

Matematikai statisztika előadás, 4. hét, március 3.

Maximumlikelihood- és momentum módszer

A matematikai statisztikában gyakran egy ismeretlen eloszlás paramétereit szeretnénk megbecsülni. Például, feltételezzük, hogy egy országban a jövedelmek eloszlása Pareto-eloszlást követ, de nem ismerjük a paramétereket. Vagy, feltételezzük, hogy az éves középhőmérséklet normális eloszlású, de nem ismerjük a várható értéket és a szórást. Az alábbiakban két általános módszert mutatunk be, amely bizonyos elvek alapján ad becslést ilyen jellegű feladatokban az ismeretlen paraméterekre.

1. Maximumlikelihood-módszer

Példa. Van két teleföntöltőnk, melyek ránézésre megkülönböztethetetlenek. A jó minden kipróbálásnál a többitől függetlenül 90%, a rossz 20% valószínűséggel működik.

Az egyik kábelt kiválasztottuk, 10 kipróbálásból pontosan 8-szor működött. Ez melyik kábel lehetett?

$$\mathbb{P}(\text{a jó kábel 8-szor működik}) = \binom{10}{8} 0,9^8 0,1^2 = \mathbf{19,8\%}.$$

$$\mathbb{P}(\text{a rossz kábel 8-szor működik}) = \binom{10}{8} 0,2^8 0,8^2 = \mathbf{0,00007\%}.$$

Inkább a jó kábel lehetett, amit kiválasztottunk.

A fenti feladat általánosítása következő, amikor már nem csak az ismeretlen paraméternek nem csak két lehetséges értéke van, hanem egy intervallumból tetszőleges értéket felvehet. Vagyis: tegyük fel, hogy van egy teleföntöltőnk, mely minden kipróbálásnál a többitől függetlenül p valószínűséggel működik.

Itt az ismeretlen paraméter: $p \in [0, 1]$.

A kábel 10 kipróbálásból pontosan 8-szor működött. Mennyi lehet p értéke?

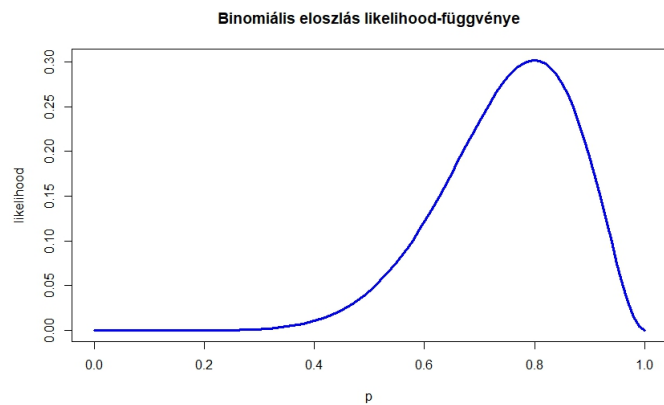
$$\mathbb{P}_p(\text{a kábel 8-szor működik}) = \binom{10}{8} p^8 (1-p)^2.$$

Kérdés: ez milyen $p \in [0, 1]$ -re lesz a legnagyobb? Hiszen az előző esetben is az alapján választottunk, hogy a megfigyelt esemény bekövetkezésének valószínűsége a paraméter mely értékénél a legnagyobb. Ez a valószínűség pedig függ p -től, ahogy láttuk, egész mást kapunk $p = 0,9$ és $p = 0,2$ esetén. Az 1. ábrán azt látjuk, hogy p -t változtatva mennyi a valószínűsége, hogy a kábel éppen 8-szor működik a tíz kipróbálásból.

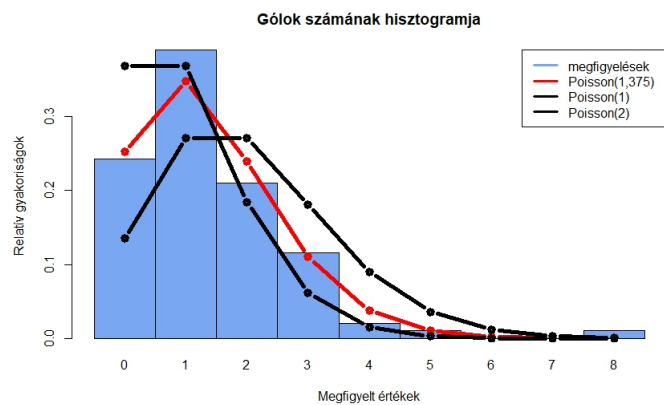
Megjegyzés. Ennek a példának egy részletesebb leírása és alkalmazása: <https://ematlap.hu/tudomany-tortenet-2020-12/992-mennyit-teszteljunk-2-v3>

1.1. A Poisson-eloszlás paraméterének becslése

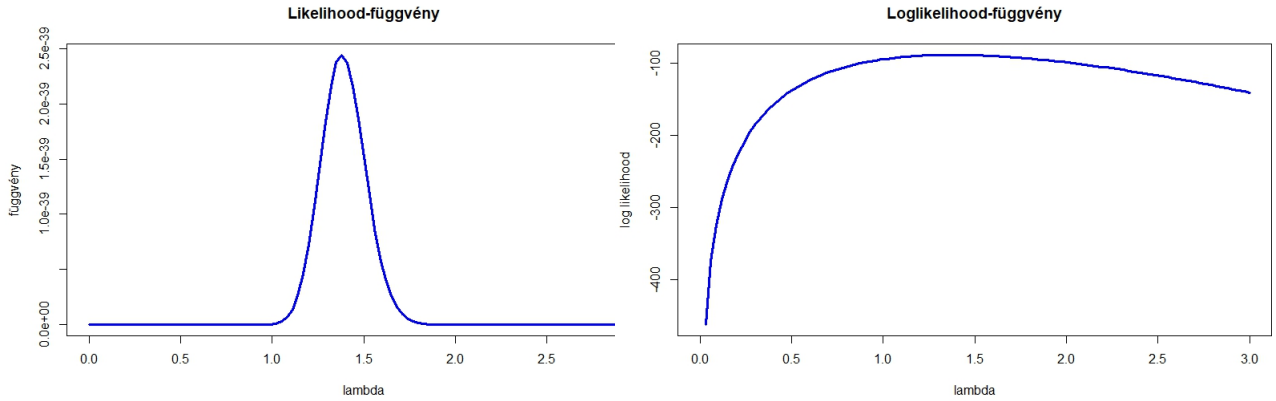
Ahogy a 2. ábra mutatja, ha egy mintára Poisson-eloszlást szeretnénk illeszteni (például további ismeretlen mennyiségek becsléséhez), akkor ki kell választanunk a legmegfelelőbb paramétert, nem mindegyik paraméterrel illeszkedik jól az eloszlás. Ehhez az előbb megismert



1. ábra. A likelihood-függvény: $\mathbb{P}_p(\text{a kábel } 8\text{-szor működik}) = \binom{10}{8} p^8 (1-p)^2$ a p ismeretlen paraméter függvényében. A maximumhelye: $\hat{p} = 0,8$.



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



3. ábra. A $\frac{\lambda^{131}}{3! \cdot 0! \cdot 2! \cdot \dots \cdot 2!} e^{-95 \cdot \lambda}$ likelihoodfüggvény, illetve ennek logaritmus a gólok száma esetén; mintaátlag: $\bar{X} = \frac{131}{95} = 1,379$

elvet követjük: ha a megfigyeléseink sorozata k_1, \dots, k_n , akkor azt a λ -t választjuk, ami mellett annak valószínűsége, hogy éppen k_1, \dots, k_n lesz a megfigyelések sorozata, a lehető legnagyobb.

Tegyük fel, hogy X_1, X_2, \dots, X_n független, azonos λ paraméterű Poisson-eloszlású minta, ahol $\lambda > 0$ ismeretlen paraméter, $n = 95$, és $\bar{X} = 1,379$.

Poisson-eloszlásnál:

$$\mathbb{P}_\lambda(X_j = k) = \frac{\lambda^k}{k!} e^{-\lambda}; \quad \mathbb{E}(X_j) = \lambda; \quad D(X_j) = \sqrt{\lambda}.$$

A megfigyelések az alábbiak (a gólok száma összesen $\sum_{j=1}^n X_j = 131$):

$$3, \quad 0, \quad 2, \quad 2, \quad 1, \quad 3, \dots, 2.$$

Annak valószínűsége λ paraméter mellett, hogy éppen ezt a sorozatot kaptuk:

$$\begin{aligned} L_{95,\lambda}(3, 0, 2, \dots, 2) &= \frac{\lambda^3}{3!} e^{-\lambda} \cdot \frac{\lambda^0}{0!} e^{-\lambda} \cdot \frac{\lambda^2}{2!} e^{-\lambda} \cdot \dots \cdot \frac{\lambda^2}{2!} e^{-\lambda} = \\ &= \frac{\lambda^{3+0+2+2\dots+2}}{3! \cdot 0! \cdot 2! \cdot \dots \cdot 2!} e^{-95 \cdot \lambda} = \frac{\lambda^{131}}{3! \cdot 0! \cdot 2! \cdot \dots \cdot 2!} e^{-95 \cdot \lambda} \end{aligned}$$

Írjuk fel ugyanezt általánosan, utána az lesz a célunk, hogy az a λ -t megkeressük, amire ez a valószínűség a legnagyobb.

X_1, \dots, X_n függetlenek, Poisson-eloszlás $\lambda > 0$ ismeretlen paraméterrel, azaz

$$\mathbb{P}(X_j = k) = \frac{\lambda^k}{k!} e^{-\lambda} \quad (k = 0, 1, 2, \dots).$$

Ekkor

$$L_{n,\lambda}(X_1, \dots, X_n) = \prod_{j=1}^n \left(\frac{\lambda^{X_j}}{X_j!} e^{-\lambda} \right) = \frac{\lambda^{X_1}}{X_1!} e^{-\lambda} \cdot \frac{\lambda^{X_2}}{X_2!} e^{-\lambda} \cdot \dots \cdot \frac{\lambda^{X_n}}{X_n!} e^{-\lambda}.$$

Vonjuk össze az azonos szorzótényezőket:

$$L_{n,\lambda}(X_1, \dots, X_n) = \lambda^{\sum_{j=1}^n X_j} e^{-n\lambda} \cdot \prod_{j=1}^n \frac{1}{X_j!}.$$

Ahogy látni fogjuk, a maximumhelyet megkeresni sokkal egyszerűbb, ha a függvény logaritmusát vesszük. Mivel a logaritmusfüggvény monoton növekvő, az a λ , amire $L_{n,\lambda}$ maximális, ugyanaz, mint amire $\log L_{n,\lambda}$ maximális. Ez látható a 3. ábrán is.

A fenti függvény logaritmusa így írható fel:

$$\log L_{n,\lambda}(X_1, \dots, X_n) = \log \lambda \sum_{j=1}^n X_j - n\lambda + \log \prod_{j=1}^n \frac{1}{X_j!}$$

Ezt deriváljuk λ szerint:

$$\frac{\partial}{\partial \lambda} \log L_{n,\lambda}(X_1, \dots, X_n) = \frac{\sum_{j=1}^n X_j}{\lambda} - n > 0 \Leftrightarrow \lambda < \bar{X}.$$

Vagyis a derivált pontosan akkor pozitív, ha $\lambda > \bar{X}$.

Ezért az ML-becslés: $\hat{\lambda} = \bar{X}$.

1.2. Likelihood-függvény és ML-becslés

Legyen $(\Omega, \mathcal{A}, \mathbb{P})$ statisztikai mező, ahol $\mathcal{P} = \{\mathbb{P}_\vartheta : \vartheta \in \Theta\}$, vagyis az ismeretlen eloszlás a ϑ paraméterrel jellemezhető.

A fenti példákat általánosítva az alábbi definíció azt fogalmazza meg, hogy ϑ paraméter esetén mennyi annak valószínűsége, hogy éppen a k_1, \dots, k_n értékeket figyeljük meg. A folytonos esetben pedig a valószínűségeket a sűrűségfüggvény t_j helyen felvett értékeivel helyettesíthetjük, ahol t_j -k a megfigyelt értékek.

1.1. 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:

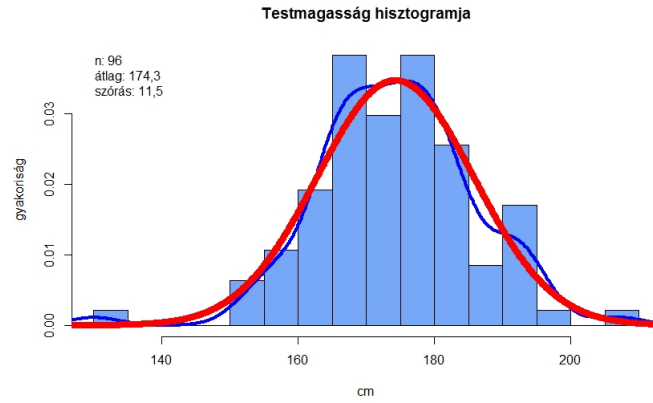
$$L_{n,\vartheta}(t_1, \dots, t_n) = \prod_{j=1}^n f_{j,\vartheta}(t_j) \quad (t_1, \dots, t_n \in \mathbb{R}).$$

1.2. Definíció (Maximum-likelihood becslés). A ϑ maximumlikelihood-becslése (ML-becslése) az X_1, \dots, X_n mintából $\hat{\vartheta}$, ha $\hat{\vartheta}$ maximalizálja a $\vartheta \mapsto L_{n,\vartheta}(X_1, \dots, X_n)$ függvényt, ahol $L_{n,\vartheta}$ a minta likelihood-függvénye. Azaz, ha

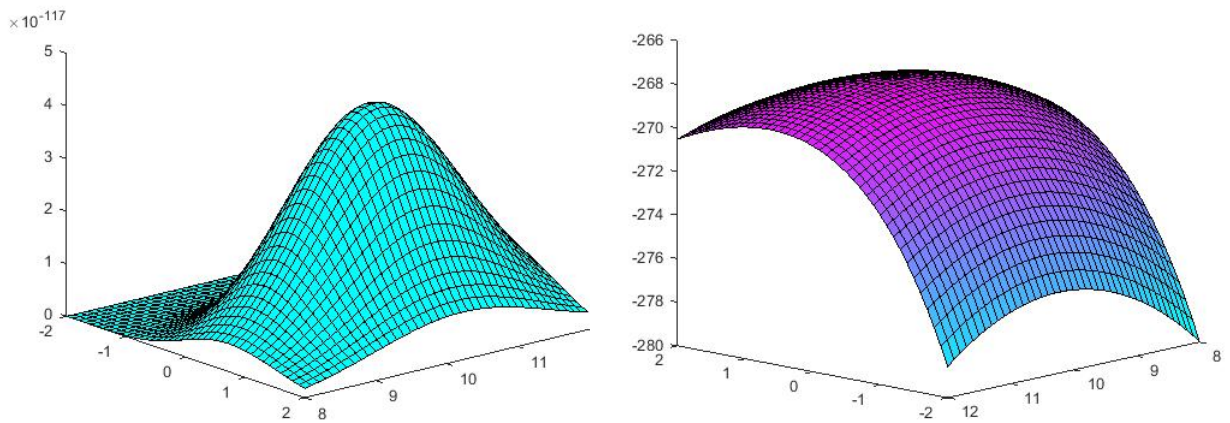
$$L_{n,\hat{\vartheta}}(X_1, \dots, X_n) \geq L_{n,\vartheta}(X_1, \dots, X_n) \quad \text{minden } \vartheta \in \Theta\text{-ra.}$$

1.3. A normális eloszlás paramétereinek becslése

Az is gyakori statisztikai feladat, hogy hogyan becsülhetjük a normális eloszlás paramétereit. Például a 4. ábrán látjuk, hogy a testmagasság hisztogramjához egy megfelelően választott normális eloszlás sűrűségfüggvénye jól illeszkedik, kérdés azonban, hogy pontosan milyen paramétereket válasszunk, hogy a legjobb illeszkedést kapjuk.



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



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

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

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

Ezt alakítsuk át:

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

Az előzőhöz hasonlóan vegyük a logaritmusát:

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

Rögzített σ mellett ez akkor maximális, ha a harmadik tagban $\sum_{j=1}^n (X_j - m)^2 = \sum_{j=1}^n X_j^2 - 2\sum_{j=1}^n X_j m + nm^2$ minimális $\Rightarrow \hat{m} = \bar{X}$.

Most már tudjuk, hogy bármi is σ , mindenképpen m helyére \bar{X} -nek kell kerülnie. Ezért a σ szerinti maximumhelyet úgy kereshetjük, hogy $m = \bar{X}$ -et behelyettesítünk.

A σ szerinti parciális derivált:

$$\frac{\partial}{\partial \sigma} \log L_{n,\sigma}(X_1, \dots, X_n) = -\frac{n}{\sigma} + \sum_{j=1}^n \frac{(X_j - \bar{X})^2}{\sigma^3}.$$

Ez pontosan akkor pozitív, ha $\sigma^2 < \frac{1}{n} \sum_{j=1}^n (X_j - \bar{X})^2 = s_n^2$.

Tehát az ML-bebecslés:

$$\hat{m} = \bar{X}; \quad \hat{\sigma} = s_n = \sqrt{\frac{1}{n} \sum_{j=1}^n X_j^2 - \bar{X}^2}.$$

Tehát normális eloszlásnál az m paraméter bebecslése a mintaátlag, a szórásé a tapasztalati szórás.

2. Egyenletes eloszlás paramétereinek bebecslése

X_1, \dots, X_n függetlenek, eloszlásuk egyenletes eloszlás az $[a, b]$ intervallumon. Ekkor

$$L_{n,a,b}(X_1, \dots, X_n) = \prod_{j=1}^n f_{j,\vartheta}(X_j) = \prod_{j=1}^n \mathbb{I}(a \leq X_j \leq b) \cdot \frac{1}{b-a}.$$

$$L_{n,a,b}(X_1, \dots, X_n) = \left(\frac{1}{b-a}\right)^n \mathbb{I}(a \leq \min_j X_j \text{ és } \max_j X_j \leq b).$$

Az első tényező legyen minél nagyobb (vagyis $b - a$ minél kisebb) úgy, hogy a második tényező nem nulla. Ebből:

$$\hat{a} = \min_j X_j; \quad \hat{b} = \max_j X_j,$$

hiszen ha a ennél kisebb, vagy b ennél nagyobb lenne, akkor az indikátor nullává válna.

2.1. Az ML-becslés tulajdonságai

- Nem minden statisztikai mezőn létezik ML-becslés.
- Az ML-becslés nem feltétlenül egyértelmű.
- Az ML-becslés nem feltétlenül torzítatlan.
- A $g(\vartheta)$ függvény ML-becslése $g(\hat{\vartheta})$, ahol $\hat{\vartheta}$ ML-becslés ϑ -ra.
- **Megfelelő feltételek** (erős regularitási feltételek mellett) az ML-becslés aszimptotikusan torzítatlan:

$$\lim_{n \rightarrow \infty} \mathbb{E}(\hat{\vartheta}) = \vartheta;$$

aszimptotikusan hatásos (vagyis minimális szórású) és **aszimptotikusan normális eloszlású**, azaz $\sqrt{n}(\hat{\vartheta}_n - \vartheta)$ normális eloszláshoz konvergál eloszlásban $n \rightarrow \infty$ esetén (a \mathbb{P}_ϑ valószínűségekre vonatkozóan).

A normális és Poisson-eloszlás teljesíti ezeket a feltételeket, az egyenletes eloszlás nem.

- Az alábbi egyenlet a maximumlikelihood-egyenlet:

$$\frac{\partial}{\partial \vartheta} \log L_{n,\vartheta}(X_1, \dots, X_n) = 0.$$

Megfelelő feltételek mellett az ML-becslés a maximumlikelihood-egyenlet megoldása (ha az ML-becslés nem számítható ki, de az egyenlet megoldható, gyakran az egyenlet megoldásával helyettesítik az ML-becslést). Az egyenletes eloszlás példa arra, amikor az ML-egyenlet nem oldható meg, az ML-becslés nem ennek a megoldása.

2.2. ML-becslés: nevezetes eloszlások

Néhány további példa:

- binomiális eloszlás ismert k renddel: $\hat{p} = \bar{X}/k$
- Poisson-eloszlás: $\hat{\lambda} = \bar{X}$
- geometriai eloszlás: $\hat{p} = 1/\bar{X}$
- normális eloszlás: $\hat{m} = \bar{X}, \hat{\sigma} = s_n$
- exponenciális eloszlás: $\hat{\lambda} = 1/\bar{X}$
- egyenletes eloszlás: $\hat{a} = \min_j X_j; \quad \hat{b} = \max_j X_j$

Egy alkalmazás a jövedelem eloszlásának becslésére: https://wisostat.uni-koeln.de/sites/statistik/user_upload/Working_Paper_Eckernkemper.pdf

3. Momentum módszer

Egy másik módszert is érdemes megismerni, például mert a likelihood-egyenlet maximumhelye nem mindig határozható meg könnyen. Ez a módszer abból indul ki, hogy (például a nagy számok törvénye alapján) a várható érték közel van az átlaghoz, így azt a paramétert választja, aminél a megfigyelt átlag pontosan megegyezne az eloszlás várható értékével. Ha pedig ez nem ad egyértelmű eredményt, ugyanezt a gondolatmenetet mondjuk el a megfigyelések négyzetére

is: válasszuk azt a ϑ -t, amivel a megfigyelések négyzetének átlaga éppen megegyezne a négyzet várható értékével. Ha szükséges, ezt folytassuk tovább.

Legyen X_1, \dots, X_n független azonos eloszlású minta.

1. Az eloszlás k . momentuma, ha ϑ az ismeretlen paraméter: $\mu_{k,\vartheta} = \mathbb{E}_\vartheta(X_1^k)$.
2. Legyen $\hat{\mu}_k = \frac{1}{n} \sum_{j=1}^n X_j^k$ az eloszlás k . tapasztalati momentuma.
3. Írjuk fel az alábbi egyenleteket a legkisebb olyan k -ig, amire az egyenletrendszer egyértelműen meghatározza ϑ -t (**bár nincs mindig ilyen k**):

$$\mathbb{E}_\vartheta(X_1) = \frac{1}{n} \sum_{j=1}^n X_j;$$

$$\mathbb{E}_\vartheta(X_1^2) = \frac{1}{n} \sum_{j=1}^n X_j^2;$$

...

$$\mathbb{E}_\vartheta(X_1^k) = \frac{1}{n} \sum_{j=1}^n X_j^k.$$

4. A ϑ momentum módszerrel kapott becslése az a $\hat{\vartheta}$, ami megoldása a fenti egyenletrendszernek. **Nem mindig létezik, nem mindig egyértelmű, nem feltétlenül hatásos.**

3.1. Példák

X_1, \dots, X_n független **Poisson-eloszlásúak** ismeretlen $\lambda > 0$ paraméterrel. A $k = 1$ -hez tartozó egyenlet:

$$\mathbb{E}_\lambda(X_1) = \bar{X}.$$

Mivel a λ paraméterű Poisson-eloszlás várható értéke λ :

$$\hat{\lambda} = \bar{X}.$$

X_1, \dots, X_n független $N(m, \sigma^2)$ eloszlású minta (azaz normális eloszlású m várható értékkel és σ szórással).

A $k = 1$ -hez és $k = 2$ -höz tartozó egyenletek:

$$\mathbb{E}_{m,\sigma}(X_1) = m = \bar{X};$$

$$\mathbb{E}_{m,\sigma}(X_1^2) = \sigma^2 + m^2 = \frac{1}{n} \sum_{j=1}^n X_j^2.$$

A másodikba beírva az elsőt: $\sigma^2 = \frac{1}{n} \sum_{j=1}^n X_j^2 - \bar{X}^2 = s_n^2$ (a tapasztalati szórásnégyzet). Tehát az első két egyenlet együtt egyértelműen oldható meg, a momentum módszerrel kapott becslés:

$$\hat{m} = \bar{X}; \quad \hat{\sigma} = s_n.$$

Legyen X_1, \dots, X_n független minta az $[a, b]$ intervallumon egyenletes eloszlásból. Ennek várható értéke $(a + b)/2$, szórása $(b - a)/\sqrt{12}$. Ezek alapján a $k = 1$ -hez és $k = 2$ -höz tartozó egyenlet:

$$\mathbb{E}_{a,b}(X_1) = \frac{a + b}{2} = \bar{X};$$

$$\mathbb{E}_{a,b}(X_1^2) = \frac{(b - a)^2}{12} + \left(\frac{a + b}{2}\right)^2 = \frac{1}{n} \sum_{j=1}^n X_j^2.$$

A másodikba beírva az első: $\frac{(b-a)^2}{12} = \frac{1}{n} \sum_{j=1}^n X_j^2 - \bar{X}^2 = s_n^2$, amiből

$$\hat{a} = \bar{X} - \sqrt{3}s_n; \quad \hat{b} = \bar{X} + \sqrt{3}s_n.$$

Hátránya: előfordulhat, hogy ezek nem is lehetséges értékek: ha \hat{a} nagyobb a legkisebb megfigyelésnél, vagy \hat{b} kisebb a legnagyobb megfigyelésnél.

Azt láthatjuk, hogy a Poisson-eloszlás és normális eloszlás esetén a két módszer ugyanarra az eredményre vezet, az egyenletes eloszlásnál viszont van eltérés.

További olvasnivaló, a Pareto-eloszlás példájával: https://www.math.arizona.edu/~jwatkins/M_moments.pdf

Házi feladat március 10., szerda, 9:00-ig Tekintsük az ismerősöktől gyűjtött mintát, és tegyük fel, hogy az, hogy egy ember hány könyvből olvasott legalább 20 oldalt az elmúlt egy hónapban, Poisson-eloszlású, ismeretlen $\lambda > 0$ paraméterrel.

- a) Ábrázoljuk a likelihood-függvényt.
- b) Mi lesz a $\hat{\lambda}$ maximum-likelihood becslés?
- c) Hasonlítsuk össze a hisztogramot a $\hat{\lambda}$ paraméterű Poisson-eloszlással (ábrázoljuk őket egy ábrán). Mennyire illeszkedik jól a becsült Poisson-eloszlás?