Matematikai statisztika előadás, 5. hét, március 10.
Elégséges statisztika, Fisher-információ, Bayes-becslés

1. Elégséges statisztika

A minta gyakran tartalmaz „felesleges” információt, amiből a paraméterre nézve semmilyen következtetést nem vonhatunk le. Például ha a megfigyelések független azonos eloszlásúak, akkor mindegy, hogy milyen sorrendben kaptuk őket: ha megkérdezünk száz embert találomra valamiről, nem hordoz semmilyen fontos információt, hogy melyik válasz hányadik volt a sorrendben. Bizonyos speciális eloszlások esetén ennél jóval többet is mondhatunk, és ezt a likelihood-függvény segítségével tudjuk pontosan megfogalmazni.

1.1. Definíció. *Legyen $X_1, \ldots, X_n$ független minta ϑ paraméterű eloszlásból. A T statisztika elégséges, ha a likelihood-függvény felírható a következő alakban megfelelő h és g függvényekkel:*

$$L_{n,\vartheta}(X_1, \ldots, X_n) = h(X_1, \ldots, X_n) \cdot g_\vartheta(T(X_1, \ldots, X_n)).$$

Nézzük a Poisson-eloszlás esetét.

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

Ekkor

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

Vonjuk össze az azonos szorzótényezőket:

$$L_{n,\lambda}(X_1, \ldots, X_n) = \prod_{j=1}^{n} \frac{1}{X_j!} \cdot \lambda^{\sum_{j=1}^n X_j} e^{-n\lambda} = h(X_1, \ldots, X_n) \cdot g_\vartheta\left(\sum_{j=1}^{n} X_j \right).$$

Az első tényező nem függ ϑ-tól, a második csak az összegtől függ, így az összeg elégséges statisztika.

- Poisson-eloszlás esetén az összeg és az átlag is elégséges statisztika.
- Normális eloszlásnál elégséges statisztika: $(\overline{X}, s_n)$, vagy $\left(\sum_{j=1}^n X_j^2, \sum_{j=1}^n X_j \right)$.
- Az elégséges statisztika nem egyértelmű.
- Ha létezik maximumlikelihood-becslés, T pedig elégséges statisztika, akkor az ML-becslés felírható $u(T(X_1, \ldots, X_n))$ alakban valamely u függvényre.

Az alábbi tétel azt mondja, hogy ha van egy torzítatlan becslésünk, abból egy hatásosabb, vagyis kisebb szórású, szintén torzítatlan becslés is készíthető, ami csak az elégséges statisztika függvénye, annak segítségével kiszámítható, tehát csak a „lényeges” információt használja.

Nézzünk először egy példát. Poisson-eloszlás paraméterének becslésénél az összeg elégséges statisztika, például X_1 (az első megfigyelés) pedig torzítatlan becslés λ-ra. Az alábbi feltételes várható érték (emlékeztetőül: `https://backhauszagi.web.elte.hu/gyak/sst_vsz4ea_l/sst_` vsz4je_l13.pdf) azt fejezi ki, hogy ha tudjuk az összeget, mit gondolunk, mi lehetett az első megfigyelés:

$$\mathbb{E}(X_1 | X_1 + \ldots + X_n).$$

Másrészt, ez nem más, mint $\overline{X}$, hiszen ha tudjuk, hogy n szám összege $X_1 + \ldots + X_n$, akkor a szimmetria miatt mindegyiknél $(X_1 + \ldots + X_n)/n$ egy jó tipp. Az alábbi állítás azt mondja (általánosabb esetben), hogy ilyenkor $\overline{X}$ szintén torzítatlan becslés, és hatásosabb, mint az eredeti, kisebb a szórása. Másrészt ez csak az elégséges statisztika függvénye: csak az összeget használja, vagyis semmilyen „felesleges" információt.

1.1. Tétel (Rao–Blackwell). *Tegyük fel, hogy a T statisztika torzítatlan becslés a ϑ paraméterre, valamint S elégséges statisztika ϑ-ra. Ekkor megadható olyan T^* becslés, melyre*

- *$T^* = h(S)$ megfelelő h függvénnyel (azaz T^* mintából való kiszámításához elég S-t ismerni);*
- *T^* torzítatlan ϑ-ra: $\mathbb{E}_\vartheta(T^*) = \vartheta$ minden $\vartheta \in \Theta$-ra;*
- *T^* hatásosabb, mint T: $D_\vartheta(T^*) \leq D_\vartheta(T)$ minden $\vartheta \in \Theta$-ra.*

Ha az S statisztika teljes is (azaz minden olyan f függvény, melyre $\mathbb{E}_\vartheta(f(S)) = 0$ minden ϑ-ra, „majdnem mindenütt" nulla), akkor T^* hatásos, azaz minden torzítatlan becslésnél hatásosabb.

A megoldás az $\mathbb{E}(T|S)$ feltételes várható érték lesz, ahogy fent láttuk.

1.1. Állítás. *Poisson-eloszlásnál normális eloszlásoknál (m-re) a mintaelemek összege teljes elégséges statisztika, a mintaátlag pedig hatásos becslése a paraméternek.*

1.1. Fisher-információ

A Fisher-információ célja, hogy megmérje, hogy egy adott minta mennyi információt tartalmaz, és ez alapján mennyire pontos becslést készíthetünk a paraméterekre. Nevezetesen, a Fisher-információ segítségével megfelelő feltételek mellett alsó becslést fogunk tudni adni torzítatlan becslések szórására, vagyis túl kevés információból nem lehet tetszőlegesen kis bizonytalanságú torzítatlan becslést adni. A Fisher-információ kiindulása szintén a likelihood-függvény

1.2. Definíció. *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 $\mathbb{P}_\vartheta$ 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 $T(X_1, \ldots, X_n)$ Fisher-információja ugyanaz, mint $(X_1, \ldots, X_n)$ 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
$$D^2(T(X_1, \ldots, X_n)) \geq \frac{1}{I_n(\vartheta)} = \frac{1}{nI_1(\vartheta)}.$$

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

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 $\vartheta = (m, \sigma)$:
$$\begin{pmatrix} \frac{1}{\sigma^2} & 0 \\ 0 & \frac{2}{\sigma^2} \end{pmatrix}.$$

1.2. A Poisson-eloszlás Fisher-információja

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 $\mathbb{E}_\lambda(X_1) = D^2_\lambda(X_1) = \lambda$, így

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

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

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

1.3. Konfidenciaintervallum a maximumlikelihood-becslés alapján

Gyakran az ismeretlen paraméter becsléseként nem csak egyetlen számot adhatunk meg, hanem például azt mondhatjuk, hogy 5 és 7 közé esik, és ebben 95%-ig biztosak vagyunk. Mivel a paraméter rögzített (ebben a felépítésben), ilyenkor valójában az teljesül, hogy $\mathbb{P}_\vartheta(T_1 \leq \vartheta \leq T_2) \geq 0,95$ igaz ϑ minden lehetséges értékére, ahol T_1 és T_2 becslések, és ez utóbbiak a valószínűségi változók (hiszen a véletlen minta alapján számoljuk ki őket).

Az eddig látott eszközökkel általános feltételek mellett tudunk konfidenciaintervallumot adni egy ismeretlen ϑ valós értékű paraméterre. Ez a maximumlikelihood-függvény alábbi tulajdonságain, illetve a Fisher-információval való kapcsolatán alapul.

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

- *létezik;*
- *aszimptotikusan torzítatlan:* $\lim_{n\to\infty}\mathbb{E}_\vartheta(\hat{\vartheta}_n) = \vartheta$ *minden* $\vartheta \in \Theta$*-ra;*

- *aszimptotikusan hatásos:* $\lim_{n\to\infty}\sqrt{nI_1(\vartheta)}D_\vartheta(\hat\vartheta_n)=1$ *minden* $\vartheta\in\Theta$*-ra;*
- *aszimptotikusan normális eloszlású:* $\sqrt{nI_1(\vartheta)}(\hat\vartheta_n-\vartheta)$ *eloszlásban tart a standard normális eloszláshoz minden* $\vartheta\in\Theta$*-ra* $n\to\infty$ *esetén.*

Itt $I_1(\vartheta)$ *az egyelemű mintából számolt Fisher-információt jelöli.*

Ez alapján **aszimptotikus** konfidenciaintervallum, ami $n\to\infty$ 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:

$$\lim_{n\to\infty}\mathbb{P}_\vartheta\left(\hat\vartheta_n-\Phi^{-1}\left(1-\frac{\alpha}{2}\right)\cdot\frac{1}{\sqrt{\hat I_n(\vartheta)}}\le\vartheta\le\hat\vartheta_n+\Phi^{-1}\left(1-\frac{\alpha}{2}\right)\cdot\frac{1}{\sqrt{\hat I_n(\vartheta)}}\right)=1-\alpha.$$

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).

2. 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.

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 találkoztunk ezzel. Ezt a becslés nem tükrözi.

Egy másik hátrány, hogy ha a paraméterről valamilyen előzetes információval rendelkezünk, azt a frekventista hozzáállásban nem tudjuk jól felhasználni, hiszen a feltételeknek minden ϑ-ra egyformán teljesülnie kell. Ha például először 100 kipróbálás alapján a mellékhatás valószínűségét $10-30\%$ közöttire becsüljük, és utána csinálunk egy nagyobb vizsgálatot 10000 résztvevővel, akkor a nagyobb vizsgálat elemzésénél nem akkora baj, ha a $p=80\%$-ra „kevésbé” teljesül egy feltétel, mint $p=20\%$-ra.

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

2.1. Bayes-becslés: a priori és a posteriori eloszlás

Példa. Tegyük fel, hogy egy biztosító háromféle ügyfelet különböztet meg a gépjármű felelősségbiztosításnál: az óvatos, átlagos és merész sofőröket. Az, hogy egy ügyfél hány balesetet okoz egy év alatt, Poisson-eloszlású, $0,1$, $0,2$ és $0,3$ paraméterekkel az egyes csoportoknál. Azt is tudjuk, hogy egy új ügyfél mindhárom csoportba azonos valószínűséggel tartozik.

Legyen a Péter által okozott balesetek száma egy év alatt Poisson-eloszlású λ paraméterrel, a fenti feltételeknek megfelelően. Adjunk becslést a λ paraméterre, ha Péter az utóbbi tíz évben egyáltalán nem okozott balesetet. Az éveket függetlennek tekinthetjük.

Itt a λ ismeretlen paraméter valószínűségi változónak tekinthető, aminek az eloszlásáról az előzetes feltételezésünk:

$$\mathbb{P}(\lambda = 0,1) = \mathbb{P}(\lambda = 0,2) = \mathbb{P}(\lambda = 0,3) = \frac{1}{3}.$$

A paraméter eloszlásáról alkotott előzetes, a minta megismerése előtt feltételezett eloszlás az **a priori eloszlás**. Ez tükrözi a paraméterre vonatkozó, a megfigyelésnél korábbi információkat.

Legyen A az az esemény, hogy Péter tíz év alatt nem okozott balesetet. Ennek valószínűsége, ha s a rá jellemző eloszlás igazi paramétere:

$$\mathbb{P}(A|\lambda = s) = (e^{-s})^{10} = e^{-10s},$$

a Poisson-eloszlás definícióját $k = 0$-val használva.

Valójában azonban nem ez a kérdés, hanem az, hogy feltéve, hogy A bekövetkezett, mennyi a valószínűsége, hogy λ, vagyis az ismeretlen paraméter egy adott értéket vesz fel. Másképpen, az a kérdés, hogy az egyes lehetséges értékek valószínűsége hogyan módosul, ha felhasználjuk a megfigyelésekből származó információt.

A Bayes-tétel alapján

$$\mathbb{P}(\lambda = 0,1|A) = \frac{\mathbb{P}(A|\lambda = 0,1)\mathbb{P}(\lambda = 0,1)}{\mathbb{P}(A|\lambda = 0,1)\mathbb{P}(\lambda = 0,1) + \mathbb{P}(A|\lambda = 0,2)\mathbb{P}(\lambda = 0,2) + \mathbb{P}(A|\lambda = 0,3)\mathbb{P}(\lambda = 0,3)} =$$
$$= \frac{e^{-10\cdot 0,1} \cdot \frac{1}{3}}{e^{-10\cdot 0,1} \cdot \frac{1}{3} + e^{-10\cdot 0,2} \cdot \frac{1}{3} + e^{-10\cdot 0,3} \cdot \frac{1}{3}} = \frac{e^{-1}}{e^{-1} + e^{-2} + e^{-3}} = 66,5\%.$$

Hasonlóképpen:

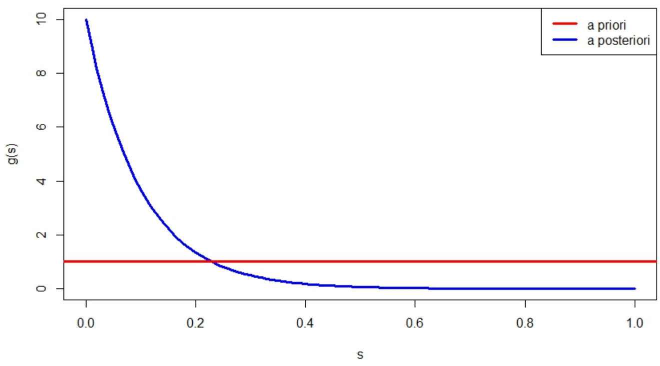
$$\mathbb{P}(\lambda = 0,2|A) = \frac{e^{-2}}{e^{-1} + e^{-2} + e^{-3}} = 24,5\%$$

$$\mathbb{P}(\lambda = 0,3|A) = \frac{e^{-3}}{e^{-1} + e^{-2} + e^{-3}} = 9\%.$$

Tehát, míg a minta ismerete nélkül a három lehetséges értéket egyformán valószínűnek tartottuk, a megfigyelések alapján, arra feltételesen, hogy Péter tíz évig nem okozott balesetet, már jóval valószínűbb a $0,1$ érték, vagyis, hogy Péter óvatos vezető, és sokkal kevésbé valószínű az, hogy a merész sofőrök közé tartozik.

A fent kiszámított valószínűségek adják a λ paraméter megfigyelésekre vonatkozó feltételes eloszlását, ami az **a posteriori eloszlás**. Ez mutatja meg a megfigyelések alapján alkotott elképzelést arról, hogy melyik érték mennyire valószínű.

Ez közvetlenül még nem ad egyetlen számot, vagyis egyetlen becslést λ-ra. A becslést ebben a példában nem is fogjuk meghatározni, nézzük meg azonban azt az esetet, amikor az a priori eloszlás nem diszkrét, hanem folytonos.



1. ábra. A λ paraméter a priori és a posteriori sűrűségfüggvénye

Példa. Tegyük fel, hogy egy biztosító minden ügyfeléről azt feltételezi, hogy az általa egy év alatt okozott balesetek száma Poisson-eloszlású, paramétere pedig az ügyfélre jellemző λ, mely egy ügyfelet tekintve egyenletes eloszlású a $[0,1]$ intervallumon. Adjunk becslést a Péterre jellemző λ paraméterre, ha Péter az utóbbi tíz évben egyetlen balesetet sem okozott.

Az alábbi egyenlet ugyanúgy érvényes, mint eddig, minden $s \in [0,1]$-re:

$$\mathbb{P}(A|\lambda = s) = (e^{-s})^{10} = e^{-10s}.$$

A Bayes-tétel (folytonos változata) azonban most ezt adja:

$$p^*(s|A) = \frac{\mathbb{P}(A|\lambda = s)p(s)}{\int_{-\infty}^{\infty} \mathbb{P}(A|\lambda = t)p(t)dt},$$

ahol p a paraméter a priori sűrűségfüggvénye. A példában $p(t) = \mathbb{I}(0 \le t \le 1)$, hiszen azt tettük fel előzetesen, hogy t egyenletes eloszlású a $[0,1]$ intervallumon.

Behelyettesítve megkapjuk az a posteriori eloszlás sűrűségfüggvényét:

$$p^*(s|A) = \frac{\mathbb{P}(A|\lambda = s)p(s)}{\int_{-\infty}^{\infty} \mathbb{P}(A|\lambda = t)p(t)dt} = \frac{e^{-10s}\mathbb{I}(0 \le s \le 1)}{\int_0^1 e^{-10t}dt} = \frac{e^{-10s}\mathbb{I}(0 \le s \le 1)}{\frac{1}{10}(1 - e^{-10})}$$

Vagyis a megfigyelések alapján s-t olyan eloszlásúnak gondolhatjuk, aminek ez a fenti függvény a sűrűségfüggvénye.

Az 1. ábrán láthatjuk az a priori és a posteriori sűrűségfüggvényt. Előzetesen egyenletes eloszlást tételeztünk fel, azonban az alapján, hogy Péter nem okozott balesetet, a λ kis értéktartományai valószínűbbek (ha a paraméter kisebb, akkor az okozott balesetek számának várható értéke is kisebb).

Ebben az esetben becslést is adhatunk (később látni fogjuk, hogy ez milyen szempontból jó becslés). Ugyanis kiszámíthatjuk az a posteriori eloszlás várható értékét, parciális integrálás segítségével:

$$\hat{\lambda} = \int_0^1 sp^*(s|A)d\lambda = \int_0^1 s \cdot \frac{e^{-10s}}{\frac{1}{10}(1 - e^{-10})} = 0,1.$$

Tehát a Bayes-módszerrel kapott becslés értéke $\hat{p} = 0,1$.

További olvasnivaló például: [1]

Házi feladat március 17., szerda, 9:00-ig Tegyük fel, hogy egy telefontöltő minden kipróbálásnál egy rá jellemző p valószínűséggel működik, ha pedig véletlenszerűen választunk egy töltőt, akkor p sűrűségfüggvénye a $[0,1]$ intervallumon $2x$, és 0 azon kívül. Tegyük fel, hogy a töltő 20 kipróbálásból 19-szer működött. Határozzuk meg az erre a megfigyelésre vonatkozó a posteriori eloszlást (sűrűségfüggvényt), és annak várható értékét.

Hivatkozások

[1] Blazejowski, M., Gazda, J., Kwiatkowski, J. (2016), Bayesian Model Averaging in the Studies on Economic Growth in the EU Regions – Application of the gretl BMA package, Economics and Sociology, Vol. 9, No 4, pp. 168-175.`https://www.economics-sociology.eu/files/E&S_9_4_Blazejowski_Gazda_Kwiatkowski.pdf`