n-1$ szabadságfokú T-eloszlású változó. Azaz $T_{n-1}\left(t_{\alpha/2,n-1}\right) = 1 - \frac{\alpha}{2}$, ahol T_{n-1} jelöli az $(n-1)$ szabadságfokú T-eloszlás eloszlásfüggvényét. Ennek az értékei is táblázatból olvashatóak ki.

Egyoldali Z-próba

Hasonló a feladat, mint kétoldali Z-próbánál, de most egy $(-\infty,\ a)$ (vagy $(a,\ \infty)$) alakú intervallumot akarunk meghatározni, amibe $1-\alpha$ valószínűséggel beleesik μ. Ekkor a megfelelő konfidenciaintervallum:

$$\left(-\infty\,,\ \overline{\mathcal{X}}_n + \frac{z_\alpha \cdot \sigma}{\sqrt{n}}\right) \qquad \left(\text{vagy } (a,\infty) \text{ esetén: } \left(\overline{\mathcal{X}}_n - \frac{z_\alpha \cdot \sigma}{\sqrt{n}}\,,\ \infty\right)\right),$$

ahol z_α az az érték, aminél csak α valószínűséggel nagyobb egy standard normális eloszlású változó, azaz $\Phi\left(z_\alpha\right) = 1 - \alpha$.

5. feladat Egy X normális eloszlású véletlen változóval generáltuk az alábbi mintát:

$$\mathcal{X}_9 = (5,\ 6,\ 7,\ 4,\ 5.5,\ 6,\ 8,\ 7,\ 10).$$

(a) Adjunk X várható értékére 95%-os konfidenciaintervallumot!

(b) Hogyan változik az intervallum, ha megtudjuk, hogy az eloszlás szórása 2?

(c) Adjuk meg mindkét esetben a 99%-os konfidenciaintervallumot is!

(d) Mekkora az az érték, aminél 90%-os valószínűséggel kisebb a várható érték? (A szórásról továbbra is tudjuk, hogy 2.)