Examen Statistică

8 Feb 2020

(11 feb 2018, 2 iunie 2018)

Exercițiul 1: Fie o variabilă aleatoare repartizată

$$\mathbb{P}_\theta(X=k) = A(k+1)\theta^k, \quad k \in \mathbb{N} \quad \text{unde } \theta \in (0,1)$$

un parametru necunoscut și $A \in \mathbb{R}$ constantă.

1. Determinați constanta A și calculați $\mathbb{E}[X]$ și $Var(X)$. Dorim să estimăm pe θ plecând de la un eșantion $X_1, X_2 \dots X_n$ de talie n din populația dată de repartiția lui X.
2. Det. estimatorul $\tilde{\theta}$ a lui θ prin metoda momentelor și calculați $\mathbb{P}_\theta(\tilde{\theta}=0)$.
3. Det. estimatorul de verosimilitate maximă $\hat{\theta}$ a lui θ și verificați dacă acesta este bine definit.
4. Studiați consistența estimatorului $\tilde{\theta}$ și det. legea la limită.

Exercițiul 2: Considerăm cuplul de variabile aleatoare (X,Y) cu densitatea:

$$f(x,y) = \frac{1}{\sqrt{8\pi}} e^{-y^2 x/2} \cdot e^{-\sqrt{x}} \quad x>0$$

1. Det. repartiția condiționată a lui Y la $X=x$.
2. Det. repartiția lui $\sqrt{X}$
3. Propuneți o metodă de simulare a unei observații din cuplul (X,Y) și scrieți un cod R care să permită acest lucru.

Scanned with CamScanner

Scanned with CamScanner

Exercițiul 3. Fie $X_1 \dots X_n$ un eșantion de talie n din populația f_θ unde $f_\theta(x) = \frac{7}{(x-\theta)^8} \mathbb{1}_{[1+\theta, +\infty)}(x)$

a) Calculați $E_\theta[X_1]$, $Var_\theta(X_1)$ și funcția de repartiție $F_\theta(x)$ a lui X_1.

b) În cazul în care $\theta = 2$, dorim să generăm 3 valori aleatoare din repartiția lui $X \sim f_\theta(x)$. Pentru aceasta dispunem de trei valori rezultate din repartiția uniformă pe $[0,1]$ $u_1 = 0.25$, $u_2 = 0.4$, $u_3 = 0.5$. Descrieți procedura.

c). Determinați estimatorul $\hat{\theta}_n^M$ a lui θ obținut prin metoda momentelor și calculați eroarea pătratică medie a acestui estimator. Care este legea la limită?

d) Exprimați în funcție de θ mediana repartiției lui X_1 și, plecând de la aceasta, găsiți un alt estimator $\hat{\theta}_n^{a}$ pt θ

e). Det. legea la limită a lui $\hat{\theta}_n^{a}$ și arătați că, asimptotic acesta este mai bun decât $\hat{\theta}_n^M$

f). Det. estimatorul de verosimilitate maximă $\hat{\theta}_n^{VM}$ a lui θ și verificați dacă este deplasat.

h) Pe care dintre cei trei estimatori îi preferați?

Ex 4 Considerăm densitatea $f(y) = \frac{1}{2\sqrt{1-y}} \mathbb{1}_{[0,1]}(y)$ unde folosim convenția $f(1) = +\infty$.

① De v.a. Y are densitatea f, care este dens. v.a. $X = \theta Y$, $\theta > 0$?

② $X_1 \dots X_n$ eșantion talie n din X. Det. estimatorul de verosimilitate max. $\hat{\theta}_n$ a lui θ.

③ Det. repart. limită a lui $n\sqrt{\theta - \hat{\theta}_n}$

④ Det. mediana repartiției v.a. X și deduceți un nou estimator $\bar{\theta}_n$. Pe care din cei doi estimatori îi preferați?

Scanned with CamScanner

Scanned with CamScanner

Ex 2

(X, Y)

a) $f(x,y) = \frac{1}{\sqrt{8\pi}} e^{-\frac{y^2 x}{2}} \cdot e^{-\sqrt{x}} \cdot 1\!\!1_{x>0}$

1. $$f_{Y|X}(y|x) \overset{def}{=} \frac{f(x,y)}{f_X(x)},$$

$$\text{unde } f_X(x) \overset{def}{=} \int_{-\infty}^{\infty} f(x,y)\, dy$$

$$f_X(x) = \frac{1}{\sqrt{8\pi}} \cdot e^{-\sqrt{x}} \cdot \int_{-\infty}^{\infty} e^{-\frac{y^2 x}{2}}\, dy$$

$$\text{S.V. } z = \frac{y\sqrt{x}}{\sqrt{2}} \qquad dz = dy \cdot \frac{\sqrt{x}}{\sqrt{2}}$$

$$\int_{-\infty}^{\infty} e^{-\frac{y^2 \cdot x}{2}}\, dy = \underbrace{\int_{-\infty}^{\infty} e^{-z^2}}_{\sqrt{\pi}} \cdot \frac{\sqrt{2}}{\sqrt{x}}\, dz$$

$$= \sqrt{\frac{2\pi}{x}}$$

Deci $f_{Y|X}(y|x) = \sqrt{4}\cdot e^{\sqrt{x}}\cdot\sqrt{2x}\cdot f(y|x)$

$$= 2\, e^{\sqrt{x}}\cdot\sqrt{x}\cdot\frac{1}{\sqrt{8\pi}}\, e^{\frac{-y^2x}{2}}\cdot e^{-\sqrt{x}}$$

$$= \sqrt{\frac{x}{2\pi}}\cdot e^{-\frac{y^2x}{2}} \sim$$

$$\sim N\left(0,\frac{1}{x}\right)$$

b) Dt. $t<0$

$\mathbb{P}(X\le t)=0$

Dt. $t>0$

$$\mathbb{P}(X\le t) = \mathbb{P}((X,Y)\in[0,t]\times\mathbb{R})$$

$$= \int_0^t\left(\int_{-\infty}^{\infty} f(x,y)\,dy\right)dx$$

$$= \int_0^t\left(\int_{-\infty}^{\infty}\frac{1}{\sqrt{8\pi}}\cdot e^{-\sqrt{x}}\cdot e^{-\frac{y^2x}{2}}\,dy\right)dx$$

$$= \int_0^t \frac{1}{2\sqrt{x}}\, e^{-\sqrt{x}}\,dx$$

S.V. $z=\sqrt{x}$ $\quad dz = \frac{1}{2\sqrt{x}}\,dx$

$$\mathbb{P}(X \le t) = \int_0^{\sqrt{t}} e^{-z}\,dz$$

$$= 1 - e^{-\sqrt{t}}$$

$$\mathbb{P}(\sqrt{X} \le t) = \mathbb{P}(X \le t^2) = 1 - e^{-t}$$

Deci $\sqrt{X} \sim Exp(1)$

c) Generăm un eșantion din distribuția X, apoi, pt fiecare valoare x a eșantionului, generăm o observație din distribuția $(Y/X = x)$.

Cod R:

```
n = 100
RX = rexp(n, 1)  // observații √X
X = RX^2
Y = Rep(0, n)
for (i in 1:n) {
  Y[i] = rnorm(1, 0, 1/X[i])
}
```

Ex 3 $X_1, X_2, \dots, X_n \sim f_\theta$

$$f_\theta(x) = \frac{7}{(x-\theta)^8} \mathbb{1}_{[1+\theta, \infty)}(x)$$

a) $$E_\theta[X_1] = \int_{-\infty}^{\infty} x \cdot \frac{7}{(x-\theta)^8} \cdot \mathbb{1}_{[1+\theta,\infty)}(x)$$

$$= \int_{1+\theta}^{\infty} x \cdot \frac{7}{(x-\theta)^8} = \int_{1+\theta}^{\infty} \frac{7}{(x-\theta)^7} + \theta \cdot \int_{1+\theta}^{\infty} \frac{7}{(x-\theta)^8}$$

$$= -\frac{7}{6\cdot(x-\theta)^6}\Big|_{1+\theta}^{\infty} + \theta \cdot \frac{1}{(x-\theta)^7}\Big|_{1+\theta}^{\infty}$$

$$= \frac{7}{6} + \theta$$

$$E_\theta[X_1^2] = \int_{1+\theta}^{\infty} x^2 \cdot \frac{7}{(x-\theta)^8} dx$$

$$= \underbrace{\int_{1+\theta}^{\infty} (x-\theta)^2 \cdot \frac{7}{(x-\theta)^8} dx}_{\frac{7}{(x-\theta)^6}} + 2\theta \cdot \int_{1+\theta}^{\infty} x \cdot \frac{7}{(x-\theta)^8} - \theta^2 \cdot \int_{1+\theta}^{\infty} \frac{7}{(x-\theta)^8}$$

$$= \frac{7}{5} + 2\theta \cdot \left(\frac{7}{6} + \theta \right) - \theta^2$$

$$= \frac{7}{5} + 2\theta \cdot \frac{7}{6} + 2\theta^2 - \theta^2$$

$$= \frac{7}{5} + 2\theta \cdot \frac{7}{6} + \theta^2$$

$$\mathrm{Var}_\theta(X_1) = \mathbb{E}_\theta[X_1^2] - \left(\mathbb{E}_\theta[X_1] \right)^2$$

$$= \frac{7}{5} + 2\theta \cdot \frac{7}{6} + \theta^2 - \left(\frac{7}{6} + \theta \right)^2$$

$$= \frac{7}{5} + 2\theta \cdot \frac{7}{6} + \theta^2 - \left(\frac{7}{6} \right)^2 - 2 \cdot \frac{7}{6} \cdot \theta - \theta^2$$

$$= \frac{7}{5} - \left(\frac{7}{6} \right)^2 = \frac{7}{5} - \frac{49}{36}$$

$$= \frac{7}{180}$$

$$F_\theta(x) = \mathbb{P}(X \leq x) = \begin{cases} 0, & x \leq 1+\theta \\ \int_{1+\theta}^{x} \frac{7}{(t-\theta)^8} \, dt \end{cases}$$

$$= \left(- \frac{1}{(x-\theta)^7} \Big|_{1+\theta}^{x} \right) \cdot \mathbb{1}_{[1+\theta, \infty)}(x)$$

$$= \left(1 - \frac{1}{(x-\theta)^7} \right) \cdot \mathbb{1}_{[1+\theta, \infty)}(x)$$

b) $\theta = 2$

Fix. $y \in (0, 1)$

$$F_2(x) = y \iff 1 - \frac{1}{(x-2)^7} = y \iff$$

$$\Rightarrow (x-2)^7 = \frac{1}{1-y}$$

$$x = 2 + \sqrt[7]{\frac{1}{1-y}}$$

Deci $F_2^{-1}(y) = 2 + \sqrt[7]{\frac{1}{1-y}}$

Dacă $U \sim U[0, 1]$

$$F_2^{-1}(U) \sim f_2$$

Deci e suficient să aplicăm F_2^{-1} pe valorile date

$$2+\sqrt[7]{\frac{1}{1-0{,}25}}\,,\quad 2+\sqrt[7]{\frac{1}{1-0{,}4}}\,,\quad 2+\sqrt[7]{\frac{1}{1-0{,}5}}$$

c) $\mathbb{E}_\theta[X_1] = \frac{7}{6} + \theta$

Deci punem $\hat{\theta}_n^m = \overline{X_n} - \frac{7}{6}$

(înlocuim $\mathbb{E}_\theta[X_1]$ cu $\overline{X_n}$)

$$MSE_\theta(\hat{\theta}_n^m) = \mathbb{E}_\theta\left[\left(\hat{\theta}_n^m - \theta\right)^2\right]$$

$$= \mathbb{E}_\theta\left[\left(\overline{X_n} - \frac{7}{6} - \theta\right)^2\right]$$

$$= \mathbb{E}_\theta\left[\overline{X_n}^2\right] - 2\cdot\left(\frac{7}{6}+\theta\right)\cdot \underbrace{\mathbb{E}[\overline{X_n}]}_{=\mathbb{E}[X_1] = \left(\frac{7}{6}+\theta\right)} + \left(\frac{7}{6}+\theta\right)^2$$

$$= \mathbb{E}_\theta\left[\overline{X_n}^2\right] - \left(\frac{7}{6}+\theta\right)^2$$

$$= \mathbb{E}_\theta\left[\overline{X_n}^2\right] - \left(\mathbb{E}_\theta[\overline{X_n}]\right)^2$$

$$= Var_\theta(\overline{X_n}) =$$

$$= Var(\overline{X}_n) = Var\left(\frac{X_1 + \dots + X_n}{n}\right)$$

$$\overset{X_1, \dots X_n \text{ i.i.d.}}{=} \frac{1}{n^2} \cdot n \cdot Var(X_1)$$

$$= \frac{1}{n} Var(X_1)$$

$$= \frac{1}{n} \cdot \frac{7}{180}$$

Din T.L.C:

$$\sqrt{n}\left(\frac{\overline{X}_n - \overbrace{\left(\frac{7}{6} + \theta\right)}^{E[X_1]}}{\underbrace{\sqrt{\frac{7}{180}}}_{\sqrt{Var}}}\right) \xrightarrow{d} N(0, 1)$$

Cum

$$\hat{\theta}_n^M = \overline{X}_n - \frac{7}{6},$$

$$\sqrt{n}\left(\hat{\theta}_n^M - \theta\right) \xrightarrow{d} N\left(0, \frac{7}{180}\right)$$

d) Mediana este

$$F_\theta^{-1}\left(\frac{1}{2}\right) = x_{\frac{1}{2}} = \text{mediana.}$$

$$F_\theta^{-1}(y) = \theta + \sqrt[7]{\frac{1}{1-y}}$$ (am demonstrat în cazul $\theta = 2$)

$$x_{\frac{1}{2}} = F_\theta^{-1}\left(\frac{1}{2}\right) = \theta + \sqrt[7]{\frac{1}{1-\frac{1}{2}}}$$

$$= \theta + \sqrt[7]{2}$$

Deci $$\hat{\theta}_n^Q = \underbrace{\hat{x}_n\left(\frac{1}{2}\right)}_{\text{mediana empirică}} - \sqrt[7]{2}$$

F_θ e derivabilă în $x_{\frac{1}{2}}$ și $f_\theta(x_{\frac{1}{2}}) > 0$,

deci

$$\sqrt{n}\,\Big(\underbrace{\hat{x}_n\left(\frac{1}{2}\right) - x_{\frac{1}{2}}}_{= \hat{\theta}_n^Q - \theta}\Big) \xrightarrow{d} N\left(0, \frac{\frac{1}{2}\left(1-\frac{1}{2}\right)}{\left(f_\theta\left(x_{\frac{1}{2}}\right)\right)^2}\right)$$

$$f_\theta\left(x_{\frac{1}{2}}\right) = \frac{7}{\left(\theta + \sqrt[7]{2} - \theta\right)^8} = \frac{7}{2^{\frac{8}{7}}}$$

$$\sqrt{n}\left(\hat{\theta}_n^Q - \theta\right) \xrightarrow{d} N\left(0, \frac{\left(2^{\frac{8}{7}}\right)^2}{4\cdot 49}\right)$$

$$\frac{7}{180} > \frac{\left(2^{\frac{8}{7}}\right)^2}{4\cdot 49} \iff \underbrace{\frac{28\cdot 49}{180}}_{} > 2^{\frac{8}{7}}$$

$$7.62\ldots > 4.87\ldots \checkmark$$

Deci, asimptotic $\hat{\theta}_n^Q$ e mai bun,
(are varianța mai mică,
deci e mai bun)

f) $L(\theta; x_1, \ldots x_n) = f_\theta(x_1 \ldots x_n)$

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

$$= \frac{7}{(x_1-\theta)^8} \cdot \frac{7}{(x_2-\theta)^8} \cdot \ldots \cdot \frac{7}{(x_n-\theta)^8} \cdot$$

$$\underbrace{\cdot\, \mathbb{1}_{[1+\theta,\infty)}(x_1) \cdot \ldots \cdot \mathbb{1}_{[1+\theta,\infty)}(x_n)}_{}$$

$$\underbrace{\mathbb{1}_{(-\infty, x_1-1]}(\theta) \cdot \ldots \cdot \mathbb{1}_{(-\infty, x_n-1]}(\theta)}_{}$$

$$= \mathbb{1}_{(-\infty, x_{(1)}-1]}(\theta)$$

Funcția

$\theta \longrightarrow \prod \frac{7}{(x_i-\theta)^8}$ e crescătoare pe $(-\infty, x_{(1)}-1)$

deci maximul fct. de verosimilitate este în $x_{(1)}-1$

$$\hat{\theta}_n^{VM} = X_{(1)} - 1$$

statistica de ordin 1.

$$\mathbb{E}[\hat{\theta}_n] = \mathbb{E}[X_{(1)}] - 1$$

$$\mathbb{P}(X_{(1)} \le x) = \mathbb{P}(X_1 \le x \text{ și } X_2 \le x \text{ și} \dots X_n \le x)$$

$$\overset{i.i.d}{=} \left(\mathbb{P}(X_1 \le x)\right)^n$$

$$= \left(1 - \frac{1}{(x-\theta)^7}\right)^n \cdot \mathbb{1}_{[1+\theta,\infty)}(x)$$

$$f_{(X_{(1)})}(x) = n \cdot \left(1 - \frac{1}{(x-\theta)^7}\right)^{n-1} \cdot \frac{7}{(x-\theta)^8} \cdot \mathbb{1}_{[1+\theta,\infty)}(x)$$

$$\mathbb{E}[X_{(1)}] = \int_{1+\theta}^{\infty} x \cdot n \cdot \left(1 - \frac{1}{(x-\theta)^7}\right)^{n-1} \cdot \frac{7}{(x-\theta)^8}$$

$$= \int_{1+\theta}^{\infty} n \cdot \left(1 - \frac{1}{(x-\theta)^7}\right)^{n-1} \cdot \frac{7}{(x-\theta)^7} +$$

$$+ \theta \cdot \underbrace{\int_{1+\theta}^{\infty} n \cdot \left(1 - \frac{1}{(x-\theta)^7}\right)^{n-1} \cdot \frac{7}{(x-\theta)^7}}_{= \mathbb{P}(X_{(1)} \leq \infty) = 1}$$

$$E[\hat{\theta}_n^{VM}] - \theta = \int_{1+\theta}^{\infty} n \cdot \left(1 - \frac{1}{(x-\theta)^7}\right)^{n-1} \cdot \frac{7}{(x-\theta)^7}\, dx -$$

$$- 1$$

$$(x-\theta)^7 \leq (x-\theta)^8$$

$$\int_{1+\theta}^{\infty} n \cdot \left(1 - \frac{1}{(x-\theta)^7}\right)^{n-1} \cdot \frac{7}{(x-\theta)^7} > \int_{1+\theta}^{\infty} n \cdot \left(1 - \frac{1}{(x-\theta)^7}\right)^{n-1} \cdot \frac{7}{(x-\theta)^8} =$$

$$= 1,$$

deci $E[\hat{\theta}_n^{VM}] - \theta > 0$, de unde

$\hat{\theta}_n^{VM}$ e DEPLASAT

$$F_{\hat{\theta}_n^{vm}-\theta}(t) =$$

$$= \mathbb{P}\left(\hat{\theta}_n^{vm} - \theta \le t\right) = \mathbb{P}\left(X_{(1)} \le 1 + \theta + t\right)$$

$$= \left(1 - \frac{1}{(1+\theta+t-\theta)^7}\right)^n \cdot \mathbb{1}_{[1+\theta,\infty)}(1+\theta+t)$$

$$= \left(1 - \frac{1}{(1+t)^7}\right)^n \cdot \mathbb{1}_{[0,\infty)}(t)$$

$$\mathbb{P}\left(\sqrt{n}\left(\hat{\theta}_n^{vm} - \theta\right) \le t\right) = \mathbb{P}\left(\left(\hat{\theta}_n^{vm} - \theta\right) \le \frac{t}{\sqrt{n}}\right)$$

$$= \left(1 - \frac{1}{\left(1 + \frac{t}{\sqrt{n}}\right)^7}\right)^n \cdot \mathbb{1}_{[0,\infty)}(t)$$

$$\ln\left(1 - \left(\frac{x}{x+t}\right)^7\right)^{x^2} =$$

$$= x^2 \ln\left(1 - \left(\frac{x}{x+t}\right)^7\right) \xrightarrow{x \to \infty} \infty \cdot (-\infty) =$$

$$= -\infty$$

Deci $\mathbb{P}\left(\sqrt{n}\left(\hat{\theta}_n - \theta\right) \leq t\right) \xrightarrow{n \to \infty} 0$,

de unde $\sqrt{n}\left(\hat{\theta}_n^{VM} - \theta\right) \xrightarrow{d} 0$

Asimptotic, $\hat{\theta}_n^{VM}$ e cel mai bun

ÎL ALEGEM PE

ACESTA