Matematikai statisztika előadás, 13. hét, május 13.

Bayes-becslések (kiegészítő anyag)

1. A frekventista és bayes-i hozzáállás összehasonlítása

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.

Ennek a módszernek egy hátránya például:

- egy szabályosnak tűnő érmével n dobásból mindegyik fej lett;
- legyen p az írás valószínűsége;
- ezt a relatív gyakorisággal becsülhetjük: $\hat{p} = 0$. Ez torzítatlan, konzisztens becslés, a frekventista hozzáállás követelményeinek megfelel.
- Ugyanakkor, ha $n = 2$ dobásból lett mindegyik fej, az egész mást jelent, mint ha $n = 200$ dobásból lett mindegyik fej. Ezt a becslés nem tükrözi.
- Továbbá nem használtuk ki azt az információt, hogy az érme szabályosnak tűnik.

Bayes-i hozzáállás: a paramétert magát is valószínűségi változónak tekintjük, ebbe beépítve valamilyen előzetes információt. Például a pénzérménél, mivel szabályosnak tűnik, azt tesszük fel, hogy nagy valószínűséggel az $1/2$-hez közel van az értéke. Például feltesszük, hogy az eloszlása beta-eloszlású a $[0,1]$ intervallumon, melynek $1/2$ a várható értéke, vagyis „nagy valószínűséggel” közel van az $1/2$-hez. A beta-eloszlást azért használjuk, mert a p paraméter értéke biztosan 0 és 1 között van, tehát olyan eloszlás kell, amiből egy valószínűségi változót sorsolva a kapott szám biztosan 0 és 1 között lesz.

1.1. Definíció. *Az X valószínűségi változó beta-eloszlású a, b paraméterekkel, ha sűrűségfüggvénye:*

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

és nulla különben.

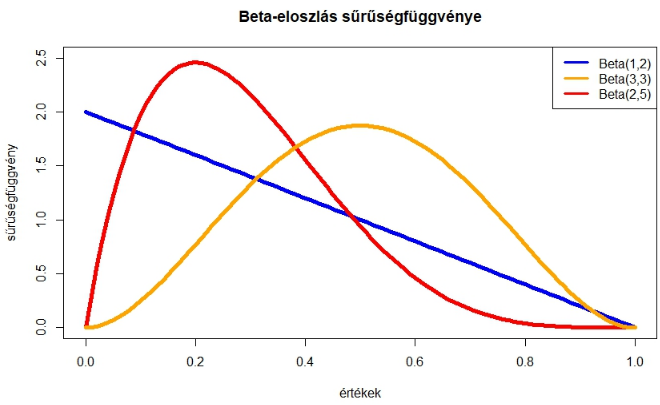
1.1. Állítás. *Az (a, b) paraméterű beta-eloszlás várható értéke és szórása:*

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

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

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

Az 1. ábrán látható néhány beta-eloszlás sűrűségfüggvénye. Ha azt mondjuk, hogy az írás dobásának valószínűsége p, és az érme szabályosnak látszik, akkor ennek például az felelhet meg, hogy a p-nek, mint a bayes-i hozzáállás szerint egy valószínűségi változónak, az eloszlása $\mathrm{Beta}(3,3)$. Az ábra mutatja ugyanis, hogy ilyen paraméterek mellett a beta-eloszlás legvalószínűbb értékei az $1/2$ körül vannak. Vagyis ezzel fejezhetjük ki a paraméterről az előzetes információt.



1. ábra. Beta-eloszlás sűrűségfüggvénye különböző paraméterpárok mellett

1.1. A priori és a posteriori eloszlás

A Bayes-becslés elkészítéséhez az alábbi fogalmakra lesz szükség.

Θ az ismeretlen ϑ paraméter összes lehetséges értékéből álló halmaz.

- **a priori eloszlás**: eloszlás a Θ paramétertéren, sűrűségfüggvénye legyen π. Ez tartalmazza a paraméterről az előzetes információt, feltevést. Ez jelenti azt, hogy a θ paramétert valószínűségi változónak tekintjük, melynek sűrűségfüggvénye π. A példában az a priori eloszlás a Beta$(3, 3)$-eloszlás, így π ennek sűrűségfüggvénye lesz. Az a priori eloszlás választása a ϑ paramétertéren értelmezett eloszlások közül tetszőleges lehet (amennyiben tükrözi a paraméterről szerzett előzetes információkat), és a becslés függni fog ettől az előzetes választástól.

- **prediktív eloszlás**: a minta eloszlása az a priori eloszlás alapján, feltétel nélkül. Sűrűségfüggvénye a teljes valószínűség tételének egy folytonos analógiájával:

$$f_\pi(x) = \int_\Theta f_\vartheta(x)\pi(\vartheta)d\vartheta,$$

  ahol f_ϑ a minta sűrűségfüggvénye a ϑ paraméter mellett (vagyis a ha paraméter ϑ, akkor erre feltételesen f_ϑ a sűrűségfüggvény.

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

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

$$\pi^*(\vartheta|\underline{X} = \underline{x}) = \frac{L_\vartheta(x_1, \ldots, x_n)\pi(\vartheta)}{L_\pi(x_1, \ldots, x_n)},$$

  ahol L_ϑ a likelihood-függvény ϑ mellett, $\underline{X}$ a minta, $\underline{x} = (x_1, \ldots, x_n)$ a megfigyelt értékek, L_π pedig az f_π-ből számolt likelihood-függvény: $L_\pi(x_1, \ldots, x_n) = \prod_{j=1}^n f_\pi(x_j)$.

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 Beta$(3, 3)$ a $\Theta = [0, 1]$ paramétertéren. Sűrűségfüggvénye az 1.1. definíció alapján:

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

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

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

hiszen π éppen a Beta$(3,3)$ eloszlás sűrűségfüggvénye, így az integrál ennek az eloszlásnak a várható értéke lesz.

A posteriori eloszlás: feltéve, hogy n dobásból k írás lett, mi a ϑ paraméter eloszlása a $[0,1]$ intervallumon? Használjuk az a posteriori eloszlás definícióját, azt, hogy diszkrét esetben a likelihood-függvény a valószínűség. Ezután azt használhatjuk, hogy ha a paraméter ϑ rögzített, akkor az írás dobások száma binomiális eloszlású, ez adja meg $\mathbb{P}_\vartheta(Y=k)$-t. A $\pi(\vartheta)$ a beta-eloszlás sűrűségfüggvénye az 1.1. definíció alapján. A nevezőben az a kérdés, hogy a prediktív eloszlás szerint mennyi annak a valószínűsége, hogy n dobásból k írás lesz. Már kiszámoltuk, hogy ebben az esetben $1/2$ az írás valószínűsége, így az n rendű, $1/2$ paraméterű binomiális eloszlással számolhatunk. Végül a törtet egyszerűsítjük és átcsoportosíthatunk:

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

Azt vehetjük észre, hogy ez is Beta-eloszlás az 1.1. definíció alapján, a paraméterei $a = k+3$ és $b = n-k+3$. Ez lesz tehát az a posteriori eloszlás: a megfigyelések alapján így változik a ϑ mint valószínűségi változó sűrűségfüggvénye. Azt is észrevehetjük, hogy ha $k < n/2$, vagyis a dobásoknak kevesebb, mint a fele lett írás, akkor az a posteriori eloszlás (mint az 1. ábrán látható Beta$(2,5)$ eloszlás, hiszen ekkor $k+2 < n-k+2$), az $1/2$-nél kisebb értékeknek ad nagyobb valószínűséget, míg ha $k > n/2$, vagyis írásból volt több, akkor a p-nek az $1/2$-nél nagyobb értékei valószínűbbek. Az a priori eloszlás szimmetrikus volt az $1/2$-re, ehhez képest változott a p-ről való feltételezett eloszlás a megfelelő irányba.

1.2. Veszteségfüggvény és Bayes-becslés

A Bayes-becslésnél is meg kell választani, hogy milyen szempont szerint legyen optimális a becsült érték. Ehhez a veszteségfüggvény fogalmát használjuk, és az a cél, hogy a veszteség várható értéke a lehető legkisebb legyen.

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 $|\vartheta-\hat{\vartheta}|$.

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

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

a priori bayesi rizikó minimális.

Itt $X_1, X_2, \dots, X_n$ a minta, és $T(X_1,\dots,X_n)$ az a statisztika, amit ebből kiszámolunk, ez lesz a becslés. A $W(T(X_1,\dots,X_n),\vartheta)$ érték tehát a becslés hibájából adódó veszteség, amit a veszteségfüggvény mond meg, ha a valódi ϑ-t a $T(X_1,\dots,X_n)$-nel becsüljük. Ezután vesszük ennek a várható értékét: az $X_1, X_2, \dots, X_n$ a minta, így ezek valószínűségi változók, a $W(T(X_1,\dots,X_n),\vartheta)$ is valószínűségi változó, rögzített ϑ mellett. Azonban a Bayes-becslés hozzáállása szerint ϑ maga is valószínűségi változó, így végül eszerint is várható értéket számolunk. A cél az, hogy a becslésből adódó veszteség ilyen módon kapott várható értéke a lehető legkisebb legyen.

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

1.1. 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:*

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

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 Beta(3, 3)-eloszlás a $\Theta = [0, 1]$ paramétertéren, mint eddig. Sűrűségfüggvénye $\pi(x) = 30x^2(1 - x)^2\mathbb{I}(0 \leq x \leq 1)$. 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:

$$\hat{\vartheta} = \int_\Theta \vartheta \cdot \pi^*(\vartheta | \underline{X} = \underline{x})d\vartheta = \int_0^1 \vartheta \cdot \pi^*(\vartheta | \underline{X} = \underline{x})d\vartheta = \frac{k + 3}{(k + 3) + (n - k + 3)} = \frac{k + 3}{n + 6},$$

hiszen éppen az $a = k + 3$ és $b = n - k + 3$ paraméterű beta-eloszlás várható értéke jelent meg, amire pedig használhatjuk az 1.1. állítást.

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

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

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

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

Itt ϑ az írás dobás valószínűsége volt. Erről eredetileg azt tételeztük fel, hogy 1/2 várható értékű Beta(3, 3)-eloszlású. Mivel nem voltak írás dobások, az új becslés kevesebb, mint 1/2, és hogy mennyivel kevesebb, abban az is számít, hogy hányelemű volt a minta. Frekventista hozzáállással mindkét esetben 0 lett volna a becslés, nem láttuk volna a mintaelemszámból adódó különbséget.

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

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

Vagyis ezzel is kevesebb, mint 1/2-nek becsüljük az írás valószínűségét, annak megfelelően, hogy kevesebb írás volt, mint fej. Ez nincs messze a relatív gyakoriságtól sem ebben az esetben.

Vegyük észre, hogy a becslésünk függ az a priori eloszlástól. Ha ugyanezzel a számolással Beta(α, β) a priori eloszlást használunk, akkor

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

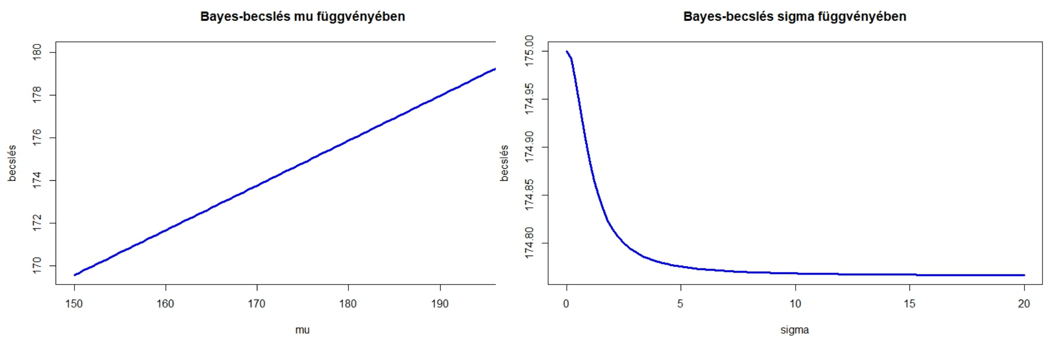
Ez más α, β-ra tipikusan különböző eredményt ad. Ebből is látható, hogy a Bayes-becslés függ az a priori eloszlástól.

1.3. 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:

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



2. ábra. 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

A priori sűrűségfüggvény az m-re:

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

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

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

Tehát 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 $x_1, \ldots, x_n$ megfigyelt értékek helyére a mintát visszaírva)

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

A 2. 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. Az adatsor $n = 94$ ember testmagasságából állt. A mintaátlag 174,77 cm.

Az ábra alapján és a képletből is látható, hogy a becslés az m-nek pozitív együtthatós lineáris függvénye. Vagyis minél nagyobb μ, vagyis előzetesen minél nagyobbnak feltételezzük az m-et (ennek a pripri eloszlása volt $N(\mu, \sigma^2)$), annál nagyobb lesz az m becslése is.

A szórást változtatva, ha $\sigma \to 0$, 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 $\sigma \to \infty$, 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 $\sigma \to \infty$ esetén kapnánk meg, vagy akkor, ha ugyan μ és σ rögzítettek, de a mintaelemszám végtelenhez tart, ilyenkor szintén a számlálóban és a nevezőben is a második tag nem számít a limeszben (kivéve, ha X_j várható értéke 0).