KLASSZIKUS ELOSZLÁSOK ML-BECSLÉSEI

1) Geom. eloszl. $\mathcal{S}=\{F_p : p\in[0,1]\}$ $\quad \mathbb{P}_p\{n\}=(1-p)^{n-1}p$

Minta: $X_1=\ell_1,\dots,X_n=\ell_n$ [PONTOSABBAN $X_k(\omega)=x_k$ VOLNA]

$$L_p=[(1-p)^{\ell_1-1}p][(1-p)^{\ell_2-1}p]\cdots[(1-p)^{\ell_n-1}p]=$$

$$=(1-p)^{\ell-n}p^n \quad \text{ahol } \ell=\sum_k \ell_k$$

$L_p\to \mathrm{MIN}_p \quad : \quad \nabla L_p=0, \quad \frac{\partial L_p}{\partial p}=0$

$$-(\ell-n)(1-p)^{\ell-n-1}p^n+(1-p)^{\ell-n}\cdot n p^{n-1}=0 \quad /:(1-p)^{\ell-n-1}p^{n-1}$$

$$-(\ell-n)\cdot p+n\cdot(1-p)=0$$

$$\boxed{p=n/\ell=n/\sum_{k=1}^{n}\ell_k}$$ NEM TORZÍTATLAN

2) Exponenciális eloszl. FOLYT. $\{f_\lambda : \lambda>0\}$ $\quad f_\lambda(x)=\lambda e^{-\lambda x}$

sűrűségfgv

Minta: $X_1=x_1,\dots,X_n=x_n$

$$\ell_\lambda=[\lambda e^{-\lambda x_1}][\lambda e^{-\lambda x_2}]\cdots[\lambda e^{-\lambda x_n}]=$$

$$=\lambda^n e^{-\lambda(x_1+\dots+x_n)}=\lambda^n e^{-n\bar{x}\lambda}$$ — ÁTLAG $(x_1+\dots+x_n)/n$

$\ell_\lambda\to \mathrm{MIN}_p \quad \nabla L_\lambda=0, \quad \frac{\partial \ell_\lambda}{\partial \lambda}=0,$

$$n\lambda^{n-1}e^{-n\bar{x}\lambda}-\lambda^n\cdot n\cdot\lambda e^{-n\lambda\bar{x}}=0$$

$$\boxed{\lambda=\frac{1}{\bar{x}}}$$ NEM TORZÍTATLAN

3) NORMÁLIS ELOSZLÁSOK $\mathcal{D} = \left\{ \underbrace{\frac{1}{\sigma\sqrt{2\pi}} e^{-(x-m)^2/(2\sigma^2)}}_{f_{m,\sigma} \text{ SŰRŰSÉGFGV}} : m \in \mathbb{R}, \sigma > 0 \right\}$

Minta: $X_1 = x_1, \dots, X_n = x_n$

$$\ell_{m,\sigma} = \prod_{k=1}^{n} \left[\frac{1}{\sigma\sqrt{2\pi}} e^{-(x_k-m)^2/(2\sigma^2)} \right]$$

$$\ell_{m,\sigma} \to \mathrm{MAX}_{m,\sigma} \Leftrightarrow \underline{\log \ell_{m,\sigma} \to \mathrm{MAX}_{m,\sigma}}$$

$$\frac{\partial \log_{m,\sigma}}{\partial m} = 0, \quad \frac{\partial \log_{m,\sigma}}{\partial \sigma} = 0 \qquad (\Leftarrow \nabla \ell_{m\sigma} = 0)$$

$$\log \ell_{m,\sigma} = -\frac{n}{2} \log(2\pi) - n \log \sigma - \sum_{k=1}^{n} (x_k - m)^2 / (2\sigma^2)$$

$$\frac{\partial \log_{m,\sigma}}{\partial m} = \frac{\partial}{\partial m} \sum_{k=1}^{n} (x_k - m)^2 / (2\sigma^2) = \sum_{k=1}^{n} 2(m - x_k) / (2\sigma^2) =$$

$$= [n \cdot m - n \cdot \underset{\uparrow \sum_k x_k}{\bar{x}}] / \sigma^2 = 0 \qquad \boxed{m = \bar{x}}$$

$$\frac{\partial \log_{m,\sigma}}{\partial \sigma} = -\frac{n}{\sigma} - \sum_{k=1}^{n} (x_k - m)^2 \frac{1}{2} (-2) \sigma^{-3} =$$

$$= -n\sigma^{-1} + \sum_{k=1}^{n} (x_k - \bar{x})^2 \sigma^{-3} = 0 \qquad / \cdot \sigma^3$$

$$\boxed{\sigma^2 = \frac{1}{n} \sum_{k=1}^{n} (x_k - \bar{x})^2} - V_n(x_1, \dots, x_n) \text{ NEM TORZÍTATLAN}$$

3*) REGRESSZIÓS FGV (GAUSS-MÓDSZER)

Alapprobléma: Keressük olyan $f_{\theta_1,\dots,\theta_r}(x)$ FGV-t, amelyre ADOTT $x_1,\dots,x_n$ (nem véletlen!) helyeken

$$f_\theta(x_1)=y_1+\varepsilon_1,\ \dots,\ f_\theta(x_n)=y_n+\varepsilon_n,$$

ahol $y_1,\dots,y_n$ szintén nem véletlen értékek, de a $\varepsilon_1,\dots,\varepsilon_n$ VÉLETLEN HIBÁK közös $N(0,\sigma)$ eloszlásúak.

Mi a legjobb $\theta=(\theta_1,\dots,\theta_r)$ VÁLASZTÁS?

[Történelemben: ELVESZETT KISBOLYGÓ]

$y_1,\dots,y_n$ MÉRT ADATOK AZ $x_1,\dots,x_n$ HELYEKEN, AZAZ AZ IGAZI $\tilde\theta=(\tilde\theta_1,\dots,\tilde\theta_r)$ MELLETT $y_k=f_{\tilde\theta}(x_k)$ $(k=1,\dots,n)$ PONTOSAN

ML-alkalmazás

A k-adik mérésnél az x_k HELYEN A MÉRÉSI EREDMÉNY

$Y_k=f_{\tilde\theta}(x_k)+\varepsilon_k$ $N(f_{\tilde\theta}(x_k),\sigma)$ eloszl. miatt

$$\text{SŰRŰSÉGFGV}=\frac{1}{\sigma\sqrt{2\pi}}e^{-[y_k-f_{\tilde\theta}(x_k)]^2/(2\sigma^2)}$$

$$L_\theta=\prod_{k=1}^{n}\left\{\frac{1}{\sigma\sqrt{2\pi}}e^{-[y_k-f_\theta(x_k)]^2/(2\sigma^2)}\right\}$$

$$\frac{\partial L_\theta}{\partial\theta_j}=0\quad(j=1,\dots,r)$$

AHÁNY PARAMÉTER VAN

EGYSZERŰBBEN:

$$\ell_\theta = \sigma^{-n}(2\pi)^{-n/2}\, e^{-\sum_k [y_k - f_\theta(x_k)]^2/(2\sigma^2)} \to MAX_\theta \Leftrightarrow$$

$$e^{-\sum_k [y_k - f_\theta(x_k)]^2} \to MAX_\theta \Leftrightarrow \underline{\sum_{k=1}^n [y_k - f_\theta(x_k)]^2 \to MIN_\theta}$$

Alappélda: LIN REGRESSZIÓ $f_{(a,b)}(x) = ax+b \quad (a,b\in\mathbb{R})$

$$(\tilde a,\tilde b) = \arg\min_{a,b} \sum_{k=1}^n [y_k-(ax_k+b)]^2$$

$$0 = \frac{\partial}{\partial a}\sum_{k=1}^n [y_k-(ax_k+b)]^2 = 2\sum_{k=1}^n [y_k-(ax_k+b)]\cdot(-x_k)$$

$$0 = \frac{\partial}{\partial b}\sum_{k=1}^n [y_k-(ax_k+b)]^2 = 2\sum_{k=1}^n [y_k-(ax_k+b)]\cdot(-1)$$

↳ $0=\sum_{k=1}^n [y_k-(ax_k+b)] \to \underline{b = \frac{1}{n}\sum_{k=1}^n (y_k - ax_k) = \bar y - a\bar x}$

$$0=\sum_{k=1}^n [y_k-(ax_k+(\bar y - a\bar x))]\,x_k$$

$$0=\sum_{k=1}^n [y_k-\bar y]\,x_k - a\sum_{k=1}^n (x_k-\bar x)\,x_k$$

$$a = \sum_{k=1}^n [y_k-\bar y]\,x_k \Big/ \sum_{k=1}^n (x_k-\bar x)\,x_k$$

SZEBB: $0 = \underbrace{\sum_k (y_k-\bar y)}_{0}\bar x - a\sum_k (x_k-\bar x)\,\bar x$

$$0=\sum_{k=1}^n (y_k-\bar y)(x_k-\bar x) - a\sum_{k=1}^n (x_k-\bar x)^2$$

$$a = \frac{\sum_k (y_k-\bar{y})(x_k-\bar{x})}{\sum_k (x_k-\bar{x})^2} = \frac{Cov_n(X,Y)}{V_n(X,Y)} \qquad b = \bar{y} - a\bar{x}$$

Megjegyzés Gauss eredeti álláspontjából LINEÁRISNÁL sokkal BONYOLULTABB KIFEJEZÉST keresett

Állítás: $\sum_{k=1}^{n} [y_k - f_\theta(x_k)]^2 \to MIN_\theta$ megoldható [illegible] NUM. MÓDSZERREL

(HF) Ellenőrizzük, hogy a $\nabla_\theta L_\theta = 0$ differenciálással kihozott eredmény $(1,2,3^*)$-ban TÉNYLEG MIN-hely
$[L_\theta > 0,\ \lim_{|\theta|\to\infty} L_\theta = 0$ alapján részletes biz.$]$

További klasszikus eloszlások ML-becslései:
[Vihsros, A stochasztika alapjai, 114–119]
POISSON, INT HELYE, HIPERGEOM. terjedelme

BECSLÉSEK MEGBÍZHATÓSÁGA, KONFIDENCIA-INTV-OK

Emlékeztető $\mathcal{F} = \{F_\theta : \theta \in G\}$ eloszláscsalád, $G \subset \mathbb{R}^r$

$\theta = (\theta_1, \dots, \theta_r)$ $\theta_1, \dots, \theta_r$ paraméterek

[SŐT $F_\theta \mapsto \varphi(\theta)$ is paraméter sok indokból]

$s : \mathbb{R}^n \to \mathbb{R}$ n-változós becslő fgv

$X_1, \dots, X_n : \Omega \to \mathbb{R}$ azonos eloszlású vál. vált., FGTL

$\omega \in \Omega$-ra $X_1(\omega), \dots, X_n(\omega)$ egy F_θ-minta, ha

$X_i \sim F_\theta$ azaz $\mathbb{P}(X_i < x) = \mathbb{P}\{\omega : X_i(\omega) < x\} = F_\theta(x)$
$\forall x \in \mathbb{R}$

$s(X_1, \dots, X_n) : \omega \mapsto s(X_1(\omega), \dots, X_n(\omega))$ egy F_θ-statisztika

Def $a : \mathbb{R}^n \to \mathbb{R}$ fgv a $\mathcal{F}$ eloszláscsalád $\varphi(\theta)$ paraméterének $(1-\varepsilon)$-megbízhatóságú alsó becslése:

$\mathbb{P}(a(X_1, \dots, X_n) \le \varphi(\theta)) \ge 1 - \varepsilon \quad \forall \theta \in G$

$\Updownarrow$

$\mathbb{P}(a(x_1, \dots, x_n) \le \varphi(\theta))$ VALAHÁNYSZOR $(x_1, \dots, x_n)$ EGY F_θ-minta, ahol $\theta \in G$

FELSŐ BECSLÉSRE HASONLÓAN $\ge$-vel

Pl $\mathcal{F} := \{\underbrace{[0,L]\text{-en egyenl. eloszl.}}_{F_L} : L > 0\}$

$b : (x_1, \dots, x_n) \mapsto 1.1 \cdot \max\{x_1, \dots, x_n\}$ mennyire megbízható becslése L-nek?

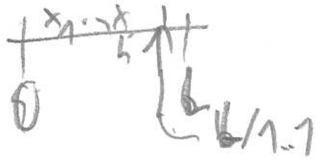
Tf. hogy $X_1,\dots,X_n$ F_L-minta

$$\mathbb{P}\{s(X_1,\dots,X_n)\le L\} = \mathbb{P}\{\max\{X_1,\dots,X_n\}\cdot 1{,}1 \le L\} =$$

$$=\mathbb{P}\{\omega: X_1(\omega),\dots,X_n(\omega)\le \tfrac{L}{1.1}\} \underset{X_1,\dots,X_n \text{ FGTL}}{=} \mathbb{P}\left(X_1(\omega)\le \tfrac{L}{1.1}\right)^n \underset{X_i\ F_{(0,L)}\text{-elo}}{=}$$

$$= \left[\frac{L/1.1}{L}\right]^n = 1.1^{-n}$$

NINCS BENNE L



Tehát $1.1\cdot\max\{X_1,\dots,X_n\}$ $\frac{1}{1.1^n}$-megbízhatóságú n-mintás alsó becslése az L jobb oldali...

TRIV: $1.1\max\{X_1,\dots,X_n\}$ 1=100% megbízhatóságú felső becslés

<u>Def</u> Legyenek $a,b:\mathbb{R}^n\to\mathbb{R}$, $a<b$ FGV-ek

$(x_1,\dots,x_n)\mapsto[a(x_1,\dots,x_n), b(x_1,\dots,x_n)]$ intv-fgv

$1-\varepsilon$ megbízhatóságú KONFIDENCIA INTERVALLUM A

$\Delta=\{F_\theta:\theta\in G\}$ eloszláscsalád $\varphi(\theta)$ paraméterére,

$\mathbb{P}\{\varphi(\theta)\in[a(X_1,\dots,X_n), b(X_1,\dots,X_n)]\}\ge 1-\varepsilon$ valahányszor

$X_1,\dots,X_n$ egy F_θ-minta, ahol θ tetszőleges $\in G$.

<u>Pl</u> $\max\{X_1,\dots,X_n\}\le L\le 1.1\max\{X_1,\dots,X_n\}$ $1-1.1^{-n}$ megbízhatóságú becslése L-nek, azaz $[\max\{X_1,\dots,X_n\}, 1.1\max\{X_1,\dots,X_n\}]$ $1-1.1^{-n}$ megbízhatóságú konfidencia L-re

Konfidencia-intervallumok előállítása

Emlékeztető $X=(X_1,\dots,X_n)$ $N(m,\sigma)$-minta

$X_1,\dots,X_n$ FGTL $\quad P(X_i<x)=F_{m,\sigma}(x)=\int_{t=-\infty}^{x}\frac{1}{\sigma\sqrt{2\pi}}e^{-(t-m)^2/(2\sigma^2)}\,dt$

$\bar X := \frac{1}{n}(X_1+\dots+X_n)\sim N(m,\sigma/\sqrt{n})$

$Var_n(X)=\frac{1}{n}\cdot\sum_{k=1}^{n}(X_k-\bar X)^2$ FGTL $\bar X$-tól, $\frac{n-1}{n}\sigma^2$ várh. ért.

$\frac{Var(X)}{\sigma\sqrt{(n-1)/n}}\sim\chi^2_{n-1}$ $(n-1)$-szabadságfokú χ^2_{n-1} eloszl.

$\frac{\bar X-m}{\left[Var(X)\cdot\frac{n}{n-1}\right]^{1/2}}$ $(n-1)$ szab. fokú Student (vagy t_{n-1}) eloszl.

$$\chi^2_{n}(x)=\int_{t=0}^{x}\frac{t^{n-1}e^{-t^2/2}}{2^{n/2-1}\Gamma(n/2)}\,dt$$

$$t_n(x)=\int_{t=-\infty}^{x}\frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi}\,\Gamma(n/2)}\left(1+\frac{t^2}{n}\right)^{-\frac{n+1}{2}}dt$$

ismert képlet sűrűségfgv-ek

[Történelem: Student = W.S. Gosset, Guiness sörgyár]

$[\bar X-d,\bar X+d]$ alakú konf. int. $\sigma\le\sigma_{MAX}$ esetén, $n=?$

tegyük $\Delta=\{F_{m,\sigma}: m\in\mathbb{R},\ \sigma\le\sigma_{max}\}$ és $(X_1,\dots,X_n)$ Δ-minta

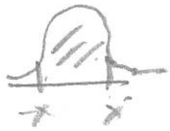
$P(m\in[\bar X-d,\bar X+d])=P(|\bar X-m|\le d)=$

$=P\left(-\frac{d}{\sigma/\sqrt{n}}\le\frac{\bar X-m}{\sigma/\sqrt{n}}\le\frac{d}{\sigma/\sqrt{n}}\right)=\int_{t=-d\sqrt{n}/\sigma}^{d\sqrt{n}/\sigma}\frac{1}{\sqrt{2\pi}}e^{-t^2/2}\,dt$

$$\mathbb{P}(m \in [\bar{X}-d, X+d]) \le \int_{t=-d\sqrt{n}/\sigma_{MAX}}^{d\sqrt{n}/\sigma_{MAX}} \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\, dt$$

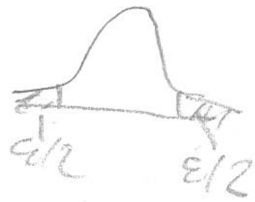
<u>Jelölés</u> $\Phi(x) = \int_{t=-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\, dt$

$$\int_{t=-x}^{x} \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\, dt = 2\Phi(x) - 1$$



<u>Tétel</u> A σ_{MAX}-nál kisebb szórású normális eloszlású $X_1, \dots, X_n$ mintából $[\bar{X}-d, \bar{X}+d]$ a $2\Phi\left(\frac{d\sqrt{n}}{\sigma_{MAX}}\right)-1$ megbízhatóságú konfidencia intv.

Adott (kicsiny) $\varepsilon > 0$ mellett válasszuk



$\Phi^{-1}\left(1-\frac{\varepsilon}{2}\right) = \delta$ értéket (erre $2\Phi(\delta)-1 = \varepsilon$)

$\frac{d\sqrt{n}}{\sigma_{MAX}} = \delta$ alapján $d := \delta \cdot \sigma_{MAX}/\sqrt{n}$ mellett

$$\left[\bar{X} - \delta \cdot \frac{\sigma_{MAX}}{\sqrt{n}}, \bar{X} + \delta \cdot \frac{\sigma_{MAX}}{\sqrt{n}}\right]$$ az $1-\varepsilon$ megbízhatóságú konf. intv.

ez az ÁTLAGRA

<u>Megjegyzés</u> Minden be vagy egyéb a később tárgyalandó <u>döntési</u> programokba — mind STANDARDIZÁLT AUTOMATIKUS ELJÁRÁSOK