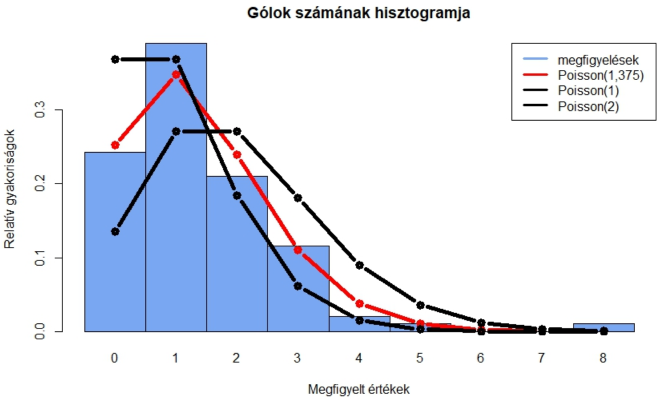
Poisson-eloszlás paraméterének becslése (3. előadás)



A gólok számának hisztogramja $n = 95$ mérkőzésen, és különböző paraméterű Poisson-eloszlások ($\mathbb{P}_\lambda(X = k) = \lambda^k/k! \cdot e^{-\lambda}$). $\lambda = 1,375$ a gólok átlagos száma.

Statisztikai mező

Definíció

Az $(\Omega, \mathcal{A}, \mathcal{P})$ *hármast* **statisztikai mezőnek** *nevezzük, ha minden* $\mathbb{P} \in \mathcal{P}$*-re* $(\Omega, \mathcal{A}, \mathbb{P})$ *Kolmogorov-féle valószínűségi mező.*

Paraméteres statisztikai mező: $\mathcal{P} = \{\mathbb{P}_\vartheta : \vartheta \in \Theta\}$. Ekkor ϑ az ismeretlen paraméter, mely egy $\Theta \subseteq \mathbb{R}^q$ ismert halmaz eleme.

Például: $\mathcal{P}$ lehet például

- a λ paraméterű Poisson-eloszlások halmaza;
- a normális eloszlások halmaza (ekkor $\vartheta = (m, \sigma)$ az ismeretlen paraméter);
- az $[a, b]$ intervallumon egyenletes eloszlások halmaza (ekkor $\vartheta = (a, b)$ az ismeretlen paraméter).

Minta és statiszika

Definíció (Minta)

Legyen $(\Omega, \mathcal{A}, \mathcal{P})$ *statisztikai mező. Egy*

$$\underline{X} = (X_1, X_2, \ldots, X_n) : \Omega \to B \subseteq \mathbb{R}^n$$

valószínűségi vektorváltozót (n elemű) **mintának** *nevezünk. Itt B a mintatér, n a minta elemszáma vagy nagysága. A minta független, ha az* $X_1, X_2, \ldots, X_n$ *valószínűségi változók függetlenek.*

Minta és statiszika

Definíció (Minta)

Legyen $(\Omega, \mathcal{A}, \mathcal{P})$ *statisztikai mező. Egy*

$$\underline{X} = (X_1, X_2, \ldots, X_n) : \Omega \to B \subseteq \mathbb{R}^n$$

valószínűségi vektorváltozót (n elemű) **mintának** *nevezünk. Itt B a mintatér, n a minta elemszáma vagy nagysága. A minta független, ha az* $X_1, X_2, \ldots, X_n$ *valószínűségi változók függetlenek.*

Definíció (Statisztika)

Legyen $T : B \to \mathbb{R}^k$ *függvény. Ekkor a* $T(X_1, X_2, \ldots, X_n)$ *valószínűségi változót statisztikának nevezzük.*

Például: $T(X_1, \ldots, X_n) = \overline{X}$ a mintaátlag, vagy $T(X_1, \ldots, X_n) = s_n^*$ a korrigált tapasztalati szórásnégyzet.

Egyenletes eloszlás paraméterének becslése

Legyen $X_1, X_2, \ldots, X_n$ független azonos eloszlású minta az $\{1, 2, \ldots, N\}$ halmazon egyenletes eloszlásból, ahol N ismeretlen paraméter https://en.wikipedia.org/wiki/German_tank_problem.

Például: $52, 38, 14, 69, 44, 28$.

Átlag: $40,83$

Hogyan becsülnénk ez alapján N-t?

Egyenletes eloszlás paraméterének becslése

Legyen $X_1, X_2, \ldots, X_n$ független azonos eloszlású minta az $\{1, 2, \ldots, N\}$ halmazon egyenletes eloszlásból, ahol N ismeretlen paraméter `https://en.wikipedia.org/wiki/German_tank_problem`.

Például: $52, 38, 14, 69, 44, 28$.

Átlag: $40,83$

Hogyan becsülnénk ez alapján N-t?

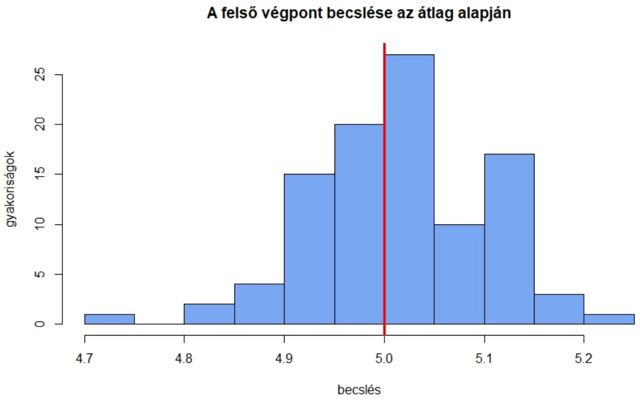
Folytonos változat: legyen $X_1, X_2, \ldots, X_n$ független azonos eloszlású minta a $[0, \vartheta]$ intervallumon egyenletes eloszlásból, ahol $\vartheta > 0$ ismeretlen paraméter.

Minta: $X_1, X_2, \ldots, X_5$, például $4,3; \quad 5,6; \quad 3,2; \quad 1,8; \quad 2,2$

Átlag: $\overline{X} = 3,42$

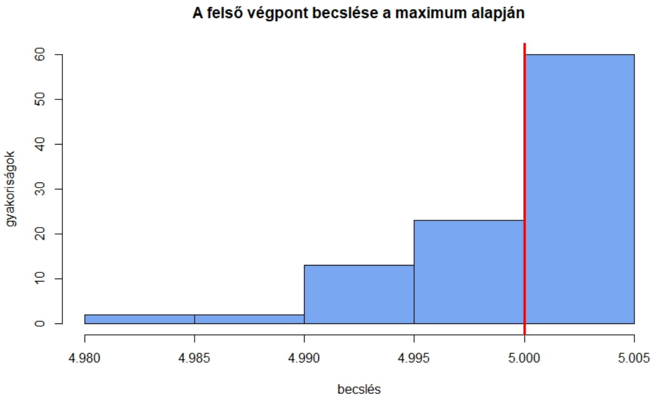
Hogyan becsülnénk ez alapján ϑ-t?

Egyenletes eloszlás paraméterének becslése



A $[0, \vartheta]$ intervallumon egyenletes eloszlás paraméterének becslése $T(X) = 2\overline{X}$-szel, $n = 1000$ elemű mintából. Az igazi paraméter $\vartheta = 5$. Száz ismétlésből a becslések átlaga: $5,015$, korrigált tapasztalati szórása: $0,086$.

Egyenletes eloszlás paraméterének becslése



A $[0, \vartheta]$ intervallumon egyenletes eloszlás paraméterének becslése a maximum alapján $T(X) = \frac{n+1}{n} \max(X_1, \ldots, X_n)$-nel, $n = 1000$ elemű mintából. Az igazi paraméter $\vartheta = 5$. Száz ismétlésből a becslések átlaga: $4,9999$, korrigált tapasztalati szórása: $0,0049 < 0,086$.

Egyenletes eloszlás paraméterének becslése

$X_1, X_2, \ldots, X_n$ független azonos eloszlású, a $[0, \vartheta]$ intervallumon egyenletes eloszlású valószínűségi változók. Itt $\vartheta > 0$ ismeretlen paraméter. Tekintsük a ϑ két különböző becslését:

$$T_1(X_1, \ldots, X_n) = 2\overline{X}; \qquad \mathbb{E}_\vartheta(T_1) = 2 \cdot \frac{\vartheta}{2} = \vartheta; \qquad D_\vartheta(T_1) = \frac{\vartheta}{\sqrt{3n}}.$$

Másrészt, felhasználva, hogy $X_n^* = \max(X_1, \ldots, X_n)$ sűrűségfüggvénye: $f_\vartheta(t) = (nt^{n-1}/\vartheta^n)\mathbb{I}(0 \leq t \leq \vartheta)$, egy másik becslést is találhatunk:

$$T_2(X_1, \ldots, X_n) = \frac{n+1}{n} \cdot X_n^*; \qquad \mathbb{E}_\vartheta(T_2) = \frac{n+1}{n} \cdot \frac{n\vartheta}{n+1} = \vartheta.$$

$$D_\vartheta(T_2) = \sqrt{\frac{n \cdot \vartheta^2}{(n+2)(n+1)^2}} \leq \frac{\vartheta}{n+1} < \frac{\vartheta}{\sqrt{3n}}.$$

Egyenletes eloszlás paraméterének becslése

$X_1, X_2, \ldots, X_n$ független azonos eloszlású, a $[0, \vartheta]$ intervallumon egyenletes eloszlású valószínűségi változók. Itt $\vartheta > 0$ ismeretlen paraméter. Tekintsük a ϑ két különböző becslését:

$$T_1(X_1, \ldots, X_n) = 2\overline{X}; \qquad \mathbb{E}_\vartheta(T_1) = 2 \cdot \frac{\vartheta}{2} = \vartheta; \qquad D_\vartheta(T_1) = \frac{\vartheta}{\sqrt{3n}}.$$

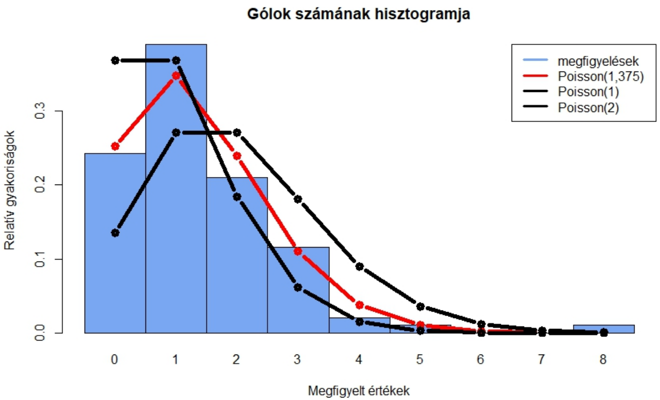
Másrészt, felhasználva, hogy $X_n^* = \max(X_1, \ldots, X_n)$ sűrűségfüggvénye: $f_\vartheta(t) = (nt^{n-1}/\vartheta^n)\mathbb{I}(0 \leq t \leq \vartheta)$, egy másik becslést is találhatunk:

$$T_2(X_1, \ldots, X_n) = \frac{n+1}{n} \cdot X_n^*; \qquad \mathbb{E}_\vartheta(T_2) = \frac{n+1}{n} \cdot \frac{n\vartheta}{n+1} = \vartheta.$$

$$D_\vartheta(T_2) = \sqrt{\frac{n \cdot \vartheta^2}{(n+2)(n+1)^2}} \leq \frac{\vartheta}{n+1} < \frac{\vartheta}{\sqrt{3n}}.$$

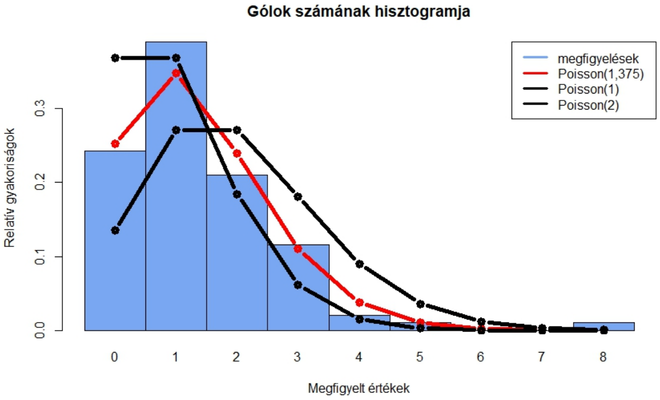
Mindkét becslés **torzítatlan** ϑ-ra, és a második **hatásosabb.**

Poisson-eloszlás paraméterének becslése



A gólok számának hisztogramja $n = 95$ mérkőzésen, és különböző paraméterű Poisson-eloszlások. Ekkor

Poisson-eloszlás paraméterének becslése



A gólok számának hisztogramja $n = 95$ mérkőzésen, és különböző paraméterű Poisson-eloszlások. Ekkor

$$\mathbb{E}(\overline{X}) = \mathbb{E}(X_1) = \lambda \qquad \text{minden } \lambda\text{-ra.}$$

Ha az átlaggal becsüljük a paramétert, akkor a becslésünk várható értéke λ.

Becslések és tulajdonságaik

- $(\Omega, \mathcal{A}, \mathcal{P})$ statisztikai mező;
- $\mathcal{P} = \{\mathbb{P}_\vartheta : \vartheta \in \Theta)$ valamely Θ halmazzal (Θ a paramétertér);
- $g : \Theta \to \mathbb{R}$ függvény.
- Cél: olyan T statisztika keresése, amire a $T(X)$ valószínűségi változó és a $g(\vartheta)$ érték valamilyen értelemben közel esnek egymáshoz.

Becslések és tulajdonságaik

- $(\Omega, \mathcal{A}, \mathcal{P})$ statisztikai mező;
- $\mathcal{P} = \{\mathbb{P}_\vartheta : \vartheta \in \Theta)$ valamely Θ halmazzal (Θ a paramétertér);
- $g : \Theta \to \mathbb{R}$ függvény.
- Cél: olyan T statisztika keresése, amire a $T(X)$ valószínűségi változó és a $g(\vartheta)$ érték valamilyen értelemben közel esnek egymáshoz.

Definíció (Torzítatlanság)

A T statisztika torzítatlan becslés $g(\vartheta)$-ra, ha minden $\vartheta \in \Theta$-ra

$$\mathbb{E}_\vartheta(T(X_1, \ldots, X_n)) = g(\vartheta).$$

A T statisztika torzítása a $b_T(\vartheta) = \mathbb{E}_\vartheta(T(X_1, \ldots, X_n)) - g(\vartheta)$ függvény.

Példa. $X_1, X_2, \ldots, X_n$ független minta a $[0, \vartheta]$ intervallumon egyenletes eloszlásból.

Becslések és tulajdonságaik

- $(\Omega, \mathcal{A}, \mathcal{P})$ statisztikai mező;
- $\mathcal{P} = \{\mathbb{P}_\vartheta : \vartheta \in \Theta)$ valamely Θ halmazzal (Θ a paramétertér);
- $g : \Theta \to \mathbb{R}$ függvény.
- Cél: olyan T statisztika keresése, amire a $T(X)$ valószínűségi változó és a $g(\vartheta)$ érték valamilyen értelemben közel esnek egymáshoz.

Definíció (Torzítatlanság)

A T statisztika torzítatlan becslés $g(\vartheta)$-ra, ha minden $\vartheta \in \Theta$-ra

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

A T statisztika torzítása a $b_T(\vartheta) = \mathbb{E}_\vartheta(T(X_1, \dots, X_n)) - g(\vartheta)$ függvény.

Példa. $X_1, X_2, \dots, X_n$ független minta a $[0, \vartheta]$ intervallumon egyenletes eloszlásból. Ekkor $2\overline{X}$ torzítatlan becslés $g(\vartheta) = \vartheta$-ra: $\mathbb{E}(2\overline{X}) = \vartheta$.

Torzítatlan becslések

Állítás (A várható érték torzítatlan becslése)

Legyen $X_1, \ldots, X_n$ *független azonos eloszlású véges várható értékű minta. Ekkor*

$$\mathbb{E}_\vartheta(\overline{X}) = \mathbb{E}_\vartheta(X_1) \qquad \textit{minden } \vartheta \in \Theta\textit{-ra},$$

vagyis a **mintaátlag** *torzítatlan becslés* ψ*-re.*

Állítás (A szórásnégyzet torzítatlan becslése)

$X_1, \ldots, X_n$ *független azonos eloszlású véges szórású minta. Ekkor Ekkor*

$$\mathbb{E}_\vartheta(s_n^{*2}) = D_\vartheta^2(X_1) \qquad \textit{minden } \vartheta \in \Theta\textit{-ra},$$

vagyis a **korrigált tapasztalati szórásnégyzet** *torzítatlan becslés a szórásnégyzet-re.*

Az átlag várható értéke

Állítás

Legyen $X_1, \ldots, X_n$ *független azonos eloszlású minta, és* $m = \mathbb{E}(X_i) < \infty$*. Ekkor*

$$\mathbb{E}(\overline{X}) = m.$$

Az átlag várható értéke

Állítás

Legyen $X_1, \ldots, X_n$ független azonos eloszlású minta, és $m = \mathbb{E}(X_i) < \infty$. Ekkor

$$\mathbb{E}(\overline{X}) = m.$$

Bizonyítás.

$$\mathbb{E}(\overline{X}) = \mathbb{E}\left(\frac{X_1 + \ldots + X_n}{n}\right) = \frac{1}{n}\mathbb{E}(X_1 + \ldots + X_n) = \frac{1}{n} \cdot nm = m.$$

Felhasználtuk a várható érték linearitását, és hogy csak eloszlástól függ:

- $\mathbb{E}(cX) = c\mathbb{E}(X)$, ha $c \in \mathbb{R}$;
- $\mathbb{E}(Y + Z) = \mathbb{E}(Y) + \mathbb{E}(Z)$;
- ha Y és Z eloszlása megegyezik, akkor $\mathbb{E}(Y) = \mathbb{E}(Z)$

Tehát a **mintaátlag** torzítatlan becslés a várható értékre.

Speciálisan: a **relatív gyakoriság** torzítatlan becslés egy esemény valószínűségére.

Az átlag szórása

Állítás

Legyen $X_1, \ldots, X_n$ *független azonos eloszlású minta, és* $\mathbb{E}(X_i^2) < \infty$ *(amiből következik, hogy a szórás véges). Ekkor*

$$D(\overline{X}) = \frac{D(X_1)}{\sqrt{n}}.$$

Az átlag szórása

Állítás

Legyen $X_1, \dots, X_n$ független azonos eloszlású minta, és $\mathbb{E}(X_i^2) < \infty$ (amiből következik, hogy a szórás véges). Ekkor

$$D(\overline{X}) = \frac{D(X_1)}{\sqrt{n}}.$$

Bizonyítás.

$$D(\overline{X}) = D\left(\frac{X_1 + \ldots + X_n}{n}\right) = \frac{D(X_1 + \ldots + X_n)}{n} = \frac{\sqrt{nD^2(X_1)}}{n} = \frac{D(X_1)}{\sqrt{n}}.$$

Felhasználtuk a szórás alábbi tulajdonságait:

- $D(cX) = |c|D(X)$, ha $c \in \mathbb{R}$ valós szám;
- $D(Y + Z) = \sqrt{D^2(Y) + D^2(Z)}$, ha Y és Z függetlenek;
- ha Y és Z eloszlása megegyezik, akkor $D(Y) = D(Z)$

Tapasztalati szórásnégyzet

Állítás (A tapasztalati szórásnégyzet másik alakja)

$$s_n^2 = \frac{1}{n}\left[\sum_{k=1}^{n} X_k^2\right] - \overline{X}^2.$$

Bizonyítás. Átrendezéssel kapjuk, hogy

$$\sum_{k=1}^{n}(X_k - \overline{X})^2 = \sum_{k=1}^{n}\left[X_k^2 - 2X_k \cdot \overline{X} + \overline{X}^2\right] = \left(\sum_{k=1}^{n} X_k^2\right) - 2n\overline{X} \cdot \overline{X} + n \cdot \overline{X}^2 =$$

$$= \left(\sum_{k=1}^{n} X_k^2\right) - n \cdot \overline{X}^2.$$

Ebből adódik, hogy

$$s_n^2 = \frac{1}{n}\left[\sum_{k=1}^{n}(X_k - \overline{X})^2\right] = \frac{1}{n}\left[\sum_{k=1}^{n} X_k^2\right] - \overline{X}^2,$$

a tapasztalati szórásnégyzet definíciója alapján.

Korrigált tapasztalati szórásnégyzet

$$s_n^{*2} = \frac{n}{n-1} s_n^2 = \frac{n}{n-1}\left[\frac{1}{n}\left[\sum_{k=1}^{n} X_k^2\right] - \overline{X}^2\right] = \frac{1}{n-1}\left[\sum_{k=1}^{n} X_k^2\right] - \frac{n}{n-1}\overline{X}^2.$$

Az első tag várható értéke a szórásnégyzet definíciója alapján:

Korrigált tapasztalati szórásnégyzet

$$s_n^{*2} = \frac{n}{n-1} s_n^2 = \frac{n}{n-1}\left[\frac{1}{n}\left[\sum_{k=1}^{n} X_k^2\right] - \overline{X}^2\right] = \frac{1}{n-1}\left[\sum_{k=1}^{n} X_k^2\right] - \frac{n}{n-1}\overline{X}^2.$$

Az első tag várható értéke a szórásnégyzet definíciója alapján:

$$\mathbb{E}_\vartheta\left(\sum_{k=1}^{n} X_k^2\right) = \sum_{k=1}^{n} \mathbb{E}_\vartheta(X_k^2) = n \cdot \mathbb{E}_\vartheta(X_1^2) = n \cdot \left[D_\vartheta^2(X_1) + \mathbb{E}_\vartheta(X_1)^2\right].$$

A második tag várható értéke az átlag szórásnégyzete alapján:

Korrigált tapasztalati szórásnégyzet

$$s_n^{*2} = \frac{n}{n-1} s_n^2 = \frac{n}{n-1}\left[\frac{1}{n}\left[\sum_{k=1}^{n} X_k^2\right] - \overline{X}^2\right] = \frac{1}{n-1}\left[\sum_{k=1}^{n} X_k^2\right] - \frac{n}{n-1}\overline{X}^2.$$

Az első tag várható értéke a szórásnégyzet definíciója alapján:

$$\mathbb{E}_\vartheta\left(\sum_{k=1}^{n} X_k^2\right) = \sum_{k=1}^{n} \mathbb{E}_\vartheta(X_k^2) = n \cdot \mathbb{E}_\vartheta(X_1^2) = n \cdot \left[D_\vartheta^2(X_1) + \mathbb{E}_\vartheta(X_1)^2\right].$$

A második tag várható értéke az átlag szórásnégyzete alapján:

$$\mathbb{E}_\vartheta\left(\overline{X}^2\right) = D_\vartheta^2(\overline{X}) + \mathbb{E}_\vartheta(\overline{X})^2 = \frac{1}{n} D_\vartheta^2(X_1) + \mathbb{E}_\vartheta(X_1)^2.$$

Korrigált tapasztalati szórásnégyzet

$$s_n^{*2} = \frac{n}{n-1} s_n^2 = \frac{n}{n-1}\left[\frac{1}{n}\left[\sum_{k=1}^{n} X_k^2\right] - \overline{X}^2\right] = \frac{1}{n-1}\left[\sum_{k=1}^{n} X_k^2\right] - \frac{n}{n-1}\overline{X}^2.$$

Az első tag várható értéke a szórásnégyzet definíciója alapján:

$$\mathbb{E}_\vartheta\left(\sum_{k=1}^{n} X_k^2\right) = \sum_{k=1}^{n} \mathbb{E}_\vartheta(X_k^2) = n \cdot \mathbb{E}_\vartheta(X_1^2) = n \cdot \left[D_\vartheta^2(X_1) + \mathbb{E}_\vartheta(X_1)^2\right].$$

A második tag várható értéke az átlag szórásnégyzete alapján:

$$\mathbb{E}_\vartheta\left(\overline{X}^2\right) = D_\vartheta^2(\overline{X}) + \mathbb{E}_\vartheta(\overline{X})^2 = \frac{1}{n} D_\vartheta^2(X_1) + \mathbb{E}_\vartheta(X_1)^2.$$

Vagyis valóban s_n^{*2} torzítatlan becslés a szórásnégyzetre:

$$\mathbb{E}_\vartheta(s_n^{*2}) = \frac{n}{n-1}\left[D_\vartheta^2(X_1) + \mathbb{E}_\vartheta(X_1)^2\right] - \frac{n}{n-1}\left[\frac{1}{n} D_\vartheta^2(X_1) + \mathbb{E}_\vartheta(X_1)^2\right] = D_\vartheta^2(X_1).$$

Becslések összehasonlítása

Definíció (Hatásosság)

Legyenek T_1, T_2 **torzítatlan** *becslései a paraméter* $\psi(\vartheta)$ *függvényének.* T_1 **hatásosabb** T_2*-nél, ha*

$$D^2_\vartheta(T_1) \leq D^2_\vartheta(T_2)$$

teljesül minden $\vartheta \in \Theta$*-ra.*

A T_1 *becslés* **hatásos** $\psi(\vartheta)$*-ra, ha* $\psi(\vartheta)$ *minden torzítatlan becslésénél hatásosabb (és ő maga is torzítatlan).*

Becslések összehasonlítása

Definíció (Hatásosság)

Legyenek T_1, T_2 **torzítatlan** *becslései a paraméter* $\psi(\vartheta)$ *függvényének.* T_1 **hatásosabb** T_2*-nél, ha*

$$D^2_\vartheta(T_1) \leq D^2_\vartheta(T_2)$$

teljesül minden $\vartheta \in \Theta$*-ra.*

A T_1 *becslés* **hatásos** $\psi(\vartheta)$*-ra, ha* $\psi(\vartheta)$ *minden torzítatlan becslésénél hatásosabb (és ő maga is torzítatlan).*

- Nem mindig létezik hatásos becslés, és lehetséges, hogy T_1 és T_2 közül egyik sem hatásosabb a másiknál.
- A várható értékre nézve a mintaátlag hatásosabb minden $\sum_{j=1}^n c_j X_j$ alakú becslésnél (ahol $\sum_{j=1}^n c_j = 1$).

Becslések összehasonlítása

Definíció (Hatásosság)

Legyenek T_1, T_2 **torzítatlan** *becslései a paraméter $\psi(\vartheta)$ függvényének. T_1* **hatásosabb** *T_2-nél, ha*

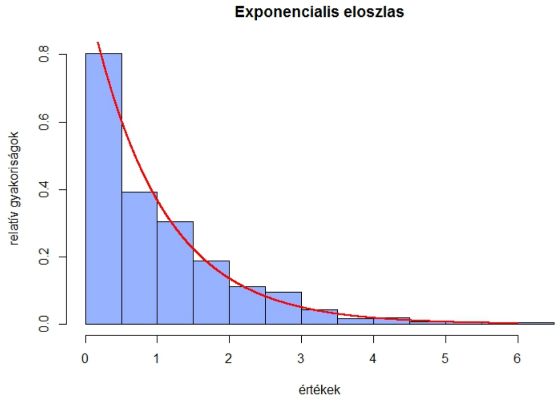
$$D^2_\vartheta(T_1) \leq D^2_\vartheta(T_2)$$

teljesül minden $\vartheta \in \Theta$-ra.

A T_1 becslés **hatásos** *$\psi(\vartheta)$-ra, ha $\psi(\vartheta)$ minden torzítatlan becslésénél hatásosabb (és ő maga is torzítatlan).*

- Nem mindig létezik hatásos becslés, és lehetséges, hogy T_1 és T_2 közül egyik sem hatásosabb a másiknál.
- A várható értékre nézve a mintaátlag hatásosabb minden $\sum_{j=1}^n c_j X_j$ alakú becslésnél (ahol $\sum_{j=1}^n c_j = 1$).
- **Bizonyos feladatokban lehet a mintaátlagnál hatásosabb becslés a várható értékre**: A $[0, b]$ intervallumon egyenletes eloszlás esetén b-re $\frac{n+1}{n}\max(X_1, \dots, X_n)$ hatásosabb a mintaátlag kétszeresénél.

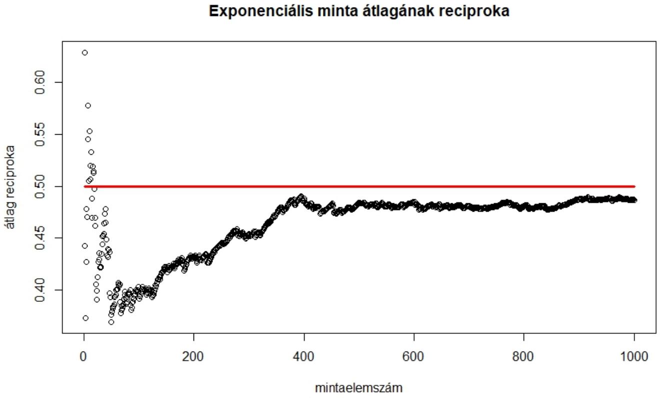
Az exponenciális eloszlású minta hisztogramja



Exponenciális eloszlású minta hisztogramja és sűrűségfüggvénye:

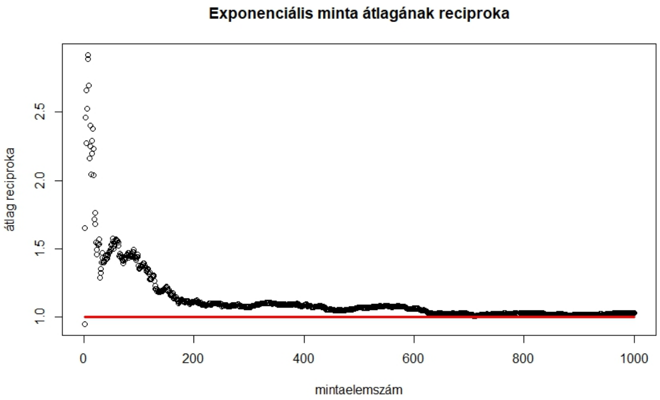
$$f(x) = \lambda \exp(-\lambda x)\mathbb{I}(x > 0); \qquad \mathbb{E}(X) = D(X) = \frac{1}{\lambda}.$$

Exponenciális eloszlás



$\lambda = 0,5$ paraméterű exponenciális eloszlást generálva a mintaátlag reciproka $0,5$-höz tart, azaz **konzisztens** becslés, hiszen ez minden λ-ra teljesül.

Exponenciális eloszlás



$\lambda = 1$ paraméterű exponenciális eloszlást generálva a mintaátlag reciproka 1-hez tart, azaz **konzisztens** becslés, hiszen ez minden λ-ra teljesül.

Konzisztencia

Definíció

A $T_n = T_n(X_1, \ldots, X_n)$ **konzisztens** *becsléssorozat* $g(\vartheta)$*-ra, ha minden* $\vartheta \in \Theta$*-ra*

$$(T_n(X_1, \ldots, X_n)) \to \psi(\vartheta)$$

$n \to \infty$ *esetén sztochasztikusan, azaz minden* $\vartheta \in \Theta$ *és* $\varepsilon > 0$*-ra teljesül, hogy*

$$\mathbb{P}_\vartheta\big(|T_n - g(\vartheta)| > \varepsilon\big) \to 0 \qquad (n \to \infty).$$

Konzisztencia

Definíció

A $T_n = T_n(X_1, \ldots, X_n)$ **konzisztens** *becsléssorozat* $g(\vartheta)$*-ra, ha minden* $\vartheta \in \Theta$*-ra*

$$(T_n(X_1, \ldots, X_n)) \to \psi(\vartheta)$$

$n \to \infty$ *esetén sztochasztikusan, azaz minden* $\vartheta \in \Theta$ *és* $\varepsilon > 0$*-ra teljesül, hogy*

$$\mathbb{P}_\vartheta\big(|T_n - g(\vartheta)| > \varepsilon\big) \to 0 \qquad (n \to \infty).$$

Elégséges feltétel:

$$\mathbb{E}_\vartheta(T_n(X)) \to g(\vartheta) \qquad \text{és} \qquad D_\vartheta(T_n(X)) \to 0$$

minden $\vartheta \in \Theta$-ra.

Példák torzítatlan, konzisztens becslésekre

$X_1, X_2, \ldots$ független azonos eloszlású minta. Ekkor

$$T_n = \frac{X_1 + X_2 + \ldots + X_n}{n} \to \mathbb{E}_\vartheta(X_1)$$

teljesül $n \to \infty$ esetén sztochasztikusan a nagy számok gyenge törvénye szerint, vagyis az **átlag** konzisztens becslés a **várható értékre**.

Speciális eset: a **relatív gyakoriság** konzisztens becslés a **valószínűségre**.

Példák torzítatlan, konzisztens becslésekre

$X_1, X_2, \ldots$ független azonos eloszlású minta. Ekkor

$$T_n = \frac{X_1 + X_2 + \ldots + X_n}{n} \to \mathbb{E}_\vartheta(X_1)$$

teljesül $n \to \infty$ esetén sztochasztikusan a nagy számok gyenge törvénye szerint, vagyis az **átlag** konzisztens becslés a **várható értékre**.

Speciális eset: a **relatív gyakoriság** konzisztens becslés a **valószínűségre**.

Nevezetes eloszlások:

- Poisson-eloszlás λ paraméterére az átlag torzítatlan, konzisztens
- a normális eloszlás m paraméterére az átlag torzítatlan és konzisztens; a σ paraméterre a tapasztalati szórás és a korrigált tapasztalati szórás konzisztensek, de nem torzítatlanok; σ^2-re s_n^{*2} torzítatlan
- exponenciális eloszlás: $1/\overline{X}$ konzisztens λ-ra, de nem torzítatlan a paraméterre
- exponenciális eloszlás: $(n+1) \cdot \min(X_1, \ldots, X_n)$ torzítatlan, de nem konzisztens a várható értékre (vagyis $1/\lambda$-ra).

Házi feladat március 5., kedd, 12:00-ig

Legyen $X_1, \ldots, X_{30}$ az $N(m, 1)$, $Y_1, \ldots, Y_{50}$ pedig az $N(m, 4)$ normális eloszlásból származó független minta. (A második paraméter a szórásnégyzetet jelöli.) Adjuk meg a legkisebb szórású $a\overline{X} + b\overline{Y}$ alakú torzítatlan becslést m-re.

Sorsoljunk (egyszer) $X_1, \ldots, X_{30}$, illetve $Y_1, \ldots, Y_{50}$ valószínűségi változókat tetszőlegesen választott m-mel, számítsuk ki a fenti becslést, és számítsuk ki a becslés abszolút hibáját (a becsült érték és az igazi m érték különbségének abszolút értékét).