Consultații Statistică - 8.05.2023

$$MSE_\theta(T_n) = Var_\theta(T_n) + \underbrace{b_\theta(T_n)^2}_{0}$$

$$\sqrt{n}\,(\underset{\bar{X}_n}{T_n} - h(\theta)) \xrightarrow{d} \mathcal{N}(0, \underbrace{Var(T_n)}_{Var(X_1)})$$

Calculați MSE, comparați estimatorii

TLC $\quad T_n = \sum Y_i \rightarrow E[Y_1] = \mu;\ Var(Y_1) = \sigma^2$

$$\frac{T_n - E[T_n]}{\sqrt{Var(T_n)}} \rightarrow \mathcal{N}(0,1)$$

$$T_n \sim \mathcal{N}(\underbrace{E[T_n]}_{n\mu}, \underbrace{Var(T_n)}_{n\sigma^2})$$

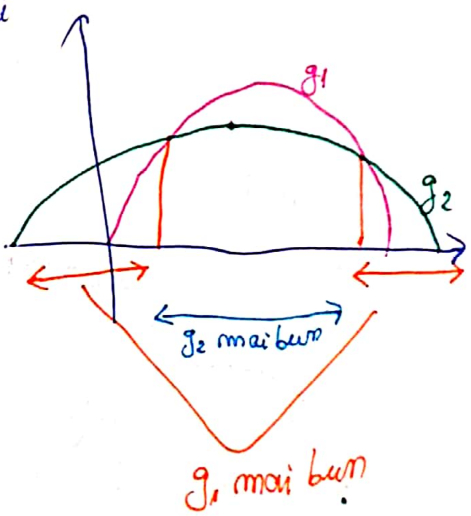
Dacă T_n este EVM pt. $h(\theta)$ atunci $\sqrt{n}\,(T_n - h(\theta)) \xrightarrow{d} \mathcal{N}\left(0, \frac{1}{I_1(h(\theta))}\right)$

Avem doi estimatori $\hat{\theta}_1, \hat{\theta}_2 \rightarrow$ comparați estimatorii

$MSE_\theta(\theta_1) = g_1(\theta)$

$MSE_\theta(\theta_2) = g_2(\theta)$

$g_1(\theta) \geq g_2(\theta) \rightarrow$ pe ce intervale?



Exemplu

$X \sim B(10, \theta)$, $\hat{\theta}_1 = \frac{X}{10}$, $\hat{\theta}_2 = \frac{X+1}{12}$

$\theta \in (0,1)$

$E_\theta[\hat{\theta}_1] = \frac{1}{10} E_\theta[X] = \frac{10\theta}{10} = \theta$ nedeplasat $\Rightarrow$

$\Rightarrow b_\theta(\hat{\theta}_1) = 0$

$E_\theta[\hat{\theta}_2] = \frac{E_\theta[X+1]}{12} = \frac{10\theta+1}{12}$

$b_\theta(\hat{\theta}_2) = \frac{10\theta+1}{12} - \theta$

$Var_\theta(\hat{\theta}_1) = Var_\theta\left(\frac{X}{10}\right) = \frac{1}{100} Var_\theta(X) = \frac{1}{100} \cdot 10\theta(1-\theta) = \frac{\theta(1-\theta)}{10}$

$MSE_\theta(\hat{\theta}_1) = \frac{\theta(1-\theta)}{10}$ (este un est. nedeplasat)

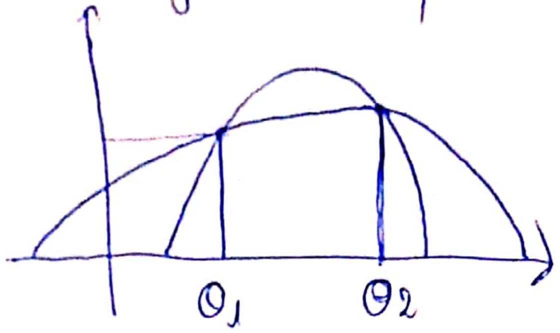
$Var_\theta(\hat{\theta}_2) = \frac{1}{144} Var_\theta(X+1) = \frac{1}{144} Var_\theta(X) = \frac{1}{144} 10\theta(1-\theta)$

$MSE_\theta(\hat{\theta}_2) = \frac{10\theta(1-\theta)}{144} + \left(\frac{10\theta+1}{12} - \theta\right)^2 \rightarrow$ media

$MSE_\theta(\hat{\theta}_1) = MSE_\theta(\hat{\theta}_2)$ ec. de grad 2 $\Rightarrow \theta_{1,2}$

[redacted]

asimptotic → ne uităm la varianță
→ cel cu varianța mai mică



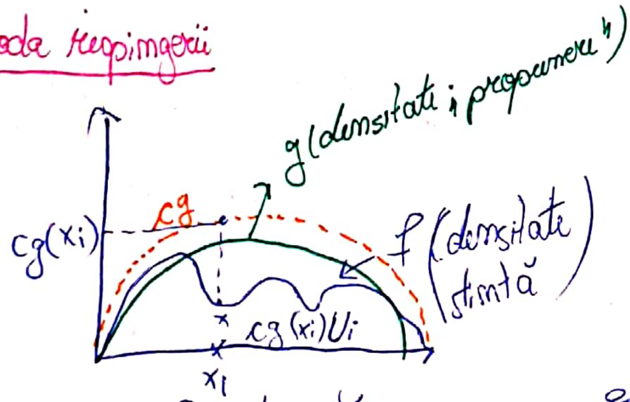
Metoda directă $X \sim f_\theta(x)$

Să calculăm $F_\theta(x)$

- Să calculăm fct. cuantilă $F_\theta^{-1}(u)$
- Th. de universalitate a rep. uniforme

gen $U \sim \mathcal{U}[0,1] \Rightarrow F_\theta^{-1}(U)$ care au aceeași rep. cu X

Metoda respingerii



Pas 1 Găsiți o densitate din care știți să generați observații („propunere")

Vrem $Y \sim f$. Știm să gen. $X \sim g$. Există const. $c \geq 1$ a.î. $f \leq cg$

Trebuie să det. const. c.

În R: trasare grafic funcție
metoda directă/respingere
histogramă

Estimatorul de verosimilitate max. atunci când f_θ, $x \in [0,\theta]$

$$X_1, X_2, \dots, X_m \sim \mathcal{U}([0,\theta]),\ \theta > 0$$

$$f_\theta(x) = \frac{1}{\theta} \cdot \mathbb{1}_{[0,\theta]}(x) \rightarrow \text{densitatea}$$

$$L_m(\theta; x_1, \dots, x_m) = \prod_{i=1}^{m} f_\theta(x_i) = \prod_{i=1}^{m} \left[\frac{1}{\theta} \mathbb{1}_{[0,\theta]}(x_i)\right] =$$

$$= \frac{1}{\theta^m} \prod_{i=1}^{m} \mathbb{1}_{[0,\theta]}(x_i)$$

$$\mathbb{1}_{[0,\theta]}(x_i) = \begin{cases} 1, & x_i \in [0,\theta] \\ 0, & \text{altfel} \end{cases} = \begin{cases} 1, & \theta \geq x_i \\ 0, & \text{altfel} \end{cases} = \mathbb{1}_{[x_i, \infty)}(\theta)$$

$$\prod_{i=1}^{m} \mathbb{1}_{[0,\theta]}(x_i) = \prod_{i=1}^{m} \mathbb{1}_{(x_i,\infty)}(\theta)$$

$$\mathbb{1}_{[x_i,\infty)}(\theta) \times \mathbb{1}_{[x_j,\infty)}(\theta) = \begin{cases} 1, & \theta \geq x_i \text{ și } \theta \geq x_j \\ 0, & \text{altfel} \end{cases} =$$

$$= \begin{cases} 1, & \theta \geq \max\{x_i, x_j\} \\ 0, & \text{altfel} \end{cases}$$

$$\prod_{i=1}^{m} \mathbb{1}_{[x_i;\infty)}(\theta) = \mathbb{1}_{[x_{(m)},\infty)}(\theta)$$

$x_{(m)} \rightarrow \max\{x_1, \dots, x_m\}$

Atunci când $f_\theta(x) = g(\theta, x)\,\mathbb{1}_{(a(\theta), b(\theta))}(x_i)$

$$\boxed{\prod_{i=1}^{m} \mathbb{1}_{(a(\theta), b(\theta))}(x_i) = \mathbb{1}_{(\ ,\ )}(\theta)}$$

$a(\theta) = \theta$, $b(\theta) = \theta+1$

$$\prod_{i=1}^{m} \mathbb{1}_{[\theta,\theta+1]}(x_i)$$

$\theta \leq x_i \leq \theta+1$, deci $x_i - 1 \leq \theta \leq x_i$

$$= \prod_{i=1}^{m} \mathbb{1}_{[x_i-1, x_i]}(\theta) = \mathbb{1}_{\bigcap_i [x_i-1, x_i]}(\theta)$$

$$\mathbb{1}_A(\theta) \times \mathbb{1}_B(\theta) = \mathbb{1}_{A \cap B}(\theta)$$

$$\bigcap_i [x_i - 1, x_i] = [x_{(m)} - 1, x_{(1)}]$$

$$\prod_{i=1}^{m} \mathbb{1}_{[\theta,\theta+1]}(x_i) = \mathbb{1}_{[x_{(m)}-1, x_{(1)}]}(\theta)$$

Revenim, $X_i \sim \mathcal{U}[0,\theta]$

$$L_m(\theta; x_1,\dots,x_m) = \frac{1}{\theta^m}\,\mathbb{1}_{[\boxed{x_{(m)}}, +\infty)}(\theta) \quad \hookrightarrow \text{max.}$$

$$\hat{\theta}_m = \underset{\theta\in\Theta}{\arg\max}\, L_m(\theta; x_1,\dots,x_m)$$

$x_1 = 2.75 \qquad [x_1-1, x_1] = [1{,}75, 2{,}75]$

$x_2 = 2.36 \qquad [x_2-1, x_2] = [1{,}36, 2.36]$

$x_3 = 2.98 \qquad [x_3-1, x_3] = [1.98, 2.98]$

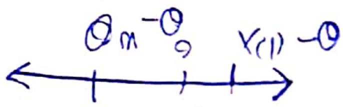
$x_{(m)}-1$ cea mai mare valoare dintre $x_1-1, x_2-1, \dots, x_m-1$

$x_{(1)}$ cea mai mică valoare dintre $x_1,\dots,x_m$

Consistența ~~TAE~~ EVM $\qquad X_1,\dots,X_m \sim \mathcal{U}[\theta,\theta+1]$

$$X_{(m)} - 1 \le \hat{\theta}_m \le X_{(1)}$$

$$X_{(m)} - 1 - \theta \le \hat{\theta}_m - \theta \le X_{(1)} - \theta$$



$$E_\theta\left[(\hat{\theta}_m-\theta)^2\right] \le (X_{(1)}-\theta)^2 + (X_{(m)}-1-\theta)^2$$

$x_1, x_2, \dots, x_m \longrightarrow \hat{\theta}_m$ observat

$\uparrow \quad \uparrow \quad \uparrow$

$X_1(\omega) \quad X_2(\omega) \quad X_m(\omega)$

$\downarrow \dfrac{x_1+x_2+\dots+x_m}{m}$

$$\hat{\theta}_m = \frac{X_1+X_2+\dots+X_m}{m}$$

Metode de construcție

1) Met. momentelor

momentele empirice = mom. teoretice

$E_\theta[X_1] = \overline{X_n}$ $\qquad g_1(\theta) \in \overline{X_n}$

$g_1(\theta)$ $\qquad \tilde{\theta}_n = g_1^{-1}(\overline{X_n})$

Obs. Dacă $\hat{\theta} \in \mathbb{R}^2$ (θ_1, θ_2)

$$\begin{cases} g_1(\theta) = E_\theta[X_1] = \overline{X_n} \\ g_2(\theta) = E[X_1^2] = \frac{\sum x_i^2}{n} \end{cases}$$

2) Metoda verosimilității maxime

$$L_n(\theta) = \prod_{i=1}^{n} f_\theta(x_i) \qquad \ell_n(\theta) = \log L_n$$

$$\hat{\theta}_n = \underset{\theta}{\arg\max}\, L_n(\theta) = \underset{\theta}{\arg\max}\, \ell_n(\theta)$$

3) Met. cuantilelor

$$\hat{x}_n(p) = F_\theta^{-1}(p) \leftarrow \text{o funcție care depinde de } \theta \quad h_p(\theta)$$

cuantila empirică = cuantila teoretică

$$\tilde{\theta}_n = h_p^{-1}(\hat{x}_n(p)) \qquad p \in \left\{\frac{1}{2}; \frac{1}{4}; \frac{3}{4}\right\}$$

$\frac{1}{2}$ → mediana; $\frac{1}{4}$ → prima cuartilă Q_1; $\frac{3}{4}$ → a treia cuartilă Q_3

$\hat{x}_n(p) \xrightarrow{a.s.} x_p$ (LNM) atunci când $F_\theta \nearrow$ (în special în x_p)

F_θ derivabilă în x_p

TLC $\sqrt{n}(\hat{x}_n(p) - x_p) \xrightarrow{d} \mathcal{N}\left(0, \frac{p(1-p)}{f^2(x_p)}\right)$

deplasarea → media

consistența → converge la val. pe care o estimează

normalitate asimpt. → de fel, se ajunge la o normală după

„compararea asimptotică" → cel cu varianța cea mai mică

$X | X > a$ (ex. 1 examen)

Să generăm $Y \sim X$ și să păstrăm $Y > a$ (while)

Pt. o obs. avem nevoie de N repetiții unde $N \sim Geom(P(Y>a))$

$$E[N] = \frac{1}{P(Y>a)}$$

~~Y = gen X~~

gen Y = function(a){
  Y = gen X
  while (Y<a)
    Y = gen X
  end.
  return(Y))
}

Dacă a este foarte mare atunci $P(Y>a)$ este foarte mică și avem nevoie de un nr. f. mare de obs. ca să obținem 1 obs. din $X | X > a$.

b) $U \sim \mathcal{U}[0,1]$

$$T = F^{-1}(\underbrace{F(a) + (1-F(a))U}_{\geq F(a) \text{ fct. crescătoare}})$$

$P(T \leq t)$

pt. $t > a$

$F(a) + (1-F(a))U \geq F(a)$

$F(a) + (1-F(a))U) \geq a$

$T = F^{-1}(F(a) + (1-F(a))U) \geq a$

$$F_T(t) = P(T \leq t) = P(F^{-1}(F(a) + (1-F(a))U) \leq t)$$
$$= P(F(a) + (1-F(a))U \leq F(t))$$
$$= P\left(U \leq \frac{F(t)-F(a)}{1-F(a)}\right)$$

$U \sim \mathcal{U}([0,1])$ $F_U(u) = u$, $u \in [0,1]$

$$F_T(t) = \frac{F(t)-F(a)}{1-F(a)} = \frac{P(X \leq t) - P(X \leq a)}{P(X > a)} = \frac{P(a < X \leq t)}{P(X>a)}$$

$$F_T(t) = \frac{P(X \le t) - P(X \le a)}{P(X > a)} = \frac{P(a < X \le t)}{P(X > a)} = P(X \le t \mid X > a)$$

$U \sim \mathcal{U}[0,1]$

Atunci $T = F^{-1}\big(F(a) + (1 - F(a))U\big) \sim X \mid X > a$