A sztochasztika alapjai

13. feladatsor: becslések

1. Egy almáskertben véletlenszerűen, egymástól függetlenül találhatók fertőzött fák. Tíz egyforma nagy, egyenként három sorból álló ültetvényben rendre 0, 3, 0, 1, 0, 0, 2, 1, 1, 2 beteg fát találtak.

(a) Adjuk meg az empirikus eloszlásfüggvényt, a mintaátlagot és az empirikus szórásnégyzetet!

(b) Tegyük fel, hogy a beteg fák száma Poisson-eloszlást követ. Adjunk maximum likelihood becslést és momentumbecslést az egy sorban található fák számának várható értékére!

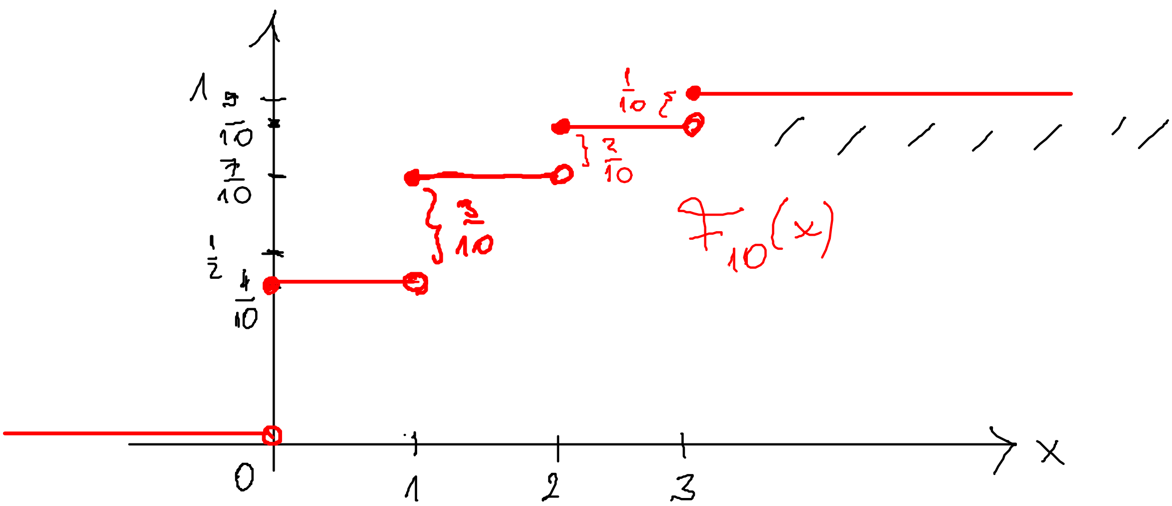
a) $x_1 = 0 \quad x_2 = 3 \quad x_3 = 0 \quad x_4 = 1 \quad \ldots, \quad x_{10} = 2$ realizáció

$n = 10$ mintaelemszám $\quad \bar{X}_{10} = \frac{1}{10}\sum_{i=1}^{10} x_i$ mintaátlag

$s_{10}^2 = \frac{1}{10}\sum_{i=1}^{10}(x_i - \bar{X}_{10})^2$ empirikus szórásnégyzet

empirikus eloszlásfgv.:

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} I(X_i \leq x) = \frac{1}{n}\#\{i : X_i \leq x\}$$



b) $X_1, X_2, \ldots \sim \text{Poisson}(\lambda)$, $\lambda \geq 0$ ismeretlen

$E(X) = \lambda$ $\qquad$ $P(X=k) = \frac{\lambda^k}{k!} e^{-\lambda}$

becsüljük λ-t! $\qquad$ $k = 0, 1, 2, \ldots$

Maximum Likelihood:

$$L_\lambda(x_1, \ldots, x_{10}) = P_\lambda((X_1, \ldots, X_{10}) = (x_1, \ldots, x_{10})) =$$

$$\stackrel{\text{fgt}}{=} \prod_{i=1}^{10} P_\lambda(X_i = x_i) = \prod_{i=1}^{10} \left(\frac{\lambda^{x_i}}{x_i!} e^{-\lambda} \right)$$

$$= e^{-10 \cdot \lambda} \cdot \lambda^{\sum_{i=1}^{10} x_i} \frac{1}{\prod_{i=1}^{10} x_i!}$$

ML elv: válasszuk azt a λ-t, ami maximalizálja ezt a valséget! $\Rightarrow \hat{\lambda}_{ML}$ $\qquad \lambda \geq 0$

$$\hat{\lambda}_{ML} = \frac{\sum_{i=1}^{10} x_i}{10} = \overline{x}_{10}.$$

Momentumbecslés:

ezt látom $\rightarrow$ $\overline{x}_{10}$: mintaátlag $\sim$ várható érték empirikus megjelenése

ismeretlen

ezt fogom számolni

$\overline{x}_{10}$: „legjobb becslés a várható értékre"

Válasszuk úgy $\hat{\lambda}_m$-et, hogy az elméleti várható érték éppen a $\bar{x}_{10}$ mintaátlag.

Mivel $E_\lambda(X) = \lambda \Rightarrow \hat{\lambda}_m = \bar{X}_{10}$.

2. Egy adatszerverre a lekérdezések exponenciális időközönként érkeznek, ahol ismeretlen paraméterrel. Az időközökre percben mérve a következő adatokat kaptuk: 1,94, 0,33, 2,51, 5,27, 1,73, és 0,61. Adjunk becslést a paraméterre a maximum likelihood és a momentumbecslés alkalmazásával!

$n = 6$ $\quad x_1 = 1{,}94 \quad x_2 = 0{,}33 \quad \dots \quad x_6 = 0{,}61$ ← realizáció

Tudom/feltesszem: $X \sim Exp(\lambda)$ $\quad \lambda > 0$

$f_\lambda(x) = \lambda e^{-\lambda x}, \; x > 0$ ahol λ <u>ismeretlen</u>

$E(X) = \frac{1}{\lambda}$.

ezt akarom becsülni

ML becslés

$P(X = 1{,}94) = 0$ mivel folytonos az elo.

likelihood:

$$L_\lambda(x_1, \dots, x_n) = f_\lambda(x_1, \dots, x_n) \stackrel{\text{fgtl}}{=} \prod_{i=1}^{n} f_\lambda(x_i) =$$

együttes sűrűség $(x_1, \dots, x_n)$ -ben, ha a valódi paraméter λ

$$= \prod_{i=1}^{n} \lambda e^{-\lambda x_i}$$

$$= \lambda^n \cdot e^{-\lambda \sum_{i=1}^{n} x_i}$$

$L_\lambda = \lambda^n \cdot e^{-\lambda \sum x_i}$ ← ennek a maximuma kell λ-ban.

$\to \hat{\lambda}_{ML}$

log-likelihood:

$$\ell(\lambda) = \ln L_\lambda = \ln\left(\lambda^n e^{-\lambda \sum x_i}\right) = n \cdot \ln \lambda - \lambda \cdot \sum_1^n x_i$$

$$\ell(\lambda) = n \cdot \ln \lambda - \lambda \cdot \sum_1^n x_i \qquad \lambda > 0.$$

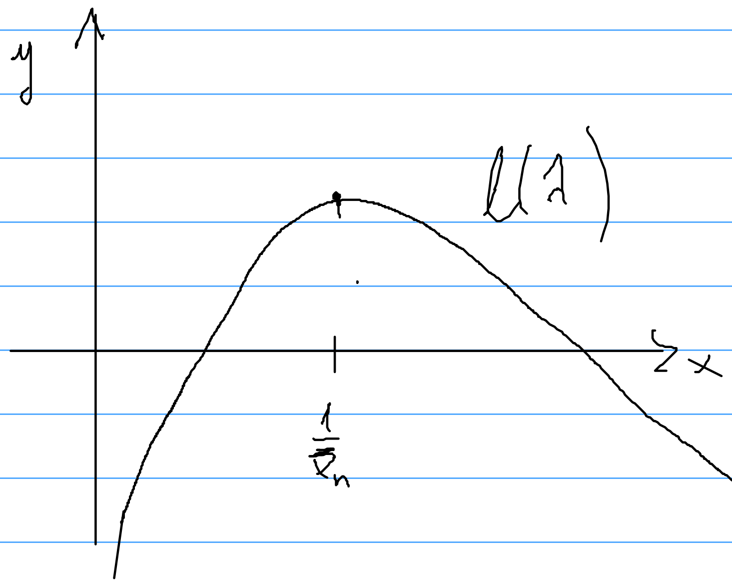
Határozzuk meg a max. helyét.

$$\ell'(\lambda) = \frac{n}{\lambda} - \sum_1^n x_i = 0 \iff \frac{n}{\lambda} = \sum_1^n x_i$$

$$\iff \lambda = \frac{n}{\sum_1^n x_i} = \frac{1}{\bar{x}_n}$$

	$\left(0, \frac{1}{\bar{x}_n}\right)$	$\left(\frac{1}{\bar{x}_n}, \infty\right)$
ℓ	$\nearrow$	$\searrow$
ℓ'	$+$	$-$

$\Rightarrow \frac{1}{\bar{x}_n}$ globális max



$$\boxed{\hat{\lambda}_{ML} = \frac{1}{\bar{x}_n}.}$$

Momentumbecslés: $X \sim Exp(\lambda)$ $\quad E(X) = \frac{1}{\lambda}$.

Azt a $\hat{\lambda}_m$ paramétert választom, amihez tartozó elméleti várható érték megegyezik a mintaátlaggal.

$$\lambda \to m_1(\lambda) = E_\lambda(X) = \frac{1}{\lambda}$$

invertálható

$$m_1(\hat{\lambda}_m) = \overline{X}_n \Rightarrow \hat{\lambda}_m = m_1^{-1}(\overline{X}_n)$$

$$\Rightarrow \hat{\lambda}_m = \frac{1}{\overline{X}_n}.$$

3. Legyenek $X_1, \dots, X_n$ független véletlen változók, $f_\theta(x) = \frac{2x}{3\theta^2}$, $\theta \le x \le 2\theta$, sűrűségfüggvénnyel. Adjunk becslést θ-ra momentum módszerrel és ML módszerrel is!

$$f_\theta(x) = \frac{2x}{3\theta^2} \qquad x \in [\theta, 2\theta]$$

Momentum:

$$E_\theta(X) = \int_{-\infty}^{\infty} x f_\theta(x)\,dx = \int_{\theta}^{2\theta} x \cdot \frac{2x}{3\theta^2}\,dx$$

$$= \frac{2}{3\theta^2} \cdot \int_{\theta}^{2\theta} x^2\,dx = \frac{2}{3\theta^2} \cdot \frac{(2\theta)^3 - \theta^3}{3} = \frac{14}{9} \cdot \theta .$$

$$m_1(\theta) = \frac{14}{9}\theta \qquad \hat{\theta}_m = m_1^{-1}(\overline{X}_n)$$

válasszuk úgy, hogy a hozzá tartozó elméleti várható érték éppen megegyezzen a mintaátlaggal.

$$\boxed{\hat{\theta}_m = \frac{9}{14} \cdot \overline{X}_n}$$

Maximum likelihood:

$$L_\theta(x_1,\dots,x_n) = f_\theta(x_1,\dots,x_n) = \prod_{i=1}^{n} f_\theta(x_i) =$$

$$= \prod_{i=1}^{n} \frac{2}{3\theta^2} x_i \cdot I(\theta \le x_i \le 2\theta)$$

$$= \left(\frac{2}{3\theta^2}\right)^n \cdot \prod_{i=1}^{n} x_i \cdot I(\theta \le \min x_i, \max x_i \le 2\theta).$$

Ezt kell maximalizálni θ-ban.

$$L(\theta) = \left(\frac{2}{3}\right)^n \cdot \prod_{i=1}^{n} x_i \cdot \theta^{-2n} \cdot I(\theta \le \min x_i, 2\theta \ge \max x_i)$$

$$I\left(\frac{1}{2}\max x_i \le \theta \le \min x_i\right)$$

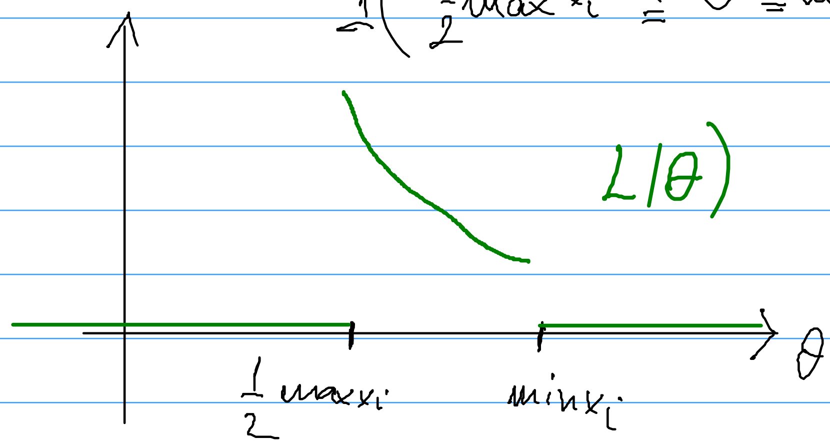
↓

nem folytonos

↓

ne deriváljunk

és nélkül



$$\hat{\theta}_{ML} = \frac{1}{2} \cdot \max_{1 \le i \le n} x_i$$

X lehetséges értékei: $[\theta, 2\theta]$

U: ffi / színtévesztő

V: nő

4. A színtévesztés leggyakoribb fajtája az X kromoszómához kapcsoltan, nemhez kötötten öröklődik. Emiatt a férfiaknál a gyakoriság p, míg a nőknél csak p^2. Adjunk maximum likelihood becslést p-re, az alapján, hogy M férfiból m, N nőből pedig n volt színtévesztő!

$$\left.\begin{matrix} M \to m \\ N \to n \end{matrix}\right\} \quad \text{becsüljük } p\text{-t!}$$

Rögzített p esetén:

U, V ftn

$$P_p(U=m, V=n) = P_p(U=m) \cdot P_p(V=n)$$

$$P_p(U=m) = p^m \cdot (1-p)^{M-m} \cdot \binom{M}{m}$$

$U \sim$ Binomiális (M, p)

M ffi pontosan m színtévesztő

$$P_p(V=n) = \binom{N}{n} \cdot (p^2)^n \cdot (1-p^2)^{N-n}$$

$V \sim$ Binomiális (N, p^2)

U, V függetlenek

$$L(p) = \binom{M}{m} p^m (1-p)^{M-m} \cdot \binom{N}{n} (p^2)^n \cdot (1-p^2)^{N-n}$$

$p \in [0, 1]$

$$= \binom{M}{m}\binom{N}{n} \cdot p^{m+2n} \cdot (1-p)^{M-m} \cdot (1-p^2)^{N-n}$$

Keressük $L(p)$ maximumát p-ben.

$$\ell(p) = \ln L(p) = \ln\left[\binom{M}{m}\binom{N}{n}\right] +$$

$$+ (m+2n) \cdot \ln p + (M-m)\ln(1-p) + (N-n)\ln(1-p^2)$$

$$\ell'(p) = (m+2n) \cdot \frac{1}{p} - (M-m)\frac{1}{1-p} - (N-n)\frac{2p}{1-p^2}$$

$$\ell'(p) = 0$$

$$\frac{m+2n}{p} = \frac{M-m}{1-p} + \frac{(N-n)\,2p}{1-p^2} \quad \Big/ \cdot p(1-p^2)$$

$\vdots$

p-ben másodfokú $\rightarrow$ $\hat{p}_{ML} = \ldots$

5. Családok jövedelmét egy olyan skálán mérjük, ahol $X = 1$ a létminimumnak felel meg. Feltételezzük, hogy a jövedelem eloszlása $f(x) = \frac{\theta}{x^{\theta+1}}$, $x \geq 1$, sűrűségfüggvénnyel adható meg. Adjunk maximum likelihood becslést θ-ra, ha 10 véletlenszerűen választott család jövedelme: 1,53, 2,76, 19,65, 4,16, 7,31, 1,21, 254,2, 5,43, 1,12, 1,63.

dolgozat: Nevezetes eloszlások

bin, Poisson, geo

exp, normális

eloszlástáblázat!!

CHT: de Moivre-Laplace

stat.: emp. ét., mintaátlag

becslés ML, momentum