

Likelihood-függvény (5. 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:

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

Likelihood-függvény (5. 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:

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

ML-becslés: Poisson-eloszlás

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

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

A likelihood-függvény:

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

Kérdés: milyen λ -ra lesz ez a függvény maximális ($\hat{\lambda} = \bar{X}$).

$$L_{n,\lambda}(X_1, \dots, X_n) = g_\lambda \left(\sum_{j=1}^n X_j \right) \cdot h(X_1, X_2, \dots, X_n)$$

alakú. A második tényező nem függ λ -tól, ezért a maximum-likelihood becsléshez elég az első tényező maximumhelyét keresni, ehhez pedig **elég az összeget ismerni**.

ML-becslés: normális eloszlás

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

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

Vagyis a likelihood-függvény felírásához és a maximumlikelihood-becslés meghatározásához **elég az összeget és a négyzetösszeget ismerni**. Láttuk, hogy

$$(\hat{m}, \hat{\sigma}) = (\bar{X}, s_n),$$

ez valóban kiszámítható az összeg és a négyzetösszeg segítségével.

Elégséges statisztika

Definíció

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

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

Elégséges statisztika

Definíció

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

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

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

Rao–Blackwell-tétel

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;
- 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. Például Poisson-eloszlásnál $\mathbb{E}(X_1 | \sum_{j=1}^n X_j) = \bar{X}$ hatásos.

Állítás

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

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

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

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

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

Házi feladat március 11., 8:15-ig: megoldás

Tekintsünk a (ϑ, ϑ^2) intervallumon egyenletes eloszlású mintát: X_1, X_2, \dots, X_n . Itt $\vartheta > 0$ ismeretlen paraméter. Adjuk meg ϑ momentum módszerrel és maximum-likelihood-módszerrel kapott becslését.

A likelihood-függvény: $L_{\vartheta}(X_1, X_2, \dots, X_n) = \frac{1}{|\vartheta^2 - \vartheta|} \mathbb{I}(\vartheta \leq X_1^*, X_n^* \leq \vartheta^2)$.

A feltétel így fogalmazható át: $\sqrt{X_n^*} \leq \vartheta \leq X_1^*$. Ha $0 \leq \vartheta \leq 1$: a $|\vartheta^2 - \vartheta|$ függvény $\vartheta = 1/2$ -re a legnagyobb, és konvex. Ezért a legnagyobb likelihood-érték ezek közül valamelyik értékre kapható meg: X_1^* és $\sqrt{X_n^*}$. Ha $\vartheta \geq 1$, akkor a függvény monoton növekvő, és $\sqrt{X_n^*}$ lesz a maximumlikelihood-becslés.

Momentum módszerrel: az első egyenlet $\bar{X} = \frac{\vartheta + \vartheta^2}{2}$. Ebből

$$\hat{\vartheta} = -\frac{1}{2} + \sqrt{1/4 + 2\bar{X}}.$$

Házi feladat március 11., 8:15-ig

```
x=c(1.8, 2.5, 3.5, 4.3, 5.8, 6.3, 8, 10, 12, 15)
ML=1:10
momentum=1:10
for (j in 1:10)
{
minta=runif(100, min=x[j], max=x[j]^2)
ML[j]=sqrt(max(minta))
momentum[j]=-0.5+sqrt(0.25+2*mean(minta))
}
```

Házi feladat március 11., 8:15-ig: megoldás

valódi paraméter	ML	momentum
1,8	1,799	1,808
2,5	2,462	2,455
3,5	3,497	3,583
4,3	4,271	4,405
5,8	5,79	5,725
6,3	6,297	6,406
8	7,983	8,108
10	9,86	10,048
12	11,997	11,955
15	14,99	15,173
relatív hibák összege	0,042	0,134
relatív hibák maximuma	0,015	0,024

Tíz ϑ értékre 100 elemű mintából való becslés a kétféle módszerrel

Konfidenciaintervallum

Példa: megmérték hatvan ember vércukorszintjét, egymástól függetlenül.

A minta egy részlete (mmol/l mértékegységben):

5,98 6,1 5,99 6,21 5,97 6,23 ... 5,85

Alapstatisztikák: $n = 60$ (méret), $\bar{X} = 5,99$ (átlag), $s_n^* = 0,18$ (korrigált tapasztalati szórás)

Konfidenciaintervallum

Példa: megmérték hatvan ember vércukorszintjét, egymástól függetlenül.

A minta egy részlete (mmol/l mértékegységben):

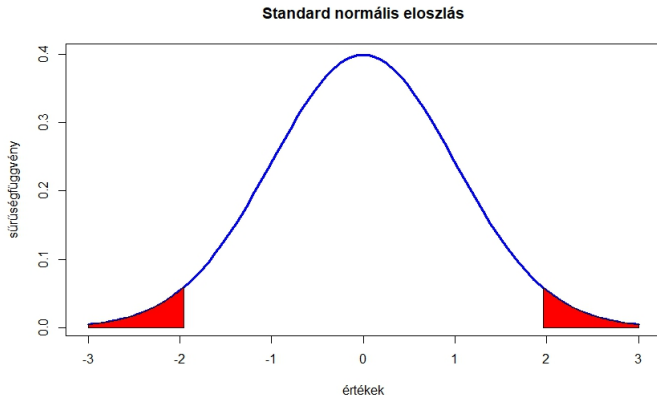
5,98 6,1 5,99 6,21 5,97 6,23 ... 5,85

Alapstatisztikák: $n = 60$ (méret), $\bar{X} = 5,99$ (átlag), $s_n^* = 0,18$ (korigált tapasztalati szórás)

Cél: a várható érték becslése az adatok alapján

pontosabban: adjunk meg egy olyan intervallumot, ami legalább 95% valószínűséggel tartalmazza az "igazi" várható értéket – ezt fogjuk **95 % megbízhatósági szintű konfidenciaintervallumnak** hívni

A kétoldali z-próba kritikus értéke



Az $\alpha = 0,05$ terjedelmű kétoldali z-próba (u -próba) kritikus értéke:
 $\Phi^{-1}(1 - \alpha/2) = \Phi^{-1}(0,975) = 1,96$. Vagyis ha $Z \sim N(0,1)$, akkor
 $\mathbb{P}(-1,96 \leq Z \leq 1,96) = 95\%$.

Konfidenciaintervallum a várható értékre

$$(T_1, T_2) = \left(\bar{X} - \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) \frac{\sigma}{\sqrt{n}}, \bar{X} + \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) \frac{\sigma}{\sqrt{n}} \right)$$

Ekkor

$$\begin{aligned} \mathbb{P}_m(T_1 \leq m \leq T_2) &= \mathbb{P}_m\left(m - \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) \frac{\sigma}{\sqrt{n}} \leq \bar{X} \leq m + \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) \frac{\sigma}{\sqrt{n}}\right) \\ &= \mathbb{P}_m\left(-\Phi^{-1}\left(1 - \frac{\alpha}{2}\right) \leq \frac{\bar{X} - m}{\sigma/\sqrt{n}} \leq \Phi^{-1}\left(1 - \frac{\alpha}{2}\right)\right) = \\ &= 2\left(1 - \frac{\alpha}{2}\right) - 1 = 1 - \alpha, \end{aligned}$$

ahol felhasználtuk, hogy mivel X_1, \dots, X_n független $N(m, \sigma^2)$ eloszlásúak, így $\bar{X} \sim N(m, \sigma^2/n)$ eloszlású. Ezért $Z = (\bar{X} - m)/(\sigma/\sqrt{n})$ standard normális eloszlású, vagyis

$$\mathbb{P}(-a \leq Z \leq a) = \Phi(a) - \Phi(-a) = \Phi(a) - (1 - \Phi(a)) = 2\Phi(a) - 1.$$

Példa konfidenciaintervallumra

Minta: megmértük hatvan ember vércukorszintjét, tegyük fel, hogy ez **normális eloszlású, valódi szórása 0,2**.

$n = 60$ (méret), $\bar{X} = 5,99$ (átlag), $s_n^* = 0,18$ (korrigált tapasztalati szórás)

Adjunk meg olyan intervallumot, ami legalább 95% valószínűséggel tartalmazza a vércukorszint valódi várható értékét.

megbízhatósági szint: $1 - \alpha = 95\%$, azaz $\alpha = 0,05$. Ebből $\Phi^{-1}(1 - \alpha/2) = \Phi^{-1}(0,975) = 1,96$.

95 %-os megbízhatósági szintű konfidenciaintervallum a várható értékre:

$$\left(5,99 - 1,96 \cdot \frac{0,2}{\sqrt{60}}, 5,99 + 1,96 \cdot \frac{0,2}{\sqrt{60}} \right) = (5,94; 6,04).$$

Kisebb mintaelemszámhoz hosszabb, nagyobb megbízhatósághoz szintén hosszabb konfidenciaintervallum tartozik.

Konfidenciaintervallumok

Legyen $(\Omega, \mathcal{A}, \mathcal{P})$ statisztikai mező, $\mathcal{P} = \{\mathbb{P}_\vartheta : \vartheta \in \Theta\}$ és $\underline{X} = (X_1, \dots, X_n)$ független azonos eloszlású minta. Tegyük fel, hogy ϑ valós paraméter, vagyis $\Theta \subseteq \mathbb{R}$.

Definíció

Azt mondjuk, hogy a $(T_1(\underline{X}), T_2(\underline{X}))$ intervallum legalább $1 - \alpha$ megbízhatósági szintű konfidenciaintervallum ϑ -ra, ha minden $\vartheta \in \mathbb{R}$ esetén teljesül, hogy

$$\mathbb{P}_\vartheta(T_1(\underline{X}) < \vartheta < T_2(\underline{X})) \geq 1 - \alpha.$$

A konfidenciaintervallum megbízhatósági szintje: $\inf_{\vartheta \in \Theta} \{\mathbb{P}_\vartheta(\vartheta \in (T_1, T_2))\}$.

Konfidenciaintervallum a várható értékre

A Φ a standard normális eloszlás eloszlásfüggvénye, azaz ha $Z \sim N(0, 1)$:

$$\Phi(t) = \mathbb{P}(Z \leq t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^t e^{-s^2/2} ds.$$

Állítás (Konfidenciaintervallum a várható értékre, **ismert szórás**)

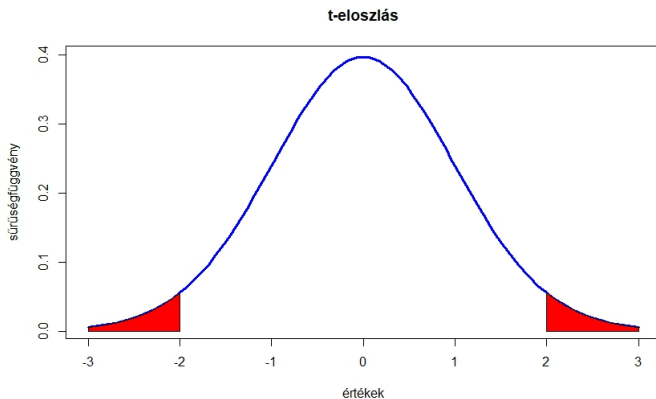
Tegyük fel, hogy X_1, \dots, X_n független azonos eloszlású **normális eloszlású** valószínűségi változók, melyek szórása, σ ismert.

Ekkor a

$$(T_1, T_2) = \left(\bar{X} - \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) \frac{\sigma}{\sqrt{n}}, \bar{X} + \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) \frac{\sigma}{\sqrt{n}} \right)$$

intervallum $1 - \alpha$ megbízhatósági szintű **kétoldali konfidenciaintervallum** az eloszlás várható értékére.

t -eloszlás kritikus értékei



Az $f = 59$ szabadsági fokú $\alpha = 0,05$ terjedelmű kétoldali t -próba kritikus értéke:
 $t_{59,0,05} = 2,001$.

Konfidenciaintervallum a várható értékre

Legyenek Z_0, Z_1, \dots, Z_n független $N(0, 1)$ eloszlásúak, és $t_{f, \alpha}$ az f szabadsági fokú α terjedelmű kétoldali t -próba kritikus értéke, azaz az f szabadsági fokú t -eloszlás $1 - \alpha/2$ -kvantilise:

$$1 - \alpha/2 = \mathbb{P}(Y \leq t_{f, \alpha}) = \mathbb{P}\left(\frac{Z_0}{\sqrt{Z_1^2 + \dots + Z_f^2}} \leq t_{f, \alpha}\right).$$

Az $Y = \frac{Z_0}{\sqrt{Z_1^2 + \dots + Z_f^2}}$ valószínűségi változó eloszlása f szabadsági fokú **t -eloszlás**.

Konfidenciaintervallum a várható értékre

Legyenek Z_0, Z_1, \dots, Z_n független $N(0, 1)$ eloszlásúak, és $t_{f, \alpha}$ az f szabadsági fokú α terjedelmű kétoldali t -próba kritikus értéke, azaz az f szabadsági fokú t -eloszlás $1 - \alpha/2$ -kvantilise:

$$1 - \alpha/2 = \mathbb{P}(Y \leq t_{f, \alpha}) = \mathbb{P}\left(\frac{Z_0}{\sqrt{Z_1^2 + \dots + Z_f^2}} \leq t_{f, \alpha}\right).$$

Az $Y = \frac{Z_0}{\sqrt{Z_1^2 + \dots + Z_f^2}}$ valószínűségi változó eloszlása f szabadsági fokú **t -eloszlás**.

Állítás (Konfidenciaintervallum a várható értékre, ismeretlen szórás)

Tegyük fel, hogy X_1, \dots, X_n független $N(m, \sigma^2)$ normális eloszlású valószínűségi változók (m, σ ismeretlenek). Ekkor a

$$(T_1, T_2) = \left(\bar{X} - t_{n-1, \alpha} \cdot \frac{S_n^*}{\sqrt{n}}, \bar{X} + t_{n-1, \alpha} \cdot \frac{S_n^*}{\sqrt{n}} \right)$$

*intervallum $1 - \alpha$ megbízhatósági szintű **kétoldali konfidenciaintervallum** az eloszlás várható értékére.*

Példa konfidenciaintervallumra

Minta: vércukorszint-értékek, tegyük fel, hogy ez **normális eloszlású**, szórása nem ismert.

$n = 60$ (méret), $\bar{X} = 5,99$ (átlag), $s_n^* = 0,18$ (korigált tapasztalati szórás)

Adjunk meg olyan intervallumot, ami legalább 95% valószínűséggel tartalmazza a vércukorszint valódi várható értékét.

megbízhatósági szint: $1 - \alpha = 95\%$, azaz $\alpha = 0,05$. Az $f = 59$ szabadsági fokú $\alpha = 0,05$ terjedelmű kétoldali t -próba kritikus értéke: $t_{59,0,975} = 2$.

95 %-os megbízhatósági szintű konfidenciaintervallum a várható értékre:

$$\left(5,99 - 2 \cdot \frac{0,18}{\sqrt{60}}, 5,99 + 2 \cdot \frac{0,18}{\sqrt{60}} \right) = (5,94; 6,04).$$

Hosszabb intervallumot kaptunk, mint ismert szórás esetén.

Konfidenciaintervallum a valószínűsége

Az A esemény valószínűsége $p \in [0, 1]$ ismeretlen paraméter. Ezt szeretnénk megbecsülni. Ha n kísérletből az A esemény X -szer következett be:

$$\mathbb{E}_p(X) = np; \quad D_p(X) = \sqrt{np(1-p)}; \quad \mathbb{E}_p\left(\frac{X}{n}\right) = p; \quad D_p\left(\frac{X}{n}\right) = \frac{\sqrt{p(1-p)}}{\sqrt{n}}.$$

$X \sim \text{Bin}(n, p)$ binomiális eloszlású, viszont a relatív gyakoriságot, X/n -t nem tudjuk egyetlen eloszlással közelíteni.

Ezért X/n eloszlását \hat{p} várható értékű, $\frac{\hat{p}(1-\hat{p})}{\sqrt{n}}$ szórású normális eloszlással közelítjük, ahol \hat{p} az esemény relatív gyakorisága a mintában.

$1 - \alpha$ megbízhatósági szintű konfidenciaintervallum p -re:

$$\left(\hat{p} - \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) \cdot \frac{\sqrt{\hat{p}(1-\hat{p})}}{\sqrt{n}}; \hat{p} + \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) \cdot \frac{\sqrt{\hat{p}(1-\hat{p})}}{\sqrt{n}} \right).$$

Házi feladat március 18., 8:15-ig

A házi feladathoz begyűjtött adatok alapján

- 1 adjunk konfidenciaintervallumot az utazási idők várható értékére külön a férfiak, külön a nők esetében, illetve összesen;
- 2 adjunk konfidenciaintervallumot annak valószínűségére, hogy egy véletlenszerűen kiválasztott ismerős legalább 5 sorozatot nézett az utóbbi egy hónapban.

Az utazási időkről tételezzük fel, hogy normális eloszlásúak (noha valójában a kerékítések miatt ez nem igazán teljesül).