

Fisher-információ (6. előadás)

Definíció (Likelihood-függvény)

Ha az (Y_1, \dots, Y_n) független minta diszkrét (a lehetséges értékeinek száma véges vagy megszámlálható sok), akkor a likelihood-függvénye:

$$L_{n,\vartheta}(k_1, \dots, k_n) = \prod_{j=1}^n \mathbb{P}_{j,\vartheta}(Y_j = k_j) \quad ((k_1, \dots, k_n) \in H).$$

Fisher-információ (6. előadás)

Definíció (Likelihood-függvény)

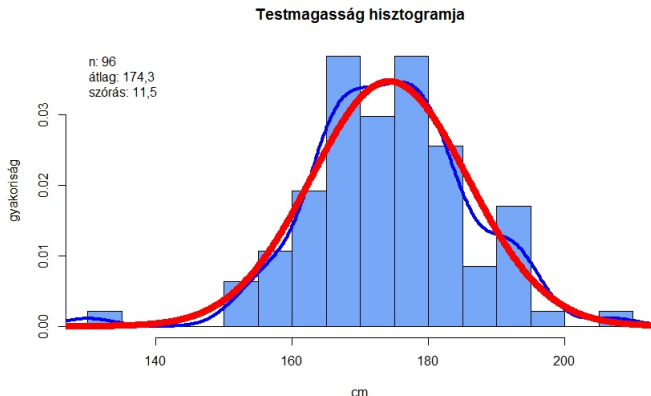
Ha az (Y_1, \dots, Y_n) független minta diszkrét (a lehetséges értékeinek száma véges vagy megszámlálható sok), akkor a likelihood-függvénye:

$$L_{n,\vartheta}(k_1, \dots, k_n) = \prod_{j=1}^n \mathbb{P}_{j,\vartheta}(Y_j = k_j) \quad ((k_1, \dots, k_n) \in H).$$

Ha az (Y_1, \dots, Y_n) független minta abszolút folytonos, és Y_j sűrűségfüggvénye (a \mathbb{P}_ϑ valószínűség mellett) $f_{j,\vartheta}$, akkor a minta likelihood-függvénye:

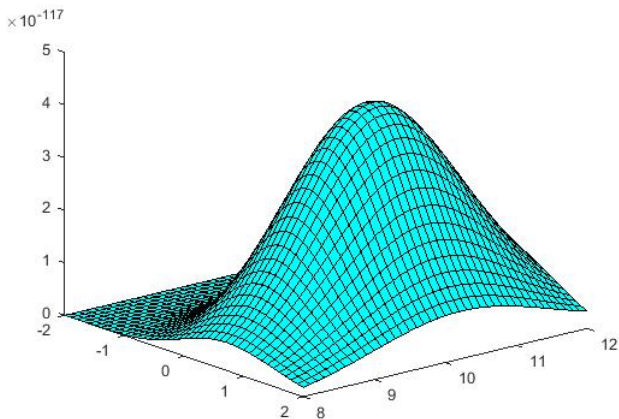
$$L_{n,\vartheta}(t_1, \dots, t_n) = \prod_{j=1}^n f_{j,\vartheta}(t_j) \quad (t_1, \dots, t_n \in \mathbb{R}).$$

Testmagasság hisztogramja



A testmagasság hisztogramja $n = 96$ elemű mintából és az $\bar{X} = 174,3$ várható értékű és $s_n^* = 11,5$ szórású normális eloszlás sűrűségfüggvénye (pirossal).

Likelihoodfüggvény



$n = 94$ elemű minta testmagasság-adatok alapján, normális eloszlást feltételezve.
Az átlag: $\bar{X} = 174,8$, a tapasztalati szórás $s_n = 10,5$.

ML-becslés: normális eloszlás

X_1, \dots, X_n függetlenek, eloszlásuk normális eloszlás $m, \sigma > 0$ paraméterekkel. Ekkor

$$L_{n,m,\sigma}(X_1, \dots, X_n) = \prod_{j=1}^n f_{j,\vartheta}(X_j) = \prod_{j=1}^n \left[\frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(X_j - m)^2}{2\sigma^2}\right) \right].$$

ML-becslés: normális eloszlás

X_1, \dots, X_n függetlenek, eloszlásuk normális eloszlás $m, \sigma > 0$ paraméterekkel. Ekkor

$$L_{n,m,\sigma}(X_1, \dots, X_n) = \prod_{j=1}^n f_{j,\vartheta}(X_j) = \prod_{j=1}^n \left[\frac{1}{\sqrt{2\pi\sigma}} \exp\left(-\frac{(X_j - m)^2}{2\sigma^2}\right) \right].$$

$$L_{n,m,\sigma}(X_1, \dots, X_n) = \left(\frac{1}{\sqrt{2\pi\sigma}}\right)^n \exp\left(-\sum_{j=1}^n \frac{(X_j - m)^2}{2\sigma^2}\right).$$

ML-becslés: normális eloszlás

X_1, \dots, X_n függetlenek, eloszlásuk normális eloszlás $m, \sigma > 0$ paraméterekkel. Ekkor

$$L_{n,m,\sigma}(X_1, \dots, X_n) = \prod_{j=1}^n f_{j,\vartheta}(X_j) = \prod_{j=1}^n \left[\frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(X_j - m)^2}{2\sigma^2}\right) \right].$$

$$L_{n,m,\sigma}(X_1, \dots, X_n) = \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n \exp\left(-\sum_{j=1}^n \frac{(X_j - m)^2}{2\sigma^2}\right).$$

$$\log L_{n,m,\sigma}(X_1, \dots, X_n) = -n \log(\sqrt{2\pi}) - n \log \sigma - \sum_{j=1}^n \frac{(X_j - m)^2}{2\sigma^2}.$$

ML-becslés: normális eloszlás

X_1, \dots, X_n függetlenek, eloszlásuk normális eloszlás $m, \sigma > 0$ paraméterekkel. Ekkor

$$L_{n,m,\sigma}(X_1, \dots, X_n) = \prod_{j=1}^n f_{j,\vartheta}(X_j) = \prod_{j=1}^n \left[\frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(X_j - m)^2}{2\sigma^2}\right) \right].$$

$$L_{n,m,\sigma}(X_1, \dots, X_n) = \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n \exp\left(-\sum_{j=1}^n \frac{(X_j - m)^2}{2\sigma^2}\right).$$

$$\log L_{n,m,\sigma}(X_1, \dots, X_n) = -n \log(\sqrt{2\pi}) - n \log \sigma - \sum_{j=1}^n \frac{(X_j - m)^2}{2\sigma^2}.$$

A likelihood-függvényben csak a $\sum_{j=1}^n X_j^2$ **négyzetösszeg** és a $\sum_{j=1}^n X_j$ **összeg** szerepel a mintából a paraméterrel együtt:

$$\sum_{j=1}^n (X_j - m)^2 = \sum_{j=1}^n X_j^2 - 2m \sum_{j=1}^n X_j + m^2.$$

Fisher-információ

Egy mintaelem **Fisher-információja**:

$$I_1(\vartheta) = \mathbb{E}_{\vartheta} \left(\left(\frac{\partial}{\partial \vartheta} \log f_{\vartheta}(X_1) \right)^2 \right),$$

ahol f_{ϑ} a likelihoodfüggvény \mathbb{P}_{ϑ} mellett.

Fisher-információ

Egy mintaelem **Fisher-információja**:

$$I_1(\vartheta) = \mathbb{E}_{\vartheta} \left(\left(\frac{\partial}{\partial \vartheta} \log f_{\vartheta}(X_1) \right)^2 \right),$$

ahol f_{ϑ} a likelihoodfüggvény \mathbb{P}_{ϑ} mellett.

- Megfelelő (regularitási) feltételek mellett a független azonos eloszlású, n elemű minta Fisher-információja: $I_n(\vartheta) = n \cdot I_1(\vartheta)$.
- Ha T elégséges statisztika, akkor $T(X_1, \dots, X_n)$ Fisher-információja ugyanaz, mint (X_1, \dots, X_n) Fisher-információja (például Poisson-eloszlásnál az átlag Fisher-információja ugyanaz, mint a teljes mintáé).
- **Cramér–Rao-egyenlőtlenség**: megfelelő (regularitási) feltételek mellett, ha T torzítatlan becslés ϑ -ra, akkor

$$D^2(T(X)) \geq \frac{1}{I_n(\vartheta)} = \frac{1}{nI_1(\vartheta)}.$$

(Ebben az értelemben feleakkora szóráshoz négyszer annyi mintaelem kell.)

Fisher-információ

Egy mintaelem **Fisher-információja**:

$$I_1(\vartheta) = \mathbb{E}_{\vartheta} \left(\left(\frac{\partial}{\partial \vartheta} \log f_{\vartheta}(X_1) \right)^2 \right),$$

ahol f_{ϑ} a likelihoodfüggvény \mathbb{P}_{ϑ} mellett.

Néhány nevezetes eloszlás Fisher-információja egy mintaelemből:

- binomiális eloszlás n renddel és p paraméterrel:

$$\frac{n}{p(1-p)}.$$

- Poisson-eloszlás λ paraméterrel: $1/\lambda$
- normális eloszlás, ha $\vartheta = (m, \sigma)$:

$$\begin{pmatrix} \frac{1}{\sigma^2} & 0 \\ 0 & \frac{2}{\sigma^2} \end{pmatrix}.$$

Fisher-információ: Poisson-eloszlás

Egy mintaelem **Fisher-információja** (az f_λ likelihood-függvény most a valószínűség):

$$\begin{aligned} I_1(\lambda) &= \mathbb{E}_\lambda \left(\left(\frac{\partial}{\partial \lambda} \log f_\lambda(X_1) \right)^2 \right) = \mathbb{E}_\lambda \left(\left(\frac{\partial}{\partial \lambda} \log \left(\frac{\lambda^{X_1}}{X_1!} e^{-\lambda} \right) \right)^2 \right) = \\ &= \mathbb{E}_\lambda \left(\left(\frac{\partial}{\partial \lambda} \left(X_1 \log \lambda - \log X_1! - \lambda \right) \right)^2 \right) = \mathbb{E}_\lambda \left(\left(\frac{X_1}{\lambda} - 1 \right)^2 \right) = \\ &= \mathbb{E}_\lambda \left(\frac{X_1^2}{\lambda^2} - 2 \frac{X_1}{\lambda} + 1 \right) = \frac{\lambda + \lambda^2}{\lambda^2} - 2 + 1 = \frac{1}{\lambda}. \end{aligned}$$

Felhasználtuk, hogy Poisson-eloszlásnál $\mathbb{E}_\lambda(X_1) = D_\lambda(X_1) = \lambda$, így

$$\mathbb{E}_\lambda(X_1^2) = D_\lambda^2(X_1) + \mathbb{E}_\lambda^2(X_1) = \lambda + \lambda^2.$$

Fisher-információ: Poisson-eloszlás

Egy mintaelem **Fisher-információja** (az f_λ likelihood-függvény most a valószínűség):

$$\begin{aligned} I_1(\lambda) &= \mathbb{E}_\lambda \left(\left(\frac{\partial}{\partial \lambda} \log f_\lambda(X_1) \right)^2 \right) = \mathbb{E}_\lambda \left(\left(\frac{\partial}{\partial \lambda} \log \left(\frac{\lambda^{X_1}}{X_1!} e^{-\lambda} \right) \right)^2 \right) = \\ &= \mathbb{E}_\lambda \left(\left(\frac{\partial}{\partial \lambda} \left(X_1 \log \lambda - \log X_1! - \lambda \right) \right)^2 \right) = \mathbb{E}_\lambda \left(\left(\frac{X_1}{\lambda} - 1 \right)^2 \right) = \\ &= \mathbb{E}_\lambda \left(\frac{X_1^2}{\lambda^2} - 2 \frac{X_1}{\lambda} + 1 \right) = \frac{\lambda + \lambda^2}{\lambda^2} - 2 + 1 = \frac{1}{\lambda}. \end{aligned}$$

Felhasználtuk, hogy Poisson-eloszlásnál $\mathbb{E}_\lambda(X_1) = D_\lambda(X_1) = \lambda$, így

$$\mathbb{E}_\lambda(X_1^2) = D_\lambda^2(X_1) + \mathbb{E}_\lambda^2(X_1) = \lambda + \lambda^2.$$

Következmény a **Cramér–Rao-egyenlőtlenség** alapján: ha T torzítatlan becslése λ -nak n elemű független mintából, akkor

$$D^2(T(X_1, \dots, X_n)) \geq \frac{\lambda}{n}.$$

Konfidenciaintervallum az ML-becslés alapján

Gyakran az ismeretlen paraméterre egyetlen becslés helyett konfidenciaintervallumot adhatunk meg, amire például az igaz, hogy

$$\mathbb{P}_{\vartheta}(T_1 \leq \vartheta \leq T_2) \geq 0,95 \quad (\vartheta \in \Theta).$$

Ez egy olyan véletlen intervallum, ami 95%-os megbízhatóságú, azaz legalább ennyi valószínűséggel tartalmazza a becsülni kívánt paramétert.

Konfidenciaintervallum az ML-becslés alapján

Gyakran az ismeretlen paraméterre egyetlen becslés helyett konfidenciaintervallumot adhatunk meg, amire például az igaz, hogy

$$\mathbb{P}_\vartheta(T_1 \leq \vartheta \leq T_2) \geq 0,95 \quad (\vartheta \in \Theta).$$

Ez egy olyan véletlen intervallum, ami 95%-os megbízhatóságú, azaz legalább ennyi valószínűséggel tartalmazza a becsülni kívánt paramétert.

Állítás

Ha likelihoodfüggvény teljesít bizonyos regularitási feltételeket, akkor a ϑ paraméternek az X_1, X_2, \dots, X_n mintából számolt $\hat{\vartheta}_n$ maximumlikelihood-becslése

- létezik;
- aszimptotikusan torzítatlan: $\lim_{n \rightarrow \infty} \mathbb{E}_\vartheta(\hat{\vartheta}_n) = \vartheta$ minden $\vartheta \in \Theta$ -ra;
- aszimptotikusan hatásos: $\lim_{n \rightarrow \infty} \sqrt{nl_1(\vartheta)} D_\vartheta(\hat{\vartheta}_n) = 1$ minden $\vartheta \in \Theta$ -ra;
- aszimptotikusan normális eloszlású: $\sqrt{nl_1(\vartheta)}(\hat{\vartheta}_n - \vartheta)$ eloszlásban tart a standard normális eloszláshoz minden $\vartheta \in \Theta$ -ra $n \rightarrow \infty$ esetén.

Itt $l_1(\vartheta)$ az egyelemű mintából számolt Fisher-információt jelöli.

Konfidenciaintervallum az ML-becslés alapján

Ez alapján **aszimptotikus konfidenciaintervallum**, ami $n \rightarrow \infty$ esetén $1 - \alpha$ -hoz tartó valószínűséggel tartalmazza ϑ -t:

$$\left(\hat{\vartheta}_n - \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) \cdot \frac{1}{\sqrt{\hat{I}_n(\vartheta)}}; \hat{\vartheta}_n + \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) \cdot \frac{1}{\sqrt{\hat{I}_n(\vartheta)}} \right),$$

ahol $\hat{I}_n(\vartheta) = n \cdot I_1(\hat{\vartheta}_n)$, vagyis a Fisher-információ kifejezésébe a maximumlikelihood-becslést írjuk be.

Másképpen felírva:

$$\lim_{n \rightarrow \infty} \mathbb{P}_{\vartheta} \left(\hat{\vartheta}_n - \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) \cdot \frac{1}{\sqrt{\hat{I}_n(\vartheta)}} \leq \vartheta \leq \hat{\vartheta}_n + \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) \cdot \frac{1}{\sqrt{\hat{I}_n(\vartheta)}} \right) = 1 - \alpha.$$

Itt is vegyük észre, hogy ϑ ismeretlen, de rögzített paraméter, és az intervallum alsó és felső végpontjai véletlenek, az eloszlásuk is ϑ -tól függ (ezt jelöli a ϑ az alsó indexben).

Konfidenciaintervallum az ML-becslés alapján

Tekintsünk egy n elemű független mintát az alábbi sűrűségfüggvénnyel, ahol $\lambda > 0$ ismeretlen paraméter:

$$f_{\lambda}(x) = \frac{\lambda^4}{6} x^3 e^{-\lambda x} \mathbb{I}(x > 0).$$

Konfidenciaintervallum az ML-becslés alapján

Tekintsünk egy n elemű független mintát az alábbi sűrűségfüggvénnyel, ahol $\lambda > 0$ ismeretlen paraméter:

$$f_{\lambda}(x) = \frac{\lambda^4}{6} x^3 e^{-\lambda x} \mathbb{I}(x > 0).$$

A likelihood-függvény, felhasználva, hogy az X_j -k 1 valószínűséggel pozitívak:

$$f(\lambda) = \prod_{j=1}^n f_{\lambda}(X_j) = \prod_{j=1}^n \left(\frac{\lambda^4}{6} X_j^3 e^{-\lambda X_j} \right) = \frac{\prod_{j=1}^n X_j^3}{6^n} \cdot \lambda^{4n} \cdot e^{-\lambda \sum_{j=1}^n X_j}.$$

Konfidenciaintervallum az ML-becslés alapján

Tekintsünk egy n elemű független mintát az alábbi sűrűségfüggvénnyel, ahol $\lambda > 0$ ismeretlen paraméter:

$$f_{\lambda}(x) = \frac{\lambda^4}{6} x^3 e^{-\lambda x} \mathbb{I}(x > 0).$$

A likelihood-függvény, felhasználva, hogy az X_j -k 1 valószínűséggel pozitívak:

$$f(\lambda) = \prod_{j=1}^n f_{\lambda}(X_j) = \prod_{j=1}^n \left(\frac{\lambda^4}{6} X_j^3 e^{-\lambda X_j} \right) = \frac{\prod_{j=1}^n X_j^3}{6^n} \cdot \lambda^{4n} \cdot e^{-\lambda \sum_{j=1}^n X_j}.$$

$$\log f(\lambda) = \log \frac{\prod_{j=1}^n X_j^3}{6^n} + 4n \log \lambda - \lambda \cdot \sum_{j=1}^n X_j.$$

$$\frac{\partial}{\partial \lambda} \log f(\lambda) = \frac{4n}{\lambda} - \sum_{j=1}^n X_j.$$

Konfidenciintervallum az ML-becslés alapján

Tekintsünk egy n elemű független mintát az alábbi sűrűségfüggvénnyel, ahol $\lambda > 0$ ismeretlen paraméter:

$$f_{\lambda}(x) = \frac{\lambda^4}{6} x^3 e^{-\lambda x} \mathbb{I}(x > 0).$$

A likelihood-függvény, felhasználva, hogy az X_j -k 1 valószínűséggel pozitívak:

$$f(\lambda) = \prod_{j=1}^n f_{\lambda}(X_j) = \prod_{j=1}^n \left(\frac{\lambda^4}{6} X_j^3 e^{-\lambda X_j} \right) = \frac{\prod_{j=1}^n X_j^3}{6^n} \cdot \lambda^{4n} \cdot e^{-\lambda \sum_{j=1}^n X_j}.$$

$$\log f(\lambda) = \log \frac{\prod_{j=1}^n X_j^3}{6^n} + 4n \log \lambda - \lambda \cdot \sum_{j=1}^n X_j.$$

$$\frac{\partial}{\partial \lambda} \log f(\lambda) = \frac{4n}{\lambda} - \sum_{j=1}^n X_j.$$

Ez pontosan akkor pozitív, ha $\lambda < \frac{4}{\bar{X}}$, és negatív különben. Ezért a $\log f(\lambda)$ függvény globális maximumhelye $\lambda = \frac{4}{\bar{X}}$, és a logaritmusfüggvény monotonitása miatt ez lesz a maximumlikelihood-becslés is.

Konfidenciaintervallum az ML-becslés alapján

Az előző alapján a Fisher-információt is kiszámíthatjuk, ha tudjuk, hogy

$$\mathbb{E}_\lambda(X_j) = \frac{4}{\lambda}, \text{ és } D_\lambda^2(X_j) = \frac{4}{\lambda^2}.$$

Ugyanis

$$\begin{aligned} I_1(\vartheta) &= \mathbb{E}_\lambda \left(\left(\frac{\partial}{\partial \lambda} \log f_\lambda(X_1) \right)^2 \right) = \mathbb{E}_\lambda \left(\left(\frac{4n}{\lambda} - \sum_{j=1}^n X_j \right)^2 \right) = \\ &= D_\lambda^2 \left(\sum_{j=1}^n X_j \right) = n \cdot D_\lambda^2(X_1) = n \cdot \frac{4}{\lambda^2}. \end{aligned}$$

Konfidenciaintervallum az ML-becslés alapján

Az előző alapján a Fisher-információt is kiszámíthatjuk, ha tudjuk, hogy

$$\mathbb{E}_\lambda(X_j) = \frac{4}{\lambda}, \text{ és } D_\lambda^2(X_j) = \frac{4}{\lambda^2}.$$

Ugyanis

$$\begin{aligned} I_1(\vartheta) &= \mathbb{E}_\lambda \left(\left(\frac{\partial}{\partial \lambda} \log f_\lambda(X_1) \right)^2 \right) = \mathbb{E}_\lambda \left(\left(\frac{4n}{\lambda} - \sum_{j=1}^n X_j \right)^2 \right) = \\ &= D_\lambda^2 \left(\sum_{j=1}^n X_j \right) = n \cdot D_\lambda^2(X_1) = n \cdot \frac{4}{\lambda^2}. \end{aligned}$$

A fentiek alapján 95%-os megbízhatósági szintű aszimptotikus konfidenciaintervallum λ -ra:

$$\left(\frac{4}{\bar{X}} - 1,96 \cdot \frac{4\bar{X}^2}{\sqrt{n}}, \frac{4}{\bar{X}} + 1,96 \cdot \frac{4\bar{X}^2}{\sqrt{n}} \right).$$

Bayes-becslés: bevezetés

Eddig a statisztikai elemzésekben a **frekventista hozzáállást** követtük. A $\vartheta \in \Theta$ paraméter számunkra ismeretlen, és csak azt tesszük fel róla, hogy egy ismert Θ halmaznak, a paramétertérnek az eleme. A Θ paramétertér rögzített halmaz saját struktúra nélkül, a tulajdonságoknak (például konzisztencia, torzítatlanság) minden $\vartheta \in \Theta$ -ra teljesülniük kell. Ezért előzetes információt nem tudunk jól felhasználni.

Bayes-becslés: bevezetés

Eddig a statisztikai elemzésekben a **frekventista hozzáállást** követtük. A $\vartheta \in \Theta$ paraméter számunkra ismeretlen, és csak azt tesszük fel róla, hogy egy ismert Θ halmaznak, a paramétertérnek az eleme. A Θ paramétertér rögzített halmaz saját struktúra nélkül, a tulajdonságoknak (például konzisztencia, torzítatlanság) minden $\vartheta \in \Theta$ -ra teljesülniük kell. Ezért előzetes információt nem tudunk jól felhasználni.

Ennek a módszernek egy hátránya például:

- egy gyógyszert n -szer kipróbálva 0 esetben lépett fel mellékhatás;
- legyen p a mellékhatás kialakulásának valószínűsége;
- ezt a relatív gyakorisággal becsülhetjük: $\hat{p} = 0$. Ez torzítatlan, konzisztens becslés, a frekventista hozzáállás követelményeinek megfelel.
- Ugyanakkor, ha $n = 20$ kipróbálásból nem alakult ki mellékhatás, az egész más, mintha $n = 2000$ kipróbálásnál sem talákoztunk ezzel. Ezt a becslés nem tükrözi.

Bayes-i hozzáállás: a paramétert magát is valószínűségi változónak tekintjük, ebbe beépítve valamilyen előzetes információt.

Beta-eloszlás és érmedobás

- van egy szabályosnak tűnő pénzérménk
- az írás valószínűsége ϑ , ismeretlen paraméter
- a priori sűrűségfüggvény:

Beta-eloszlás és érmedobás

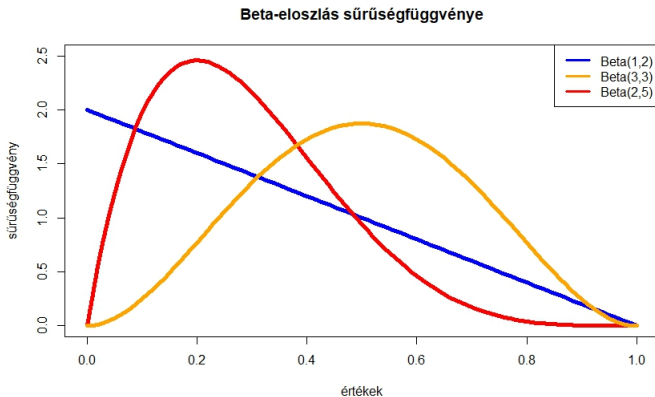
- van egy szabályosnak tűnő pénzérménk
- az írás valószínűsége ϑ , ismeretlen paraméter
- a priori sűrűségfüggvény: a $[0, 1]$ intervallumon kívül 0, és az $1/2$ környékén a legnagyobb
- például a $\text{Beta}(3,3)$ eloszlást választhatjuk a priori eloszlásnak

$$p(t) = ct^2(1-t)^2\mathbb{I}(0 \leq t \leq 1),$$

megfelelő $c > 0$ számmal (hogy 1 legyen az integrál).

- Simon Jackman: Bayesian analysis for the social sciences, Wiley, 2009.

Beta-eloszlás



Beta-eloszlás sűrűségfüggvénye különböző paraméterpárok mellett:

$$f(x) = x^{a-1}(1-x)^{b-1}\mathbb{I}(0 \leq x \leq 1).$$

Beta-eloszlás

Definíció

Az X valószínűségi változó beta-eloszlású a, b paraméterekkel, ha sűrűségfüggvénye:

$$f(x) = \frac{(a+b-1)!}{(a-1)!(b-1)!} x^{a-1}(1-x)^{b-1}, \quad \text{ha } x \in [0, 1],$$

és nulla különben.

Például: $a = 1, b = 1$: egyenletes eloszlás a $[0, 1]$ intervallumon; $a = 2, b = 1$: két független, a $[0, 1]$ intervallumon egyenletes eloszlású valószínűségi változó közül a kisebbnek az eloszlása.

Állítás

Az (a, b) paraméterű beta-eloszlás várható értéke és szórása:

$$\mathbb{E}(X) = \frac{a}{a+b}; \quad D(X) = \sqrt{\frac{ab}{(a+b)^2(a+b+1)}}.$$

Beta-eloszlás

Állítás

Az (a, b) paraméterű beta-eloszlás várható értéke és szórása:

$$\mathbb{E}(X) = \frac{a}{a+b}; \quad D(X) = \sqrt{\frac{ab}{(a+b)^2(a+b+1)}}.$$

Legyen X_1, X_2, \dots, X_n független minta a $[0, 1]$ intervallumon egyenletes eloszlásból. Ekkor X_k^* -nak, vagyis a nagyság szerint növekvő sorban a k . mintaelemnek az eloszlása beta-eloszlás $a = k$ és $b = n - k + 1$ paraméterekkel. Következmény:

$$\mathbb{E}(X_k^*) = \frac{k}{n+1}; \quad D(X_k^*) = \sqrt{\frac{k(n-k+1)}{(n+1)^2(n+2)}}.$$

Például $\mathbb{E}(X_n^*) = \frac{n}{n+1}$, ehhez hasonlóan adódott, hogy ha $[0, \vartheta)$ intervallumon egyenletes eloszlásból van n elemű mintánk, akkor $\frac{n+1}{n}X_n^*$ torzítatlan becslés volt ϑ -ra:

$$\mathbb{E}\left(\frac{n+1}{n}X_n^*\right) = \frac{n+1}{n}\mathbb{E}(X_n^*) = \frac{n+1}{n} \cdot \frac{n}{n+1}\vartheta = \vartheta$$

minden n -re és minden ϑ -ra. Másrészt az X_n^* aszimptotikusan torzítatlan ϑ -ra. ↻↻↻

A priori és prediktív eloszlás

A Bayes-becslés elkészítéséhez az alábbi fogalmakra lesz szükség.

A θ paramétert valószínűségi változónak tekintjük. Ennek segítségével építjük be az előzetes információkat. A választás tetszőleges lehet, de ettől függni fog a becslés is.

Θ az ismeretlen ϑ paraméter összes lehetséges értékéből álló halmaz.

A priori eloszlás: eloszlás a Θ paraméterterén, sűrűségfüggvénye legyen p .

A priori és prediktív eloszlás

A Bayes-becslés elkészítéséhez az alábbi fogalmakra lesz szükség.

A θ paramétert valószínűségi változónak tekintjük. Ennek segítségével építjük be az előzetes információkat. A választás tetszőleges lehet, de ettől függni fog a becslés is.

Θ az ismeretlen ϑ paraméter összes lehetséges értékéből álló halmaz.

A priori eloszlás: eloszlás a Θ paraméterterén, sűrűségfüggvénye legyen p .

Prediktív eloszlás: a minta eloszlása az a priori eloszlás alapján, feltétel nélkül (kisorsoljuk ϑ -t, majd a mintát a ϑ -hoz tartozó eloszlásból).

Sűrűségfüggvénye a teljes valószínűség tételének egy folytonos analógiájával:

$$f_p(x) = \int_{\Theta} f_{\vartheta}(x)p(\vartheta)d\vartheta,$$

ahol f_{ϑ} a minta sűrűségfüggvénye a ϑ paraméter mellett.

Ha a megfigyelések diszkrét, akkor $f_{\vartheta}(x)$ helyett $\mathbb{P}_{\vartheta}(X = x)$ (ugyanúgy, mint a likelihood-függvénynél). A prediktív eloszlás a paramétertől nem függ.

A posteriori eloszlás

A posteriori eloszlás: amikor megfigyeljük a mintát, a paraméterről kapunk információt. Erre feltételes valószínűséget számolva bizonyos paramétertartományok valószínűbbek, mások kevésbé valószínűek lesznek, mint az előzetes, a priori eloszlás esetében.

Az a posteriori eloszlás azt adja meg, hogy a minta megfigyelt értékeire feltételesen mi lesz a ϑ paraméter eloszlása a Θ paraméterterén. Sűrűségfüggvénye a Bayes-tétel analógiája szerint:

$$p^*(\vartheta | \underline{X} = \underline{x}) = \frac{L_{\vartheta}(x_1, \dots, x_n) p(\vartheta)}{L_p(x_1, \dots, x_n)},$$

ahol L_{ϑ} a likelihood-függvény ϑ mellett, \underline{X} a minta, $\underline{x} = (x_1, \dots, x_n)$ a megfigyelt értékek, L_p pedig az f_p -ből számolt likelihood-függvény: $L_p(x_1, \dots, x_n) = \prod_{j=1}^n f_p(x_j)$, ahol $f_p(x) = \int_{\Theta} f_{\vartheta}(x) p(\vartheta) d\vartheta$.

A priori és a posteriori eloszlás: példa.

Egy érmével dobva n dobásból k írás lett. Az írások száma legyen Y . Az írás (I) valószínűsége ϑ . Az **a priori eloszlás** legyen $\text{Beta}(3, 3)$ a $\Theta = [0, 1]$ paraméterterén. Sűrűségfüggvénye az 2. definíció alapján:

$$p(x) = \frac{5!}{2! \cdot 2!} x^2 (1-x)^2 \mathbb{I}(0 \leq x \leq 1) = 30x^2(1-x)^2 \mathbb{I}(0 \leq x \leq 1).$$

A priori és a posteriori eloszlás: példa.

Egy érmével dobva n dobásból k írás lett. Az írások száma legyen Y . Az írás (I) valószínűsége ϑ . Az **a priori eloszlás** legyen $\text{Beta}(3, 3)$ a $\Theta = [0, 1]$ paraméterterén. Sűrűségfüggvénye az 2. definíció alapján:

$$p(x) = \frac{5!}{2! \cdot 2!} x^2 (1-x)^2 \mathbb{I}(0 \leq x \leq 1) = 30x^2(1-x)^2 \mathbb{I}(0 \leq x \leq 1).$$

Prediktív eloszlás: annak valószínűsége, hogy írást dobunk egy dobásnál. Ha a paraméter ϑ , akkor ez éppen ϑ valószínűséggel történik. A teljes valószínűség tételének egy folytonos változatát alkalmazhatjuk:

$$\mathbb{P}_p(I) = \int_{\Theta} \mathbb{P}_{\vartheta}(I) p(\vartheta) d\vartheta = \int_0^1 \vartheta \cdot p(\vartheta) d\vartheta = \frac{3}{6} = \frac{1}{2}.$$

Itt $\mathbb{P}_{\vartheta}(I)$ az írás dobás valószínűsége, ha az igazi paraméter ϑ , ez definíció szerint ϑ .

A priori és a posteriori eloszlás: példa

A posteriori eloszlás: feltéve, hogy n dobásból k írás lett, mi a ϑ paraméter eloszlása a $[0, 1]$ intervallumon?

Ha a paraméter ϑ rögzített, akkor az írás dobások száma binomiális eloszlású.

A nevezőben a prediktív eloszlás szerint számolhatunk: ott is binomiális eloszlás jelenik meg, $1/2$ paraméterrel.

$$\begin{aligned} p^*(\vartheta | Y = k) &= \frac{L_{\vartheta}(x_1, \dots, x_n) p(\vartheta)}{L_p(x_1, \dots, x_n)} = \frac{\mathbb{P}_{\vartheta}(Y = k) p(\vartheta)}{\mathbb{P}_p(Y = k)} = \\ &= \frac{\binom{n}{k} \vartheta^k (1 - \vartheta)^{n-k} \cdot 30 \cdot \vartheta^2 (1 - \vartheta)^2}{\binom{n}{k} \left(\frac{1}{2}\right)^k \left(1 - \frac{1}{2}\right)^{n-k}} = \\ &= 30 \cdot 2^n \cdot \vartheta^{k+2} (1 - \vartheta)^{n-k+2}. \end{aligned}$$

A priori és a posteriori eloszlás: példa

A posteriori eloszlás: feltéve, hogy n dobásból k írás lett, mi a ϑ paraméter eloszlása a $[0, 1]$ intervallumon?

Ha a paraméter ϑ rögzített, akkor az írás dobások száma binomiális eloszlású.

A nevezőben a prediktív eloszlás szerint számolhatunk: ott is binomiális eloszlás jelenik meg, $1/2$ paraméterrel.

$$\begin{aligned} p^*(\vartheta | Y = k) &= \frac{L_{\vartheta}(x_1, \dots, x_n) p(\vartheta)}{L_p(x_1, \dots, x_n)} = \frac{\mathbb{P}_{\vartheta}(Y = k) p(\vartheta)}{\mathbb{P}_p(Y = k)} = \\ &= \frac{\binom{n}{k} \vartheta^k (1 - \vartheta)^{n-k} \cdot 30 \cdot \vartheta^2 (1 - \vartheta)^2}{\binom{n}{k} \left(\frac{1}{2}\right)^k \left(1 - \frac{1}{2}\right)^{n-k}} = \\ &= 30 \cdot 2^n \cdot \vartheta^{k+2} (1 - \vartheta)^{n-k+2}. \end{aligned}$$

Az a posteriori eloszlás Beta-eloszlás $a = k + 3$ és $b = n - k + 3$ paraméterekkel. Ha például $k < n/2$, azaz a dobások kevesebb mint fele lett írás, akkor az a posteriori sűrűségfüggvény is az $1/2$ -nél kisebb tartományon lesz nagyobb.

Veszteségfüggvény és Bayes-becslés

A Bayes-becslésnél is meg kell választani, hogy milyen szempont szerint legyen optimális a becsült érték.

Veszteségfüggvény: a veszteség, ha az igazi ϑ paraméter helyett annak $\hat{\vartheta}$ becslését használjuk, ez $W(\vartheta, \hat{\vartheta})$. Ez nemnegatív, $\vartheta - \hat{\vartheta}$ függvénye, például $(\vartheta - \hat{\vartheta})^2$ vagy $|\vartheta - \hat{\vartheta}|$.

Definíció

A $\vartheta \in \Theta$ paraméter Bayes-becslése az a $\hat{\vartheta} = T(X_1, \dots, X_n)$ becslés, melyre az

$$R_Q(T) = \mathbb{E}_{\vartheta}(\mathbb{E}(W(T(X_1, \dots, X_n), \vartheta)))$$

a priori bayesi rizikó minimális.

Vagyis a cél olyan $T(X_1, X_2, \dots, X_n)$ statisztika keresése, melyre a becslés hibájából adódó veszteség várható értéke a lehető legkisebb. A várható érték a megfigyelésekre és a paraméter sorsolására is vonatkozik.

Négyzetes veszteségfüggvény

Abban a speciális esetben, amikor a négyzetes veszteségfüggvényt használjuk, általánosan is megfogalmazható, hogy hogyan találhatjuk meg az optimális becslést. Erről szól az alábbi tétel.

Tétel

Ha a $W(x, y) = (x - y)^2$ négyzetes veszteségfüggvényt használjuk, akkor a paraméter $g(\vartheta)$ függvényének Bayes-becslése az a g várható értéke az a posteriori eloszlás szerint:

$$\widehat{g(\vartheta)} = \int_{\Theta} g(\vartheta) p^*(\vartheta | \underline{X} = \underline{x}) d\vartheta.$$

Bayes-becslés: példa

Egy érmével dobva n dobásból k írás lett. Az írások száma legyen Y . Az írás (I) valószínűsége ϑ .

Az **a priori eloszlás** legyen $\text{Beta}(3, 3)$ -eloszlás a $\Theta = [0, 1]$ paraméterterén, mint eddig. Sűrűségfüggvénye:

$$p(x) = 30x^2(1-x)^2\mathbb{I}(0 \leq x \leq 1).$$

Azt láttuk korábban, hogy az **a posteriori eloszlás** beta-eloszlás $a = k + 3$ és $b = n - k + 3$ paraméterekkel.

$W(x, y) = (x - y)^2$ négyzetes veszteségfüggvény esetén a ϑ paraméter becslése a tétel szerint az a posteriori eloszlás várható értéke:

$$\begin{aligned}\hat{\vartheta} &= \int_{\Theta} \vartheta \cdot p^*(\vartheta | \underline{X} = \underline{x}) d\vartheta = \int_0^1 \vartheta \cdot p^*(\vartheta | \underline{X} = \underline{x}) d\vartheta = \\ &= \int_0^1 \vartheta \cdot 30 \cdot 2^n \cdot \vartheta^{k+2} (1 - \vartheta)^{n-k+2} d\vartheta = \frac{k+3}{(k+3) + (n-k+3)} = \frac{k+3}{n+6},\end{aligned}$$

hiszen éppen az $a = k + 3$ és $b = n - k + 3$ paraméterű beta-eloszlás várható értéke jelent meg, amire pedig használhatjuk a 3. állítást.

Bayes-becslés: példa

Tehát például ha $n = 10$ dobásból $k = 0$ írás van, és $\text{Beta}(3, 3)$ az a priori eloszlás, akkor

$$\hat{\vartheta} = \frac{3}{16} = 18,8\%.$$

Ha $n = 100$ dobásból $k = 0$ írás van, akkor

$$\hat{\vartheta} = \frac{3}{106} = 2,8\%.$$

Ugyanezen feltételek mellett, ha $n = 100$ dobásból $k = 32$ írás van, akkor

$$\hat{\vartheta} = \frac{35}{106} = 33\%.$$

Ez nincs messze a relatív gyakoriságtól (32%), viszont 0 írás dobás esetén is pozitív becslést kapunk, amiben benne van a mintaelemek száma.

Bayes-becslés: példa

A becslésünk **függ az a priori eloszlástól**. Ha $\text{Beta}(\alpha, \beta)$ a priori eloszlást használunk, akkor

$$\hat{\vartheta} = \int_{\Theta} \vartheta \cdot p^*(\vartheta | \underline{X} = \underline{x}) d\vartheta = \int_0^1 \vartheta \cdot p^*(\vartheta | \underline{X} = \underline{x}) d\vartheta = \frac{k + \alpha}{n + \alpha + \beta}.$$

Ez más α, β -ra tipikusan különböző eredményt ad.

Bayes-becslés: példa

A becslésünk **függ az a priori eloszlástól**. Ha $\text{Beta}(\alpha, \beta)$ a priori eloszlást használunk, akkor

$$\hat{\vartheta} = \int_{\Theta} \vartheta \cdot p^*(\vartheta | \underline{X} = \underline{x}) d\vartheta = \int_0^1 \vartheta \cdot p^*(\vartheta | \underline{X} = \underline{x}) d\vartheta = \frac{k + \alpha}{n + \alpha + \beta}.$$

Ez más α, β -ra tipikusan különböző eredményt ad.

Példa arra, amikor téves előzetes feltevés rossz eredményre vezet. Tegyük fel, hogy

- a priori eloszlás az írás valószínűségére: $\text{Beta}(2000, 200)$
- ennek várható értéke $2000/2200 = 0,91$, tehát előzetesen sokkal valószínűbbnek tekintjük az írás dobást, mint a fejet
- tegyük fel, hogy $n = 100$ dobásból egy írás sem volt

Bayes-becslés: példa

A becslésünk **függ az a priori eloszlástól**. Ha $\text{Beta}(\alpha, \beta)$ a priori eloszlást használunk, akkor

$$\hat{\vartheta} = \int_{\Theta} \vartheta \cdot p^*(\vartheta | \underline{X} = \underline{x}) d\vartheta = \int_0^1 \vartheta \cdot p^*(\vartheta | \underline{X} = \underline{x}) d\vartheta = \frac{k + \alpha}{n + \alpha + \beta}.$$

Ez más α, β -ra tipikusan különböző eredményt ad.

Példa arra, amikor téves előzetes feltevés rossz eredményre vezet. Tegyük fel, hogy

- a priori eloszlás az írás valószínűségére: $\text{Beta}(2000, 200)$
- ennek várható értéke $2000/2200 = 0,91$, tehát előzetesen sokkal valószínűbbnek tekintjük az írás dobást, mint a fejet
- tegyük fel, hogy $n = 100$ dobásból egy írás sem volt
- $k = 0$, és így

$$\hat{\vartheta} = \frac{2000}{100 + 2000 + 200} = 0,87.$$

Vagyis a 100 dobás nem volt elég arra, hogy az előzetes feltételezésünk hibáját korigálja.

Bayes-becslés a normális eloszlásra

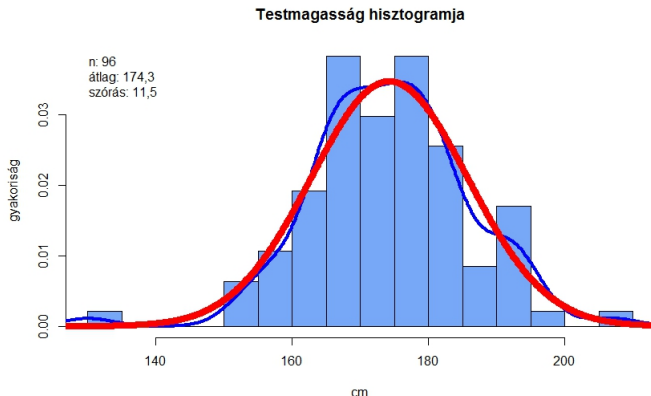
- tegyük fel, hogy az emberek testmagassága normális eloszlású, várható értéke m ismeretlen paraméter, szórása $s = 10$
- az m paraméterről tegyük fel, hogy az a priori eloszlása normális, várható értéke μ , szórása σ
- célunk, hogy meghatározzuk az m paraméter Bayes-becslését egy valós mintából (ez μ -nek és σ -nak egy függvénye lesz).
- az eloszlás sűrűségfüggvénye, ha a paraméter m :

$$f_m(x) = \frac{1}{\sqrt{2\pi} \cdot 10} \exp\left(-\frac{(x - m)^2}{200}\right).$$

- a priori sűrűségfüggvény az m -re:

$$p(m) = \frac{1}{\sqrt{2\pi} \cdot \sigma} \exp\left(-\frac{(m - \mu)^2}{2\sigma^2}\right).$$

Testmagasság hisztogramja



A testmagasság hisztogramja $n = 96$ elemű mintából és az $\bar{X} = 174,3$ várható értékű és $s_n^* = 11,5$ szórású normális eloszlás sűrűségfüggvénye (pirossal).

A posteriori eloszlás a normális eloszlásra

Az a posteriori eloszlás sűrűségfüggvénye (itt L_m a likelihood-függvény, f_p a prediktív eloszlás sűrűségfüggvénye, és az m -től nem függő részeket nem számoljuk ki)

$$\begin{aligned} p^*(m|X = \underline{x}) &= \frac{L_m(x_1, \dots, x_n)p(m)}{f_p(x_1, \dots, x_n)} = \\ &= \frac{1}{f_p} \left(\frac{1}{\sqrt{2\pi}10} \right)^n \exp \left(- \frac{\sum_{j=1}^n (x_j - m)^2}{200} \right) \cdot \frac{1}{\sqrt{2\pi}\sigma} \exp \left(- \frac{(m - \mu)^2}{2\sigma^2} \right) = \\ &= C_{x_1, \dots, x_n} \exp \left(- \frac{\sum_{j=1}^n (x_j - m)^2 \cdot \sigma^2 + (m - \mu)^2 \cdot 100}{200 \cdot \sigma^2} \right) = \\ &= C'_{x_1, \dots, x_n} \exp \left(- \frac{m^2(\sigma^2 n + 100) - 2m(\sum_{j=1}^n x_j \sigma^2 + 100\mu)}{200 \cdot \sigma^2} \right) = \\ &= C'_{x_1, \dots, x_n} \exp \left(- \frac{m^2 - 2m \frac{\sum_{j=1}^n x_j \sigma^2 + 100\mu}{\sigma^2 n + 100}}{200 \cdot \sigma^2 \cdot (\sigma^2 n + 100)} \right) = \\ &= C''_{x_1, \dots, x_n} \exp \left(- \frac{\left(m - \frac{\sum_{j=1}^n x_j \sigma^2 + 100\mu}{\sigma^2 n + 100} \right)^2}{200 \cdot \sigma^2 \cdot (\sigma^2 n + 100)} \right) \end{aligned}$$

Bayes-becslés a normális eloszlásra

Az a posteriori eloszlás egy olyan normális eloszlás, melynek várható értéke

$$\frac{\sum_{j=1}^n x_j \sigma^2 + 100\mu}{\sigma^2 n + 100},$$

szórása

$$10\sigma \sqrt{\sigma^2 n + 100}.$$

Bayes-becslés a normális eloszlásra

Az a posteriori eloszlás egy olyan normális eloszlás, melynek várható értéke

$$\frac{\sum_{j=1}^n x_j \sigma^2 + 100\mu}{\sigma^2 n + 100},$$

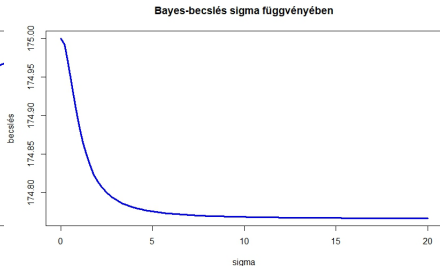
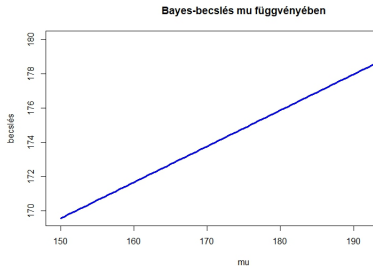
szórása

$$10\sigma \sqrt{\sigma^2 n + 100}.$$

Négyzetes veszteségfüggvény esetén a Bayes-becslés az a posteriori eloszlás várható értéke, vagyis (az x_1, \dots, x_n megfigyelt értékek helyére a mintát visszaírva)

$$\hat{m} = \frac{\sum_{j=1}^n X_j \sigma^2 + 100\mu}{\sigma^2 n + 100} = \frac{\sum_{j=1}^n X_j + 100 \frac{\mu}{\sigma^2}}{n + \frac{100}{\sigma^2}}.$$

Bayes-becslés: példa



Bayes-becslés a várható értékre μ és σ függvényében, ahol az a priori eloszlás $N(\mu, 2^2)$ az első esetben és $N(175, \sigma^2)$ a második esetben

Bayes-becslés a normális eloszlás várható értékére

- Az adatsor $n = 94$ ember testmagasságából állt, mintaátlag 174,77 cm.
- Az első ábrán azt láthatjuk, hogy hogyan változik \hat{m} , azaz a várható érték becslése, ha az a priori eloszlás paramétereit, a μ várható értéket és a σ szórást változtatjuk.

Bayes-becslés a normális eloszlás várható értékére

- Az adatsor $n = 94$ ember testmagasságából állt, mintaátlag $174,77$ cm.
- Az első ábrán azt láthatjuk, hogy hogyan változik \hat{m} , azaz a várható érték becslése, ha az a priori eloszlás paramétereit, a μ várható értéket és a σ szórást változtatjuk.
- A becslés az m -nek pozitív együtthatós lineáris függvénye. Vagyis minél nagyobb μ , vagyis előzetesen minél nagyobboknak feltételezzük az m -et, annál nagyobb lesz az m becslése is.

Bayes-becslés a normális eloszlás várható értékére

- Az adatsor $n = 94$ ember testmagasságából állt, mintaátlag $174,77$ cm.
- Az első ábrán azt láthatjuk, hogy hogyan változik \hat{m} , azaz a várható érték becslése, ha az a priori eloszlás paramétereit, a μ várható értéket és a σ szórást változtatjuk.
- A becslés az m -nek pozitív együtthatós lineáris függvénye. Vagyis minél nagyobb μ , vagyis előzetesen minél nagyobbban feltételezzük az m -et, annál nagyobb lesz az m becslése is.
- A szórást változtatva, ha $\sigma \rightarrow 0$, akkor a tört első alakjából látszik, hogy \hat{m} limesze μ lesz, vagyis ilyenkor a becslés a minta értékeitől függetlenül az a priori várható érték lesz lényegében. Ez figyelhető meg az ábrán is. Ha $\sigma \rightarrow \infty$, akkor pedig a tört második alakjából látható, hogy a limesz az átlag lesz (a számlálóban és a nevezőben is a második tag határértéke 0 , és csak az összeg, illetve a mintaelemszám marad).
- A frekventista hozzáállással a becslés (például a maximumlikelihood-becslés) az átlag lett volna. Ezt tehát $\sigma \rightarrow \infty$ esetén kapnánk meg, vagy akkor, ha ugyan μ és σ rögzítettek, de a mintaelemszám végtelenhez tart.

Házi feladat március 25., kedd, 10:15-ig

Adjunk Bayes-becslést a Poisson-eloszlás paraméterére az X_1, X_2, \dots, X_n , ha az a priori eloszlás exponenciális eloszlás $\lambda = 1$ paraméterrel.