További konf intv konstrukciók [Viharos, jegyzet]

1) $\mathcal{P} = \{ F_{m,\sigma} : m \in \mathbb{R}, \sigma > 0 \}$ szórású normális eloszlások

$$Chi_n(\sigma) := \frac{n V_n(X)}{\sigma^2} = \sum_{k=1}^{n} \left[\frac{X_k - \frac{1}{n}(X_1 + \dots + X_n)}{\sigma / \sqrt{n}} \right]^2 \quad \chi^2_{n-1} \text{ eloszl}$$

(HF) $\delta := \chi^2_{n-1}(1-\frac{\varepsilon}{2})$ segítségével $(1-\varepsilon)$-MEGB. KONF INTV σ-RA

2) $X = (X_1, \dots, X_{n_1})$ $Y = (Y_1, \dots, Y_{n_2})$ FGTL normális eloszl minták UGYAZON SZÓRÁSSAL

$X_i \sim N(m_1, \sigma)$ $Y_j \sim N(m_2, \sigma)$

KONF INTV $1-\varepsilon$ MEGB $|m_1 - m_2|$-re:

$$\frac{(\bar{X} - \bar{Y}) - (m_1 - m_2)}{\sqrt{\frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2}}} = Z(m_1 - m_2) \sim N(0,1) \text{ standard norm.}$$

(HF) $\delta := \Phi^{-1}(1 - \frac{\varepsilon}{2})$ segítségével $(1-\varepsilon)$-konf intv $|m_1 - m_2|$-re

2') UAZ, MINT 2), DE σ ismeretlen

$$\frac{(\bar{X} - \bar{Y}) - (m_1 - m_2)}{D^*(X,Y)} : t_{n_1 + n_2 - 2} \text{ - eloszl.}$$

$$D^*(X,Y)^2 = \frac{V(X_1)/(n_1 - 1)}{n_1} + \frac{V(X_2)/(n_2 - 1)}{n_2}$$

(HF) $\delta := t^{-1}_{n_1 + n_2 - 2}(1 - \varepsilon)$ segítségével konf intv $|m_1 - m_2|$-re

Finomítás az σ ism. $\sigma > 0$ lehetőségre

$\mathcal{D} = \{F_{m,\sigma} : m\in\mathbb{R}, \sigma>0\}$ σ ism. normális eloszl.

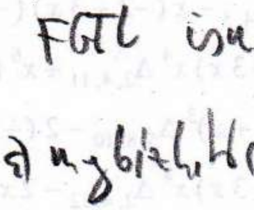
Tudjuk: $X_1,\dots,X_n$ függ. $F_{m,\sigma}$-szerinti eloszl.

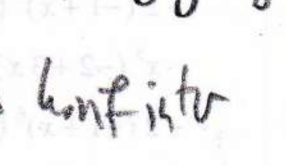
$$T_{n-1}(m) = \frac{\bar X - m}{\left[\frac{1}{n-1}V_n(X)\right]^{1/2}} = \frac{\frac{1}{n}(X_1+\dots+X_n)-m}{\left[\sum_{k=1}^{n}\left(X_k - \frac{1}{n}(X_1+\dots+X_n)\right)^2\right]^{1/2}}\sqrt{n-1} \quad \text{Student}_{n-1} \text{ eloszl.}$$

Megj. Gosset statisztikus volt egy sörgyárban, kellett álnéven közölni, hogy STUDENT NÉV alatt ismert sörgyári ügyek

Algoritmus: Adott $\varepsilon>0$-hoz $(1-\varepsilon)$ megbízhatóságú konf. intv.

$$\delta := t_{n-1}^{-1}\left(1-\frac{\varepsilon}{2}\right)$$





Ha m X valódi várható értéke, akkor mivel

$$T_{n-1}(m) = \frac{\bar X - m}{\sqrt{V_n(X)}}\sqrt{n-1} \quad t_{n-1}\text{-eloszlású},$$

$$\mathbb{P}\left(|T_{n-1}(m)|\le\delta\right) = 1-\varepsilon$$

$$|T_{n-1}(m)|\le\delta \iff -\delta\le T_{n-1}(m)\le\delta \iff -\delta \le \frac{\bar X-m}{\sqrt{V_n(X)}}\sqrt{n-1}\le\delta$$

$$\Updownarrow$$

$$-\frac{\delta\sqrt{V_n(X)}}{\sqrt{n-1}} \le \bar X - m \le \frac{\delta\sqrt{V_n(X)}}{\sqrt{n-1}} \iff$$

$$m\in\left[\bar X - \frac{\delta\sqrt{V_n(X)}}{\sqrt{n-1}},\ \bar X + \frac{\delta\sqrt{V_n(X)}}{\sqrt{n-1}}\right] \leftarrow \text{CONF INTV } 1-\varepsilon \text{ MEGBÍZH.}$$

3) $\mathbb{X}=(X_1,\dots,X_{n_1})$, $\mathbb{Y}=(Y_1,\dots,Y_{n_2})$ FGTL $N(m_1,\sigma_1)$, $N(m_2,\sigma_2)$ minták

$(1-\varepsilon)$-KONF INTV σ_1/σ_2-re

$$F_{n_1-1,n_2-1}(\sigma_1/\sigma_2) := \frac{\left[\frac{1}{n_1-1}V_{n_1}(\mathbb{X})\right] / \left[\frac{1}{n_2-1}V_{n_2}(\mathbb{Y})\right]}{(\sigma_1/\sigma_2)^2} \qquad \underline{F_{n_1-1,n_2-1}\ \text{elo}}$$

(MF) $\alpha := F^{-1}_{n_1-1,n_2-1}\left(\frac{\varepsilon}{2}\right)$ $\beta := F^{-1}_{n_1-1,n_2-1}\left(1-\frac{\varepsilon}{2}\right)$ mellett

$$\mathbb{P}\left(\alpha < \frac{\frac{n_2-1}{n_1-1}\, V_{n_1}(X)/V_{n_2}(Y)}{(\sigma_1/\sigma_2)^2} < \beta\right) = 1-\varepsilon$$

$$\mathbb{P}\left(\left[\frac{1}{\beta}\cdot\frac{n_2-1}{n_1-1}\cdot\frac{V_{n_1}(X)}{V_{n_2}(Y)}\right]^{1/2} \le \frac{\sigma_1}{\sigma_2} \le \left[\frac{1}{\alpha}\,\frac{n_2-1}{n_1-1}\,\frac{V_{n_1}(X)}{V_{n_2}(Y)}\right]^{1/2}\right) = 1-\varepsilon$$

ebből $(1-\varepsilon)$-KONF INTV σ_1/σ_2-re

<u>Megjegyzés</u> EXCEL-program is készíthető 1), 2), 3)-ra

STATISZTIKAI PRÓBÁK (TESZTEK)

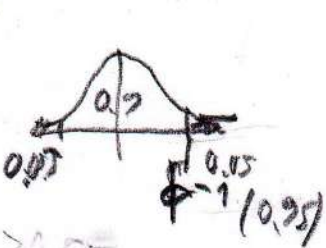
Példa: Mennyire lehetünk biztosak abban, hogy Magyarországon a felnőttek átlagmagassága 175 cm? [Megjegyzés: 175 helyett elvárhatjuk 175+ε közeli ε-nal is elfogadhatónak. Pl. 90%-os biztonsággal elfogadható lehet a [175−ε, 175+ε]-ba eső bármely érték]. MODELL: $X \sim N(m,\sigma)$ a magasság, $\sigma \le \sigma_{max} = 15$ (cm), $X_1,\dots,X_n$ X-minta

$x_1 = X_1(\omega), \dots, x_n = X_n(\omega)$ konkrét mérések

Döntés u-próba (u-teszt) segítségével: $\mathcal{D} = \{F_{m,\sigma} : m\in\mathbb{R}, \sigma\le\sigma_{max}\}$

Tudjuk: $\bar X = \frac{1}{n}(X_1+\dots+X_n)$-re $\frac{\bar X - m}{\sigma/\sqrt n} \sim \underbrace{N(0,1)}_{F_{0,1}}$ $\forall F_{m,\sigma}\in\mathcal D$

HA $m=175$ $P\left(-\delta \le \frac{\bar X - 175}{\sigma/\sqrt n} \le \delta\right) \ge 0.9 \iff \delta \ge \Phi^{-1}(0.95)$



Akkor $P\left(|\bar X - 175| \le \Phi^{-1}(0.95)\cdot\frac{\sigma_{max}}{\sqrt n}\right) \ge 0.9$

$|\bar x - 175| \ge \Phi^{-1}(0.95)\frac{\sigma_{max}}{\sqrt n}$ csak ritkán, az esetek $0.1 = (1-0.9)$ részében teljesül.

Döntés u-teszt segítségével: a átlagmag. = 175 90%-os szinten elfogadjuk / nem elfogadjuk aszerint, hogy $|\bar x - 175| \underset{\ge}{\le} \Phi^{-1}(0.95)\frac{15}{\sqrt n}$

Megjegyzés 1) $\mathcal D = \{F_{m,\sigma}: m\in\mathbb R, \sigma\le\sigma_{max}\}$ esetén láttuk:

$$\left[\bar X - \Phi^{-1}(0.95)\frac{\sigma_{max}}{\sqrt n},\ \bar X + \Phi^{-1}(0.95)\frac{\sigma_{max}}{\sqrt n}\right]$$ 0.9 megbízhatóságú konf. intv. m-re

HA $m=175$ ebben van, akkor elfogadás

2) Minden konfidencia intv konstrukcióhoz lehet így szerkeszteni egy tesztet. Pl. student → t-próba

(HF) programba a tanult konf intv alapján

Általános alapfogalmak $\mathcal{D} = \{F_\theta : \theta \in \Theta\}$ eloszláscsalád

$\mathcal{D} = \mathcal{D}_0 \dot\cup \mathcal{D}_1$ $\mathcal{D}_j = \{F_\theta : \theta \in \Theta_j\}$ $\Theta_0 \dot\cup \Theta_1 = \Theta$ FELOSZTÁS, 2 RÉSZRE

ELDÖNTENDŐ KÉRDÉS (kérdés) : MILYEN MEGBÍZHATÓSÁGGAL MONDHATJUK EGY $X_1, \dots, X_n$ MINTAVÉTEL ($\sim F_\theta$-eloszlású valószínűségi változók valamely $\theta \in \Theta$-ra) ALAPJÁN HOGY $\theta \in \Theta_1$?

$H_0 : \theta \in \Theta_0$ NULLHIPOTÉZIS, $H_1 : \theta \in \Theta_1$ ELLENHIPOTÉZIS

H_0, H_1 „nem szimmetrikus" :

1-ső fajú hiba : ELVETJÜK H_0-t, PEDIG $\theta \in \Theta_0$

2-od fajú hiba : ELFOGADJUK H_0-t, PEDIG $\theta \in \Theta_1$

Példa COVID 19 -nél $H_0 \simeq$ FERTŐZÖTT, $H_1 \simeq$ NEM FERTŐZÖTT

(Az a rosszabb, ha elengedjük a fertőzöttet — 1. fajú hiba)

Statisztikai DÖNTÉS $\varepsilon > 0$ és $n = 1, 2, \dots$ mellett képezünk olyan $K_n(\varepsilon) \subset \mathbb{R}^n$ halmazokat, hogy

$\mathbb{P}\{\omega : (X_1(\omega), \dots, X_n(\omega)) \in K_n(\varepsilon)\} \le \varepsilon$ valahányszor $X_1, \dots, X_n \sim F_\theta$ és $\theta \in \Theta_0$

Döntés : ELVETJÜK H_0-t, ha $(\underbrace{X_1(\omega)}_{x_1}, \dots, \underbrace{X_n(\omega)}_{x_n}) \in K_n(\varepsilon)$

$\boxed{1-\varepsilon \text{ SZINTEN}}$ ELFOGADJUK H_0-t ha $\notin K_n(\varepsilon)$

Megjegyzés A döntés 1-ső fajú hibája $\le \varepsilon$ a konstrukció miatt

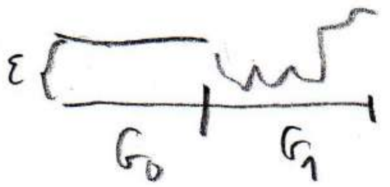
Példa: (TRIV) $K_n(\varepsilon) = \emptyset$ — (mindenkit fertőzöttnek nyilvánítunk)

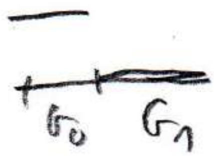
Ekkor nincs 1-ső hiba, SEMMITMONDÓ

Erőfüggvény $E_n(\varepsilon, \theta) = [1\text{-ső hiba valószínűsége ha } \theta \in \Theta_0,\ 1-(2\text{-od hiba valószínűsége}) \text{ ha } \theta \in \Theta_1]$

$= \mathbb{P}((X_1, \dots, X_n) \in K_n(\varepsilon))$ ha $X_1, \dots, X_n \sim F_\theta$, $\theta \in \Theta$

Adott $\varepsilon>0$ mellett $E_n(\varepsilon,\cdot)$



Pl 

A TRIV. DÖNTÉSSEL

Cél: Fix $\varepsilon>0$, $n\to\infty$ mellett $E_n(\varepsilon,\theta)\to 1$ teszt konzisztens

<u>Pl</u> u-próba $\mathcal{P}=\{F_{m,\sigma}: m\in\mathbb{R}, \sigma<\sigma_{max}\}$

$H_0: m=m_0$ $\quad G_0=\{(m_0,\sigma):\sigma<\sigma_{max}\}$

$K_n(\varepsilon)=\{(x_1,\dots,x_n):$ H_0 elutasítva $(X_1=x_1,\dots,X_n=x_n)$ a legjobb $F_{m_0,\sigma}$ esetén$\}=$

$=\left\{(x_1,\dots,x_n): \bar x\notin \left[m_0-\Phi^{-1}(1-\tfrac{\varepsilon}{2})\frac{\sigma_{max}}{\sqrt n},\ m_0+\Phi^{-1}(1-\tfrac{\varepsilon}{2})\frac{\sigma_{max}}{\sqrt n}\right]\right\}$

HA $m\neq m_0$ és $F_{m,\sigma}$ AZ IGAZI ELOSZLÁS, akkor 2. hiba jelentése

$\bar x\in\left[m_0+\Phi^{-1}(1-\tfrac{\varepsilon}{2})\frac{\sigma_{max}}{\sqrt n},\ \Phi^{-1}(1+\tfrac{\varepsilon}{2})\frac{\sigma_{max}}{\sqrt n}\right] = \left[m_0+\frac{d}{\sqrt n}, m_0-\frac{d}{\sqrt n}\right]$

ENNEK VALÓSZ.

$$1-E_n(\varepsilon,(m,\sigma)) = \mathbb{P}_{m,\sigma}\Big(m_0-\frac{d}{\sqrt n}\le \underbrace{\bar X}_{\frac1n(X_1+\dots+X_n)}\le m_0+\frac{d}{\sqrt n}\Big)$$

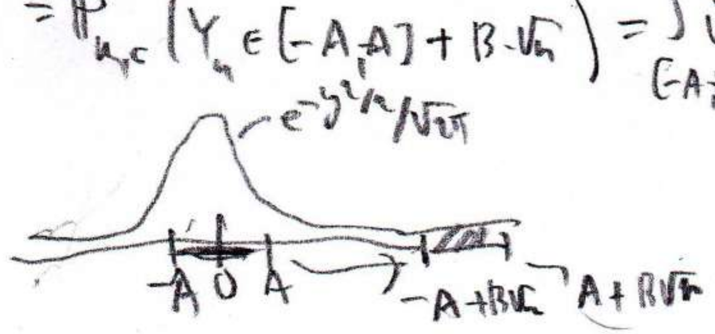
$$X\sim F_{m,\sigma}\Rightarrow \bar X\sim F_{m,\sigma/\sqrt n},\ \frac{\bar X-m}{\sigma/\sqrt n}\sim N(0,1)$$

$$E_n(\varepsilon,(m,\sigma)) = \mathbb{P}_{m,\sigma}\Big(\frac{m_0-d/\sqrt n-m}{\sigma/\sqrt n}\le Y_n\le \frac{m_0+d/\sqrt n-m}{\sigma/\sqrt n}\Big)=$$

$$=\mathbb{P}_{m,\sigma}\Big(Y_n\in\Big[\underbrace{-\frac d\sigma}_{A\neq0}+\underbrace{\frac{m_0-m}{\sigma}}_{B\neq0}\sqrt n,\ \frac d\sigma+\frac{m_0-m}{\sigma}\sqrt n\Big]\Big)=$$

$$=\mathbb{P}_{m,\sigma}\big(Y_n\in[-A,A]+B\sqrt n\big)=\int_{[-A,A]+B\sqrt n}\frac{1}{\sqrt{2\pi}}e^{-y^2/2}\,dy\to 0$$

$[-A,A]+B\sqrt n\to\infty$



Illeszkedésvizsgálat χ^2-próbával

Kérdés: $A_1,\dots,A_r$ teljes eseményrendszer $\left(A_i\cap A_j=\emptyset\ (i\neq j),\ \sum_j P(A_j)=1\right)$

$(p_1,\dots,p_r)$-rel milyen megbízhatósággal mondhatjuk, hogy $p_1=P(A_1),\dots,p_r=P(A_r)$ és $X_1,\dots,X_n:\Omega\to\{1,\dots,r\}$ minta elemei, ha $\omega\in\Omega$ esetén $X_k(\omega)=[j:\omega\in A_j]$

Pl. kockadobás, $r=6$, $X_k=[j.\text{ dobás eredménye}]$

Megjegyzés: MAX-LIKELIHOOD $(p_1,\dots,p_r)=\left(\frac{\#\{k:X_k=1\}}{n},\dots,\frac{\#\{k:X_k=r\}}{n}\right)$

Emlékeztető: χ^2_n eloszlás $=\Big[Z_1^2+\dots+Z_n^2$ eloszlása, ha $Z_1,\dots,Z_n$ FGTL $N(0,1)$ eloszl.$\Big]$

χ^2_n sűrűségfüggvénye

$$f_{\chi^2_n}(x)=\frac{d}{dx}\mathbb{P}(Z_1^2+\dots+Z_n^2<x)=\frac{d}{dx}\mathbb{P}\big((Z_1,\dots,Z_n)\in x\underbrace{B_n}_{n\text{ dim egységgömb}}\big)=$$

$$=\frac{d}{dx}\int_{(t_1,\dots,t_n)\in xB_n}\left(\frac{1}{\sqrt{2\pi}}e^{-t_1^2/2}\right)\cdots\left(\frac{1}{\sqrt{2\pi}}e^{-t_n^2/2}\right)dt_1\dots dt_n \underset{r=\sqrt{t_1^2+\dots+t_n^2}}{\overset{\text{gömbi koord.}}{=}}$$

$$=\frac{d}{dx}\int_{r=0}^{x}\left(\frac{1}{\sqrt{2\pi}}\right)^n e^{-r^2/2}\cdot r^{n-1}\,\mathrm{Vol}_{n-1}(\partial B_n)\,dr=$$

$$=\frac{x^{(n/2)-1}e^{-x/2}}{2^{n/2}\Gamma\left(\frac{n}{2}\right)}\qquad(x>0)$$

$\Gamma(N)=(N-1)!$

$\Gamma\left(\frac{1}{2}\right)=\sqrt{\pi}$

$\Gamma(z)=(z-1)\Gamma(z-1)$

Tétel (NEHÉZ) $X_1, X_2, \ldots : \Omega \to \{1, \ldots, r\}$ (kísérlet r kimenettel)

$\nu_j(n) := \omega \mapsto \#\{k : k \leq n,\ X_k(\omega) = j\}$ n-ből hány j. kimenetel

$$K_{n-1} := \sum_{j=1}^{r} \frac{[\nu_j(n) - np_j]^2}{np_j}, \quad \text{ahol } p_j = \mathbb{P}(j.\ \text{kimenetel})$$

EKKOR $\mathbb{P}(K_{n-1} < x) \to \int_{t=0}^{x} f_{\chi^2_{r-1}}(t)\,dt \quad (n \to \infty,\ x > 0)$.

Megjegyzés NAGY n-re VEHETŐ K_{n-1} χ^2_{r-1}-eloszlásúnak

Közelítő teszt (az eloszlások megfelelőségére)

$H_0 : p_1 = p_1^0, \ldots, p_r = p_r^0$ $1-\varepsilon$ SZIGNIFIKANCIÁVAL, n KÍSÉRLETBŐL ELUTASÍTÁS, HA

$$\sum_{j=1}^{r} \frac{[\#\{j.\ \text{kimenetel}\} - np_j^0]^2}{np_j^0} > h_{r-1}(1-\varepsilon)$$

ahol $h_{r-1}(\delta) = \left[s > 0 : \int_{t=0}^{s} f_{\chi^2_{r-1}}(t)\,dt = \delta\right]$