Eroarea pătratică medie
(mean square error)

Def: Fie $X_1, X_2, ..., X_n$ din f_θ și $\hat{\theta}_n$ estimator. Definim eroarea pătratică:

$$MSE_\theta(\hat{\theta}_n) = E_\theta[(\hat{\theta}_n - \theta)^2]$$

Spunem că estimatorul $\hat{\theta}_1$ este mai bun decât estimatorul $\hat{\theta}_2$ (în sensul MSE) dacă $MSE_\theta(\hat{\theta}_1) < MSE_\theta(\hat{\theta}_2)$

(P) Avem că $\boxed{MSE_\theta(\hat{\theta}) = Var_\theta(\hat{\theta}) + b_\theta(\hat{\theta})^2}$

În particular dacă $\hat{\theta}$ nedeplasat pt θ, atunci

$MSE_\theta(\hat{\theta}) = Var_\theta(\hat{\theta})$

Obs: În general vrem să găsim estimatori pt. care MSE este cât mai mică.

Ex: $X_1, ..., X_n \sim N(\mu, \sigma^2)$ și scopul este să estimăm varianța σ^2.

Avem statistica $T_n = \sum_{i=1}^{n} (X_i - \bar{X}_n)^2$ și căutăm un estimator de forma $\hat{\sigma}_a^2 = aT_n$, $a = a(n) > 0$.

Până acum am văzut $\begin{cases} a = \frac{1}{n} \to \text{varianța empirică } V_n^2 \\ a = \frac{1}{n-1} \to \text{varianța eșantionului} \end{cases}$

Vreau să determin a pentru care $MSE_{\sigma^2}(\hat{\sigma}_a^2)$ minimă

$$MSE_{\sigma^2}(\hat{\sigma}_a^2) = Var_{\sigma^2}(\hat{\sigma}_a^2) + b_{\sigma^2}(\hat{\sigma}_a^2)^2$$

Reamintim în cazul populației normale

$$\frac{(n-1)S_n^2}{\sigma^2} = \frac{\sum_{i=1}^{n}(X_i - \bar{X}_n)^2}{\sigma} \sim \chi^2(n-1)$$

$\Rightarrow \frac{T^n}{\sigma^2} \sim \chi^2(n-1)$

Știm $\begin{cases} E[\chi^2(\nu)] = \nu \\ Var(\chi^2(\nu)) = 2\nu \end{cases}$

$E_{\sigma^2}\left(\frac{T_n}{\sigma^2}\right) = n-1 \Rightarrow E_{\sigma^2}[T_n] = (n-1)\sigma^2$

$Var_{\sigma^2}\left(\frac{T_n}{\sigma^2}\right) = 2(n-1) \Rightarrow Var_{\sigma^2}(T_n) = 2(n-1)\sigma^4$

Astfel

$$MSE_{\sigma^2}(\hat{\sigma}_a^2) = Var_{\sigma^2}(\overbrace{aT_n}^{\hat{\sigma}_a^2}) + b_{\sigma^2}(aT_n)^2 =$$
$$= a^2 Var_{\sigma^2}(T_n) + [E_{\sigma^2}(aT_n) - \sigma^2]^2 =$$
$$= 2a^2(n-1)\sigma^4 + [a(n-1)\sigma^2 - \sigma^2]^2 =$$
$$= \sigma^4[2a^2(n-1) + (a(n-1)-1)^2].$$

Derivând după a obținem

$$\frac{d}{da} MSE_{\sigma^2}(\hat{\sigma}_a^2) = \sigma^4[4a(n-1) + 2(a(n-1)-1)(n-1)]$$
$$= 2\sigma^4(n-1)[2a + a(n-1) - 1]$$

$$\frac{d}{da} MSE_{\sigma^2}(\hat{\sigma}_a^2) = 0 \Leftrightarrow 2a + a(n-1) - 1 = 0 \Rightarrow a = 1/n+1$$

Prin urmare printre estimatorii lui σ^2 de forma $a\sum_{i=1}^{n}(X_i - \overline{X}_n)^2$, estimatorul cu MSE cea mai mică este $\frac{1}{n+1}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2$

Estimatorul $S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2$ plătește un preț pentru a fi nedeplasat.

① Dacă $MSE_\theta(\hat{\theta}_n) \to 0$ pt. $n \to \infty$, atunci $\hat{\theta}_n$ e un estimator consistent pentru θ

~~Ex~~

Ex. Compararea a 2 estimatori nedeplasați după MSE

Fie $X_1, X_2, \dots, X_n \sim Pois(\theta)$, și $\hat{\theta}_1 = \overline{X}_n$, și $\hat{\theta}_2 = S_n^2$ doi estimatori nedeplasați pentru θ. Vrem să determinăm care este mai bun din punctul de vedere al MSE.

$$MSE_\theta(\hat{\theta}_1) \ (?) \ MSE_\theta(\hat{\theta}_2)$$
$$\overset{\shortparallel}{Var_\theta(\hat{\theta}_1)} \qquad \overset{\shortparallel}{Var_\theta(\hat{\theta}_2)}$$

Metode de construcție a estimatorilor

1) Metoda momentelor:

- sfârșitul sec. XIX, începutul sec. XX (Karl Pearson)

Fie $X_1, X_2, \dots, X_n$ un eșantion de volum n dintr-o populație de densitate f_θ, $\theta \in \Theta \subseteq \mathbb{R}^k$, $\theta = (\theta_1, \theta_2, \dots, \theta_k)$

Dacă $X \sim f_\theta$, atunci momentele de ordin j, $1 \le j \le k$ sunt $E[X^j] = \int_{-\infty}^{+\infty} x^j f_\theta(x)\,dx$, $j \in \{1, 2, \dots, k\}$

În general

$$E[X^1] = g_1(\theta_1, \dots, \theta_k)$$
$$\dots$$
$$E[X^k] = g_k(\theta_1, \dots, \theta_k)$$

Metoda momentelor presupune rezolvarea sistemului următor:

$$\begin{cases} \frac{1}{n}\sum_{i=1}^{n} X_i = g_1(\theta_1, \dots, \theta_k) \\ \dots \\ \frac{1}{n}\sum_{i=1}^{n} X_i = g_k(\theta_1, \dots, \theta_k) \end{cases}$$

Soluția este estimatorul obținut prin metoda momentelor.

Exp: $X_1, X_2, \ldots, X_n \sim Exp(\lambda)$, $\theta = \lambda \in (0, +\infty)$

Cum $\theta \in \mathbb{R} \Rightarrow k = 1$

Momentul empiric de ordin 1 = Momentul teoretic de ordin 1.

$$\overline{X_n} = E_\lambda[X_1] = \frac{1}{\lambda} \Rightarrow \boxed{\tilde{\lambda}_n = \frac{1}{\overline{X_n}}}$$

Exp: Fie $x_1 = 0{,}42$, $x_2 = 0{,}10$, $x_3 = 0{,}65$, $x_4 = 0{,}23$ un ~~eșat~~ eșantion de volum $n = 4$ și populație

$$f_\theta(x) = \theta x^{\theta - 1}, \quad 0 \le x \le 1$$

$\theta \in (0, 1] \subseteq \mathbb{R} \Rightarrow k = 1$

Met. momentelor presupune rezolvarea:

$$\overline{X_n} = E_\theta[X_1] = \int_{-\infty}^{+\infty} x f_\theta(x)\,dx = \int_0^1 x\,\theta x^{\theta-1}\,dx = \frac{\theta}{\theta+1}$$

$$\Rightarrow \overline{X_n} = \frac{\tilde{\theta}_n}{\tilde{\theta}_n - 1} \Rightarrow \boxed{\tilde{\theta}_n = \frac{\overline{X_n}}{1 - \overline{X_n}}} \qquad \overline{X_4} = 0{,}35$$

$$\tilde{\theta}_4 = \frac{0{,}35}{1 - 0{,}35} = 0{,}54$$

Exp: $X_1, X_2, \ldots, X_n \sim N(\mu, \sigma^2)$

În acest caz $\theta = (\theta_1, \theta_2) = (\mu, \sigma^2)$

$\theta \in \Theta = \mathbb{R} \times (0, \infty) \subseteq \mathbb{R}^2$, $k = 2$

Sistem cu 2 ecuații și 2 necunoscute

$$\begin{cases} \overline{X_n} = E_\theta[X_1] \\ \frac{1}{n}\sum_{i=1}^{n} X_i^2 = E_\theta[X_1^2] \end{cases}$$

Dacă $X_1 \sim N(\mu, \sigma^2) \Rightarrow E_\theta[X_1] = \mu$

$E_\theta[X_1^2] = \sigma^2 + \mu^2$

Sistemul devine

$$\begin{cases} \overline{X_n} = \mu \\ \frac{1}{n}\sum_{i=1}^{n} X_i^2 = \sigma^2 + \mu^2 \end{cases} \Rightarrow \begin{cases} \tilde{\mu}_n = \overline{X_n} \\ \tilde{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \overline{X_n}^2 \end{cases}$$

$$\Rightarrow \begin{cases} \tilde{\mu}_n = X_n \\ \tilde{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^{n} (X_i - \overline{X_n})^2 \end{cases}$$

<u>Obs</u>: Ce se întâmplă în cazul unei repartiții $B(k,p)$

$$\begin{cases} \overline{X_n} = kp \\ \frac{1}{n}\sum_{i=1}^{n} X_i^2 = kp(1-p) + k^2p^2 \end{cases} \qquad \begin{cases} p \in (0,1) \\ k \in \mathbb{N} \end{cases}$$

2) Metoda verosimilității maxime (Maximum likelyhood)

Fie $X_1, X_2, \dots, X_n \sim f_\theta(x)$, $\theta \in \Theta \subseteq \mathbb{R}^k$

Densitatea (masa) comună a $(X_1, X_2, \dots, X_n)$ este

$$f_\theta(X_1, X_2, \dots, X_n) = \prod_{i=1}^{n} f_\theta(X_i)$$

Def: Se numește funcție de verosimilitate asociată eșantionului $X_1, X_2, \dots, X_n$ funcția

$$L(\theta \mid x_1, x_2, \dots, x_n) = f_\theta(X_1, \dots, X_n) = \prod_{i=1}^{n} f_\theta(x_i)$$

și este văzută ca o funcție de θ

$$l(\theta \mid x_1, x_2, \dots, x_n) = \log L(\theta \mid x_1 \dots x_n)$$

<u>Obs</u>: uneori putem întâlni doar notația $L(\theta)$, și $l(\theta)$ sau $L(\theta \mid x)$, și $l(\theta \mid x)$, $x = (x_1, \dots, x_n)$

<u>Obs</u>: Funcția de verosimilitate nu este o densitate de probabilitate pentru θ

<u>Obs</u>: Dacă $L(\theta_1, x) > L(\theta_2, x)$ (echivalent cu)

$$l(\theta_1, x) > l(\theta_2, x)$$

atunci spunem că θ_1 este mai probabil decât θ_2 să fi produs $X_1, X_2, \dots, X_n$

Cu alte cuvinte f_{θ_1} reprezintă un model mai bun decât f_{θ_2} în ceea ce privește fixarea datelor observate

Def: Fie $X_1, X_2, \dots, X_n \sim f_\theta(x)$ și $L(\theta/x)$, $\ell(\theta/x)$ sunt funcția de verosimilitate, respectiv log fct de verosimilitate

Numim estimator de verosimilitate maximă (MLE) pentru θ

$$\hat{\theta}_n = \underset{\theta \in \Theta}{\arg\max}\, L(\theta/x) = \underset{\theta \in \Theta}{\arg\max}\, \ell(\theta/x)$$

$$\hat{\theta}_n = \hat{\theta}_n(X_1, X_2, \dots, X_n)$$

Obs: Dacă funcția de verosimilitate este diferențiabilă, atunci posibilii candidați pentru ~~estimatorul~~ MLE sunt soluțiile sistemului

$$\frac{\partial}{\partial \theta_i} L(\theta/x) = 0, \quad i = 1, \dots, k$$

Exp: $X_1, X_2, \dots, X_n \sim \text{Bernoulli}(\theta)$, $\theta \in (0,1)$

$$L(\theta/x) = \prod_{i=1}^{n} f_\theta(x_i)$$

$$f_\theta(x) = \begin{cases} \theta, & x = 1 \\ 1-\theta, & x = 0 \end{cases} = \theta^{x}(1-\theta)^{1-x}$$

$$L(\theta/x) = \prod_{i=1}^{n} \theta^{x_i}(1-\theta)^{1-x_i}$$

$$= \theta^{\sum_{i=1}^{n} x_i}(1-\theta)^{n-\sum_{i=1}^{n} x_i}$$

Avem $\ell(\theta/x) = \log L(\theta/x) =$

$$= \left(\sum_{i=1}^{n} x_i\right)\log\theta + \left(n - \sum_{i=1}^{n} x_i\right)\log(1-\theta)$$

Dacă $0 < \sum_{i=1}^{n} x_i < n$ atunci derivând

$$\frac{d}{d\theta} l(\theta/x) = \frac{\sum_{i=1}^{n} x_i}{\theta} - \left(n - \sum_{i=1}^{n} x_i\right)\frac{1}{1-\theta} = 0$$

$$\Leftrightarrow \frac{\sum_{i=1}^{n} x_i}{\theta} = \frac{1}{1-\theta}\left(n - \sum_{i=1}^{n} x_i\right)$$

$$\Leftrightarrow \frac{\overline{x_n}}{\theta} = \frac{1}{1-\theta}(1 - \overline{x_n}) \Rightarrow \hat{\theta}_n = \overline{x_n}$$

Dacă $\sum_{i=1}^{n} x_i = 0$ sau $\sum_{i=1}^{n} x_i = n$

$$l(\theta/x) = \begin{cases} n\log(1-\theta), & \sum_{i=1}^{n} x_i = 0 \\ n\log(\theta), & \sum_{i=1}^{n} x_i = n \end{cases}$$

și $l(\theta/x)$ monotonă în θ, și maximul se atinge ~~în~~

~~în cazul~~ pt $\left.\begin{array}{ll} \theta = 0 & \text{dacă } \sum x_i = 0 \\ \theta = 1 & \text{dacă } \sum x_i = n \end{array}\right) \Rightarrow \hat{\theta}_n = \overline{x_n}$

<u>Obs</u>: Determinați estimatorul prin metoda momentelor pentru acest caz

Exp: $X_1, X_2, \ldots, X_n \sim N(\mu, \sigma^2)$ cu μ și σ^2 necunoscute

$\theta = (\theta_1, \theta_2) = (\mu, \sigma^2)$

Funcția de verosimilitate:

$$L(\mu, \sigma^2 / x) = \prod_{i=1}^{n} f_\theta(x_i)$$

$$f_\theta(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

$$L(\mu, \sigma^2/x) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{1}{2}\left(\frac{x_i-\mu}{\sigma}\right)^2}$$

$$= \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{n}{2}} e^{-\frac{1}{2}\sum_{i=1}^{n}\left(\frac{x_i-\mu}{\sigma}\right)^2}$$

$$\ell(\mu, \sigma^2 / x) = \log L(\mu, \sigma^2 / x)$$

$$= -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log\sigma^2 - \frac{1}{2}\sum_{i=1}^{n}\frac{(x_i-\mu)^2}{\sigma^2}$$

Avem sistemul

$$\begin{cases} \frac{\partial}{\partial\mu}\ell(\mu, \sigma^2 / x) = 0 \\ \frac{\partial}{\partial\sigma^2}\ell(\mu, \sigma^2 / x) = 0 \end{cases}$$

$$\hat{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2$$

Prin urmare $\overline{X_n}$ și $\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X}_n)^2$ sint estimatorii de verosimilitate maximă.

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

$$f_\theta(x) = \begin{cases} e^{-(x-\theta)}, & x \geq \theta,\ \theta > 0 \\ - \end{cases}$$

Funcția de verosimilitate:

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

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

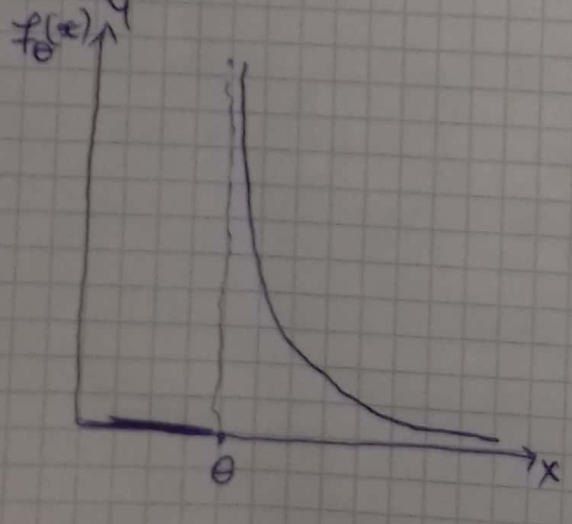
$$= e^{n\theta - \sum_{i=1}^{n} x_i} \cdot \underbrace{1\!\!1_{[\theta,+\infty)}(x_1) \cdot 1\!\!1_{[\theta,+\infty)}(x_2) \cdot \dots \cdot 1\!\!1_{[\theta,+\infty)}(x_n)}$$

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

Deci $L(\theta/x) = e^{n\theta - \sum_{i=1}^{n} x_i} \cdot 1\!\!1_{[\theta,+\infty)}(\min(x_i))$

Dacă $\theta > \min(x_i) \Rightarrow L(\theta/x) = 0$

altfel $L(\theta/x) = e^{n\theta - \sum_{i=1}^{n} x_i}$



Obs. că $\dfrac{\partial \ln L(\theta/x)}{\partial \theta} = n \neq 0$ (funcția nu e derivabilă)

Observăm că pentru $\theta < \min(x_i)$ 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 pentru $\hat{\theta}_n = \min(x_i)$

Exp: Estimatorul de verosimilitate maximă nu este unic

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

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

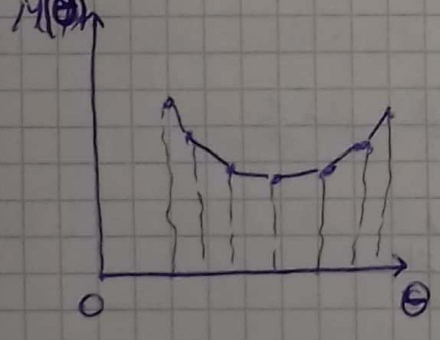
$$L(\theta / x_1 \dots x_n) = \prod_{i=1}^{n} f_\theta(x_i) = \prod_{i=1}^{n} e^{-\frac{|x_i-\theta|}{2}}$$

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

Cum e^x crescătoare $\Rightarrow L(\theta/x)$ este maximă atunci când $\sum_{i=1}^{n}|x_i-\theta|$ minim

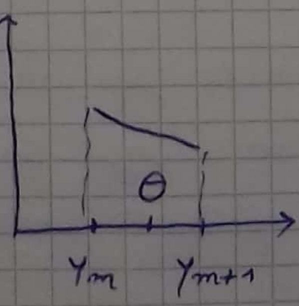
$$\underset{\theta \in \mathbb{R}}{\operatorname{argmax}} \; L(\theta/x) = \underset{\theta \in \mathbb{R}}{\operatorname{argmin}} \sum_{i=1}^{n}|x_i-\theta|$$

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



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

Dacă $y_m < \theta \le y_{m+1}$



atunci cum $y_i \le y_m$ pt $i \le m$

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

$y_j \ge y_{m+1}$

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

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

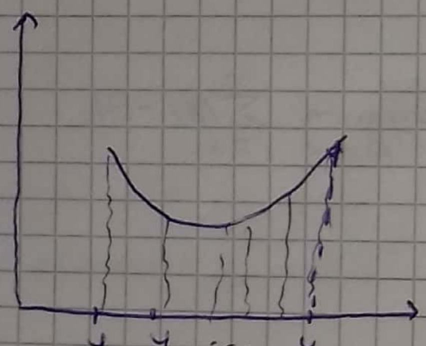
$$= \sum_{i=1}^{m} (\theta - y_i) + \sum_{i=m+1}^{n} (y_i - \theta)$$

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

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

Dacă $m < \frac{n}{2} \Rightarrow M'(\theta) < 0 \Rightarrow M$ descrescătoare

$m > \frac{n}{2} \Rightarrow M'(\theta) > 0 \Rightarrow M$ crescătoare



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

și atunci pt $\theta < y_{\ell+1}$ $M(\theta)$ este $\searrow$

$\theta > y_{\ell+1}$, $M(\theta)$ este $\nearrow$

În acest caz $\arg\min \sum_{i=1}^{n} |x_i - \theta| = y_{\ell+1}$

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

Dacă n este par, $n = 2\ell$ atunci minimul se atinge în orice punct din $[y_\ell, y_{\ell+1}]$

Alegem mijlocul intervalului

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

Finalizând, ~~estimatorul~~

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

mediana eșantionului (empirică a eșantionului)

Intervale de încredere:

Obs: Dacă până acum în partea de estimare punctuală am ~~plecat~~ determinat plecând de la un eșantion cu o valoare cât mai apropiată de parametrul necunoscut care a generat observația, în această secțiune ne propunem să găsim un interval de valori a.î. 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 $\alpha \in (0,1)$, ~~și~~ iar funcțiile $A_\alpha, B_\alpha : \mathbb{R}^n \to \mathbb{R}$ ~~ai~~ a.î.

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

Intervalul $[A_\alpha(x_1 \dots x_n), B_\alpha(x_1, \dots, x_n)]$ s.n. interval de încredere pt. θ cu coeficientul de încredere $1-\alpha$

$$\mathbb{P}_\theta([A_\alpha(X_1 \dots X_n), B_\alpha(X_1 \dots X_n)] \ni \theta) \ge 1-\alpha$$

Intervalul este aleator, nu θ

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 $[A_\alpha(X_1, \dots X_n), +\infty)$

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

Metoda pivotului.

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

a) Repartiția lui $g(X_1,\dots,X_n,\theta)$ nu depinde de θ

b) Pentru orice $u_1 \leq u_2$, $u_1, u_2 \in \mathbb{R}$ inecuația $u_1 \leq g(X_1,\dots,X_n,\theta) \leq u_2$ se poate rezolva în θ conducând la soluția

$$a(X_1 \dots X_n) \leq \theta \leq b(X_1 \dots X_n)$$

Def: Fie $p \in (0,1)$. Se numește cuantilă de ordin p asociată funcției de repartiție F valoarea x_p care verifică

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

Obs: Cuantila de ordin p este acea valoare pentru care $p\%$ din observații sunt mai mici și $1-p$ mai mari

Dacă $p = \frac{1}{2}$, atunci cuantila de ordin $\frac{1}{2}$ e mediană.

$x_{1/2}$ - mediana (Q_2)

$$\mathbb{P}(X \leq x_{1/2}) = \mathbb{P}(X > x_{1/2}) = \frac{1}{2}$$

Dacă $p = \frac{1}{4}$ (resp. $\frac{3}{4}$), atunci cuantila de ordin $\frac{1}{4}$ se numește prima cuantila / a treia cuantilă
Q_1 / Q_3

Ideea: Un estimator pt θ pt. care cunoaștem cum este repartizat (nu depinde de θ)

u_1 și u_2 cuantilele de ordin $\frac{\alpha}{2}$ și $1-\frac{\alpha}{2}$

Intervalul de încredere va fi $[u_1, u_2]$ cu probabilitatea $1-\alpha$

Intervale de încredere pentru populații normale

a) Int. înc. pentru media unei populații normale atunci când varianța este cunoscută

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

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

$$\bar{X}_n \sim N(\mu, \sigma^2/n) \Rightarrow \sqrt{n}\cdot\frac{\bar{X}_n - \mu}{\sigma} \sim N(0,1)$$

Funcția $\underbrace{g(X_1, \dots, X_n, \mu)}_{T_n} = \sqrt{n}\cdot\frac{\bar{X}_n - \mu}{\sigma}$ este o fct. pivot

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

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

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

$z_{\alpha/2} = z_{1-\alpha/2}$ în N

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

b) varianța necunoscută

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

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

Funcția $g(X_1, \dots, X_n, \mu) = \sqrt{n}\cdot\frac{\bar{X}_n - \mu}{S_n}$ este o funcție pivot.

Pt. $\alpha \in (0,1)$ fie $t_{\alpha/2}$ și $t_{1-\alpha/2}$ cuantilele de ordin $\alpha/2$ și $1-\alpha/2$. Atunci

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

Din simetria repartiției student: $t_{\alpha/2} = -t_{1-\alpha/2}$

$$\underline{I}C^{1-\alpha}(\mu) = \left[\overline{X_n} - t_{1-\alpha/2}\frac{S_n}{\sqrt{n}}, \overline{X_n} + t_{1-\alpha/2}\frac{S_n}{\sqrt{n}}\right]$$

Consultație

① $X \sim U([0,1])$