Curs 19

Exp: Fie $X_1, X_2, \dots, X_n \sim f_\theta$

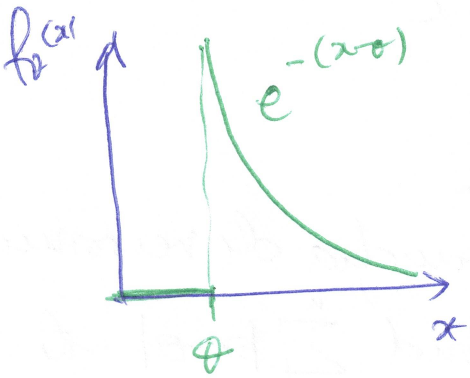
$$f_\theta(x) = e^{-(x-\theta)}, \quad x \geq \theta, \quad \theta > 0$$

Funcția de verosimilitate:

$$L(\theta|x) = \prod_{i=1}^{n} f_\theta(x_i) = \prod_{i=1}^{n} e^{-(x_i-\theta)} \cdot \mathbb{1}_{[\theta,+\infty)}(x_i)$$

$$= e^{-\sum_{i=1}^{n}(x_i-\theta)} \cdot \prod_{i=1}^{n} \mathbb{1}_{[\theta,+\infty)}(x_i)$$

$$= e^{n\theta-\sum_{i=1}^{n} x_i} \underbrace{\mathbb{1}_{[\theta,+\infty)}(x_1) \cdot \mathbb{1}_{[\theta,+\infty)}(x_2) \cdot \ldots \cdot \mathbb{1}_{[\theta,+\infty)}(x_n)}$$



$$= \begin{cases} 1, & x_1 \geq \theta, x_2 \geq \theta, \ldots, x_n \geq \theta \\ 0, & \text{altfel} \end{cases}$$

$$= \begin{cases} 1, & \min_i(x_i) \geq \theta \\ 0, & \text{altfel} \end{cases}$$

Deci

$$L(\theta|x) = e^{n\theta-\sum_{i=1}^{n} x_i} \, \mathbb{1}_{[\theta,+\infty)}\left(\min_i(x_i)\right)$$

Dacă $\theta > \min_i(x_i) \Rightarrow L(\theta|x) = 0$

$\theta < \min_i(x_i) \Rightarrow L(\theta|x) = e^{-\sum_{i=1}^{n} x_i + n\theta}$

Obs. că $\frac{\partial \ln L(\theta|x)}{\partial \theta} = n \neq 0$ (Deci? - fct nu este derivabilă)

Observăm că pentru $\theta < \min_i (x_i)$ funcția $L(\theta|x)$, cu $L(\theta|x) = e^{n\theta - \sum_{i=1}^n x_i}$ este crescătoare (în θ)

deci maximul se atinge pt $\boxed{\hat{\theta}_n = \min_i (x_i)}$

Astfel $\hat{\theta}_n = \min_i (x_i)$ este estimatorul de verosimilitate maximă

Exp: (Estimatorul de verosimilitate maximă nu este unic)

Fie $X_1, X_2, \dots X_n \sim f_\theta$ cu

$$f_\theta(x) = e^{-\frac{|x-\theta|}{2}}, x \in \mathbb{R} \quad \text{(Repartiția Laplace)}$$

Funcția de verosimilitate:

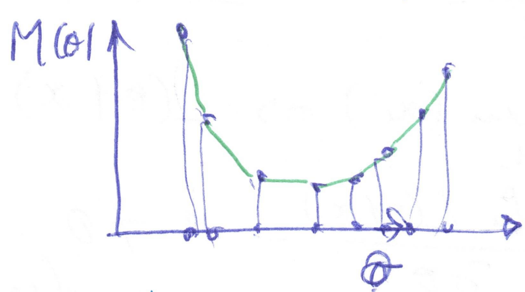
$$L(\theta|x_1 \dots x_n) = \prod_{i=1}^n f_\theta(x_i) = \prod_{i=1}^n e^{-\frac{|x_i-\theta|}{2}}$$

$$= e^{-\frac{1}{2}\sum_{i=1}^n |x_i-\theta|}$$

Cum e^x este crescătoare atunci funcția de verosimilitate $L(\theta|x)$ este maximă (în θ) atunci când $\sum_{i=1}^n |x_i-\theta|$ este minim

$$\arg\max_{\theta\in\mathbb{R}} L(\theta|x) = \arg\min_{\theta\in\mathbb{R}} \sum_{i=1}^n |x_i-\theta|$$

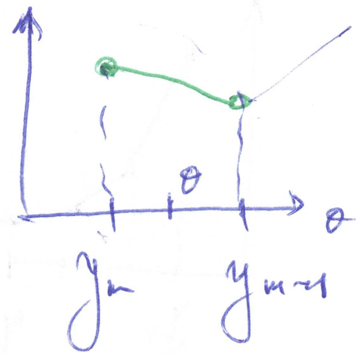
Fie $M(\theta) = \sum_{i=1}^n |x_i-\theta|$ — continuă și liniară pe porțiuni



Fie $y_1, y_2, \dots, y_n$ valorile lui $x_1, x_2, \dots x_n$ ordonate crescător astfel y_1 este $\min(x_i)$, $y_n = \max(x_i)$

Atunci $M(\theta) = \sum_{i=1}^{n} |x_i - \theta| = \sum_{i=1}^{n} |y_i - \theta|$

Dacă $y_m < \theta \leq y_{m+1}$ atunci



cum $y_i \leq y_m$ pt $i \leq m$

avem $|y_i - \theta| = \theta - y_i$

iar $y_j \geq y_{m+1}$, $m+1 \leq j \leq n$

$$|y_j - \theta| = y_j - \theta$$

$$|z| = \begin{cases} z, & z \geq 0 \\ -z, & z < 0 \end{cases}$$

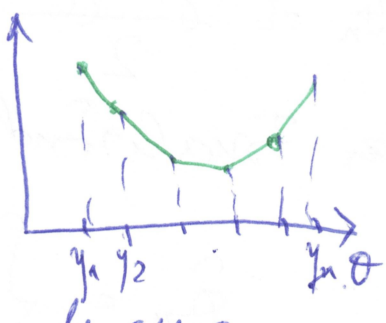
$$M(\theta) = \sum_{i=1}^{n} |y_i - \theta| = \sum_{i=1}^{m} |y_i - \theta| + \sum_{i=m+1}^{n} |y_i - \theta|$$

$$= \sum_{i=1}^{m} (\theta - y_i) + \sum_{i=m+1}^{n} (y_i - \theta)$$

f-ț cont de θ
și derivabilă

$$\frac{d}{d\theta} M(\theta) = m - (n-m) = 2m - n$$

$$\frac{d}{d\theta} M(\theta) = 0 \iff m = n/2$$



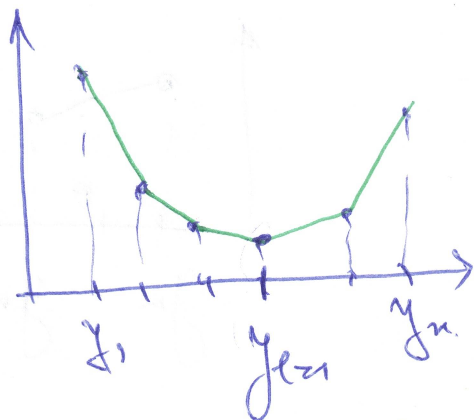
Dacă $m < n/2 \Rightarrow M'(\theta) < 0 \Rightarrow M$ este descresc.

$m > n/2 \Rightarrow M'(\theta) > 0 \Rightarrow M$ este cresc.

Dacă n este impar, $n = 2l+1 \Rightarrow \frac{n}{2} = l + \frac{1}{2}$

și atunci pt $\theta < y_{l+1}$ fct. $M(\theta)$ este $\downarrow$

$\theta > y_{l+1}$ fct. $M(\theta)$ este $\uparrow$.



În acest caz

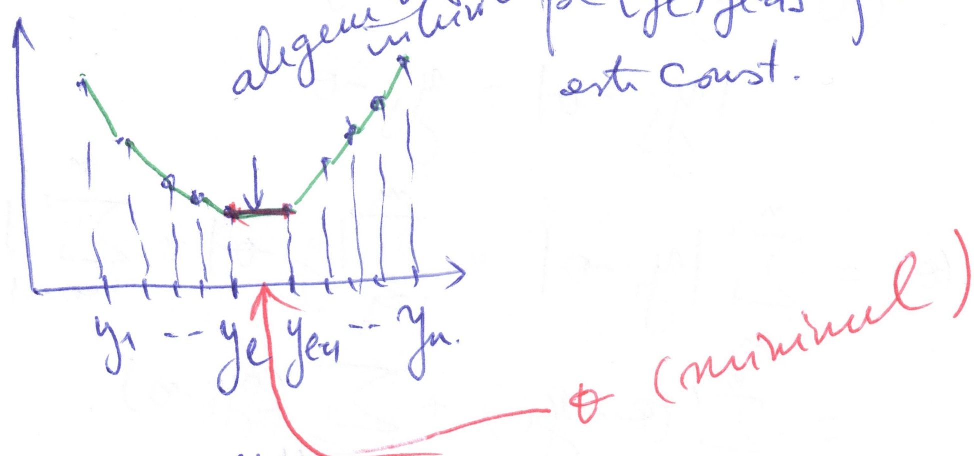
$$\arg\min \sum_{i=1}^{n} |x_i - \theta| = y_{l+1}.$$

$$l+1 = \frac{n-1}{2} + 1 = \frac{n+1}{2}$$

Dacă n este par, $n = 2l$ atunci

alegem mijlocul intervalului

pe $[y_l, y_{l+1}]$ fct $M(\theta)$ este const.



$$\hat{\theta}_n = \frac{y_l + y_{l+1}}{2} = \frac{y_{n/2} + y_{n/2+1}}{2}$$

~~Da~~ Finalizând

$$\hat{\theta}_n = \begin{cases} y_{\frac{n+1}{2}} & \text{, dacă } n \text{ este impar} \\ \dfrac{y_{n/2} + y_{n/2+1}}{2} & \text{, dacă } n \text{ este par} \end{cases}$$

mediana eșantionului

Intervale de încredere

Obs: Dacă până acum, în partea de estimare punctuală am determinat plecând de la un eșantion o valoare cât mai apropiată de parametrul necunoscut care a generat observațiile, în această secțiune ne propunem să găsim un interval de valori astfel încât probabilitatea ca acest interval să conțină adevărata valoare a parametrului să fie cât mai mare.

Def: Fie $X_1, X_2, \dots X_n \sim f_\theta$, $\theta \in \Theta \subseteq \mathbb{R}$ și $\alpha \in (0,1)$ iar funcțiile $A_\alpha, B_\alpha : \mathbb{R}^n \longrightarrow \mathbb{R}$ a.î.

$$A_\alpha(x_1, \dots x_n) \leq B_\alpha(x_1, \dots x_n), \quad \forall (x_1, \dots x_n) \in \mathbb{R}^n$$

Intervalul $[A_\alpha(X_1, X_2, \dots X_n), B_\alpha(X_1, X_2, \dots X_n)]$ s.n. interval de încredere pt. θ cu coeficientul de încredere $1-\alpha$:

$$\mathbb{P}_\theta\left([A_\alpha(X_1, \dots X_n), B_\alpha(X_1, \dots X_n)] \ni \theta\right) \geq 1-\alpha, \quad \forall \theta \in \Theta$$

Obs: Intervalul $[A_\alpha(X_1 \dots X_n), B_\alpha(X_1, \dots X_n)]$ este un interval aleator! Nu θ este aleator! bilateral

Obs: În practică $\alpha = 0.05$ și atunci $1-\alpha = 0.95$

Obs: Este posibil să fim interesați de intervale unilaterale de tipul $(-\infty, B_\alpha(X_1, \dots X_n)]$ sau $[A_\alpha(X_1, \dots X_n), +\infty)$

Obs: Dacă am observat $x_1, x_2, \dots x_n$ atunci găsim $A_\alpha(x_1, \dots x_n)$ și $B_\alpha(x_1, \dots x_n)$ valori numerice și în acest caz vb. de intervalul de încredere de nivel de încredere $1-\alpha$ pt θ

$$IC_{1-\alpha}(\theta) = [A_\alpha(x_1, \dots x_n), B_\alpha(x_1, \dots x_n)]$$

Metoda pivotului:

Def: Fie $X_1, X_2, \dots X_n \sim f_\theta$, $\theta \in \Theta \subset \mathbb{R}$. O funcție $g: \mathbb{R}^n \times \Theta \to \mathbb{R}$ se numește fct pivot dacă

a) Repartiția lui $g(\underbrace{X_1, \dots X_n, \theta}_{\mathbb{R}^n \times \Theta})$ nu depinde de θ

b) Pentru orice $u_1 \le u_2$, $u_1, u_2 \in \mathbb{R}$ inecuația

$$u_1 \le g(x_1, \dots x_n, \theta) \le u_2$$

se poate rezolva în θ conducând la sol:

$$a(x_1, \dots x_n) \le \theta \le b(x_1, \dots x_n)$$

Def: Fie $p \in (0,1)$. Se numește cuantila de ordin p asociată fct. de rep. F, valoarea x_p care verifică

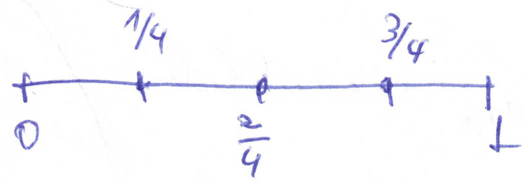
$$x_p = F^{-1}(p) = \inf\{x \mid F(x) \ge p\}$$

Obs: ~~Ex~~ Cuantila de ordin p este acea valoare pentru care $p\%$ dintre obs. sunt mai mici decât ea și $(1-p)\%$ dintre obs sunt mai mari

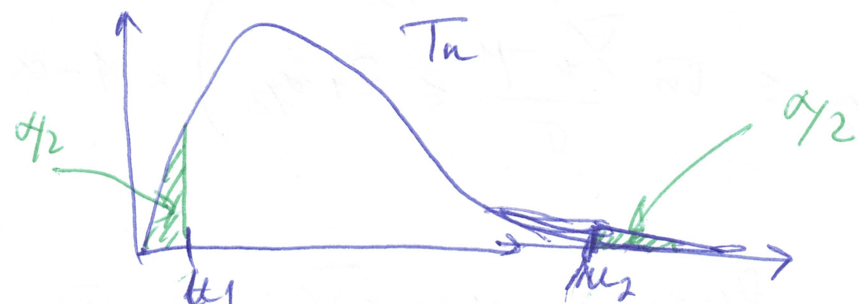
Dacă $p = 1/2$ atunci cuantila de ordin $1/2$ s.n. <u>mediană</u>

$x_{1/2}$ – mediana (q_2) : $\mathbb{P}(X \le x_{1/2}) = \mathbb{P}(X > x_{1/2}) = \frac{1}{2}$

Dacă $p = 1/4$ (resp $3/4$) atunci cuantila de ordin $1/4$ s.n. prima <u>cuartilă</u> q_1 (resp. a 3-a cuartilă q_3)



<u>Ideea</u>: – considerăm un estimator pt θ, T_n, pt care cunoaștem cum este repartizat (nu depinde de θ)



– considerăm u_1 și u_2 cuantilele de ordin $\alpha/2$ și resp. $1-\alpha/2$

– intervalul de încredere va fi

$$\mathbb{P}_\theta([u_1, u_2] \ni \theta) = 1-\alpha$$

<u>Intervale de încredere pt populații normale</u>

a) <u>Int. de încredere pt media unei pop. normale atunci când varianța este cunoscută</u>

Fie $X_1, X_2, \dots, X_n \sim N(\mu, \sigma^2)$ cu μ – necunoscut, σ^2 – cunoscut

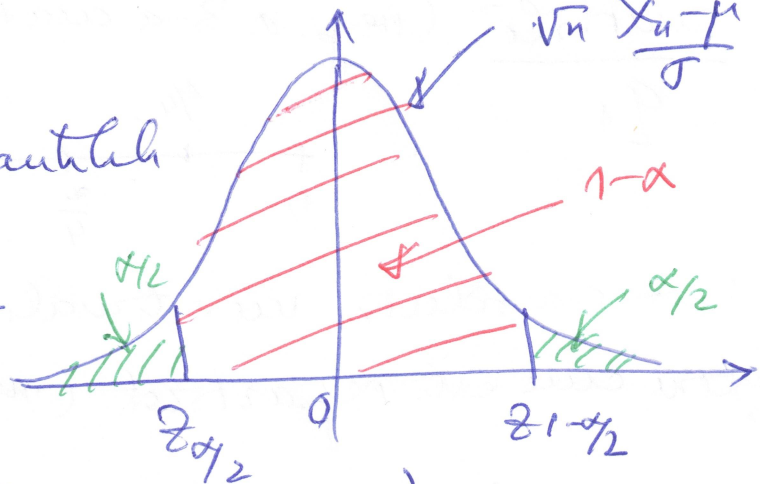
Am văzut că $\bar{X}_n$, estimatorul pt medie, verifică

$$\bar{X}_n \sim N(\mu, \sigma^2/n)$$

$$\Rightarrow \quad \sqrt{n}\,\frac{\bar{X}_n - \mu}{\sigma} \sim N(0,1)$$

Funcția $\underbrace{g(x_1, \ldots, x_n, \mu)}_{T_n} = \sqrt{n}\,\frac{\bar{x}_n - \mu}{\sigma}$ este o fct pivot

Fie $z_{\alpha/2}$ și $z_{1-\alpha/2}$ cuantilele de ordin $\alpha/2$ și resp $1-\alpha/2$ din $N(0,1)$



$$\mathbb{P}_\mu\left(z_{\alpha/2} \le \sqrt{n}\,\frac{\bar{X}_n - \mu}{\sigma} \le z_{1-\alpha/2} \right) = 1-\alpha$$

$$\Leftrightarrow \mathbb{P}_\mu\left(\bar{X}_n - z_{1-\alpha/2}\cdot\frac{\sigma}{\sqrt{n}} \le \mu \le \bar{X}_n - z_{\alpha/2}\,\sigma/\sqrt{n} \right) = 1.$$

În cazul unei rep. normale $z_{\alpha/2} = -z_{1-\alpha/2}$ (din simetrie)

$$\boxed{IC^{1-\alpha}(\mu) = \left[\bar{x}_n - z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}},\ \bar{x}_n + z_{1-\alpha/2}\,\sigma/\sqrt{n} \right]}$$

b) Int. de încredere pentru media unei pop. normale atunci când varianța este necunoscută

Fie $X_1, X_2, \dots X_n \sim N(\mu, \sigma^2)$, μ, σ^2 - necunoscute

În acest caz am văzut că statistica

$$T_n = \sqrt{n}\,\frac{\bar{X}_n - \mu}{S_n} \sim t_{n-1} \quad (t\text{-Student})$$

Funcția $g(x_1, \dots x_n, \mu) = \sqrt{n}\cdot\frac{\bar{x}_n - \mu}{s_n}$ este o fct pivot

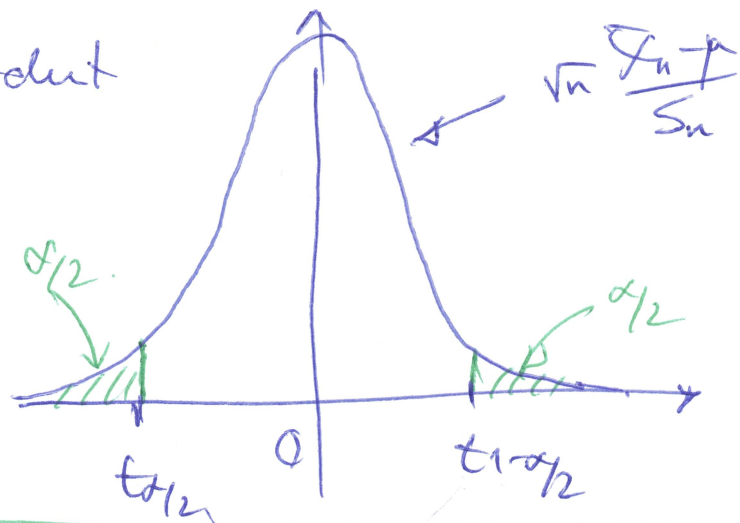
Pt. $\alpha \in (0,1)$ fie $t_{\alpha/2}$ și $t_{1-\alpha/2}$ cuantilele de ordin $\alpha/2$ și resp. $1-\alpha/2$ din rep. Student cu $n-1$ grade de libertate

Atunci

$$P_\mu\left(t_{\alpha/2} \le \sqrt{n}\,\frac{\bar{X}_n - \mu}{S_n} \le t_{1-\alpha/2}\right) = 1-\alpha$$

Din simetria rep. Student

$t_{\alpha/2} = -t_{1-\alpha/2}$



$$\boxed{IC^{1-\alpha}(\mu) = \left[\bar{x}_n - t_{1-\alpha/2}\frac{s_n}{\sqrt{n}},\ \bar{x}_n + t_{1-\alpha/2}\frac{s_n}{\sqrt{n}}\right]}$$