ESTIMAREA PARAMETRILOR

Prin alegerea modelului:

- forma functionala specificata
- existenta unor parametri necunoscuti

"Model parametric"

$$P_\theta \circ X^{-1}, \theta \in \Theta \sqsubseteq R^k, k \geq 1$$

$$X : \Omega \longrightarrow S, \quad \text{v.a.}, \quad S = A \text{ sau } S = R$$

Presupunem modelul "corect": valoarea adevarata, necunoscuta $\theta_0 \in \Theta$.

Observatiile $X_1, ..., X_n$ v.a.i.i.r. $P_\theta \circ X^{-1}$
Spatiul de selectie $n-$dimensional $\left(S^n, \mathcal{S}^n, \bigotimes_{i=1}^{n} P_\theta \circ X_i^{-1}\right)$

$$\left(A^n, (\mathcal{P}(A))^n, \bigotimes_{i=1}^{n} P_\theta \circ X_i^{-1}\right)$$
$$\left(R^n, \mathcal{B}^n, \bigotimes_{i=1}^{n} P_\theta \circ X_i^{-1}\right)$$

Definitie:
Fie o functie masurabila $\widehat{\theta} : S^n \longrightarrow \Theta$. Atunci $\widehat{\theta}(X_1, ..., X_n)$ se numeste estimator al parametrului θ.

$$\Omega \overset{(X_1,...,X_n)}{\longrightarrow} S^n \overset{\widehat{\theta}(x_1,...,x_n)}{\longrightarrow} \Theta$$

Pentru datele statistice $(x_1, ..., x_n)$, valoarea $\widehat{\theta}(x_1, ..., x_n)$ se numeste estimatie a lui θ.

Notatii (presupunand ca toate mediile de mai jos exista):

$$\theta = (\theta_1, ..., \theta_k)'$$
$$\widehat{\theta} = \left(\widehat{\theta}_1, ..., \widehat{\theta}_k\right)'$$

$$M_\theta\left(\widehat{\theta}\right) = \left(M_\theta\left(\widehat{\theta}_1\right), ..., M_\theta\left(\widehat{\theta}_k\right)\right)'$$

$$Cov_\theta\left(\widehat{\theta},\widehat{\theta}\right)=\left\|cov_\theta\left(\widehat{\theta}_i,\widehat{\theta}_j\right)\right\|_{i,j=1,\ldots,k}$$
$$=\left\|M_\theta\left(\left(\widehat{\theta}_i-M_\theta\left(\widehat{\theta}_i\right)\right)\left(\widehat{\theta}_j-M_\theta\left(\widehat{\theta}_j\right)\right)\right)\right\|_{i,j=1,\ldots,k}$$

$$\text{Pentru. } k=1,\quad M_\theta\left(\widehat{\theta}\right),D_\theta^2\left(\widehat{\theta}\right)$$

Definitii:

- $\widehat{\theta}\left(X_1,...,X_n\right)$ este estimator nedeplasat daca
$$M_\theta\left(\widehat{\theta}\left(X_1,...,X_n\right)\right)=\theta,\quad \forall\theta\in\Theta$$
- $\widehat{\theta}\left(X_1,...,X_n\right)$ este estimator nedeplasat, de dispersie minima (ENDM) daca este nedeplasat si pentru orice alt estimator nedeplasat $g\left(X_1,...,X_n\right)$ matricea
$$Cov_\theta\left(g,g\right)-Cov_\theta\left(\widehat{\theta},\widehat{\theta}\right)$$
este semipozitiv definita, $\forall\theta\in\Theta$.

Comentariu:
Pentru $k=1$, $\widehat{\theta}\left(X_1,...,X_n\right)$ este ENDM daca
$$M_\theta\left(\widehat{\theta}\right)=\theta,\quad \forall\theta\in\Theta$$
$$D_\theta^2\left(\widehat{\theta}\right)\le D_\theta^2\left(g\right),\quad \forall\theta\in\Theta$$
pentru orice alt estimator nedeplasat $g\left(X_1,...,X_n\right)$.
DEPLASAREA estimatorului $\widehat{\theta}$
$$Bias\left(\widehat{\theta}\right)=M_\theta\left(\widehat{\theta}\right)-\theta$$

EROAREA MEDIE PATRATICA a estimatorului $\widehat{\theta}$
$$M_\theta\left(\widehat{\theta}-\theta\right)^2=D_\theta^2\left(\widehat{\theta}\right)+\left(Bias\left(\widehat{\theta}\right)\right)^2$$

Definitie:
Fie un sir de observatii i.i.r., $\left(X_n\right)_n$ si fie $\left(\widehat{\theta}\left(X_1,...,X_n\right)\right)_n$.
Spunem ca $\widehat{\theta}$ este un estimator consistent daca
$$\widehat{\theta}\left(X_1,...,X_n\right)\xrightarrow{P_\theta}\theta\ \text{ pentru } n\to\infty,\quad\forall\theta\in\Theta$$

"Estimatori buni" $\Longleftrightarrow$ nedeplasati, ENDM, consistenti.

Metode:

- metoda momentelor
- metoda verosimiltatii maxime (maximum likelihood)
- metoda celor mai mici patrate (least squares)
- metoda lui Bayes

METODA MOMENTELOR

utila cand semnificatia lui θ este direct legata de momentele lui X

Momentele lui X (presupunem ca exista)

$$\mu_r = M\left(X^r\right), \quad r \in N^*$$
$$\mu_1 = M\left(X\right)$$

Momentele centrate ale lui X (presupunem ca exista)

$$\overline{\mu_r} = M\left(\left(X - \mu_1\right)^r\right), \quad r \in N^*$$
$$\overline{\mu_2} = D^2\left(X\right)$$

Pentru observatiile i.i.d. $X_1, ..., X_n$, definim momentele de selectie

$$\widehat{\mu_r} = \frac{1}{n}\sum_{i=1}^{n} X_i^r, \quad r \in N^*$$
$$\widehat{\mu_1} = \overline{X}$$

$$\widehat{\overline{\mu_r}} = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^r, \quad r \in N^*$$
$$\widehat{D^2\left(X\right)} = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2$$

Proprietatea 1

$$M\left(\widehat{\mu_r}\right) = \mu_r \quad \text{(estimator nedeplasat)}$$
$$M\left(\widehat{D^2(X)}\right) = \frac{n-1}{n} \cdot D^2(X) \quad \text{(estimator deplasat)}$$

Demonstratie:

$$M\left(\widehat{\mu_r}\right) = \frac{1}{n}\sum_{i=1}^{n} M\left(X_i^r\right) = \frac{1}{n} \cdot n\mu_r = \mu_r$$

$$M\left(\overline{X}\right) = M(X)$$

$$D^2\left(\overline{X}\right) = \frac{1}{n^2}\sum_{i=1}^{n} D^2(X_i) = \frac{1}{n^2} \cdot nD^2(X) = \frac{1}{n}D^2(X)$$

$$\begin{aligned}\widehat{D^2(X)} &= \frac{1}{n}\sum_{i=1}^{n}\left((X_i - M(X)) - \left(\overline{X} - M(X)\right)\right)^2 = \\ &= \frac{1}{n}\left\{\sum_{i=1}^{n}(X_i - M(X))^2 - n\left(\overline{X} - M(X)\right)^2\right\}\end{aligned}$$

$$M\left(\widehat{D^2(X)}\right) = \frac{1}{n}\left\{nD^2(X) - nD^2\left(\overline{X}\right)\right\} = \frac{n-1}{n} \cdot D^2(X)$$

■

Un estimator nedeplasat pentru $D^2(X)$ este

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2 = \frac{n}{n-1}\widehat{D^2(X)}$$

Cat poate sa fie dispersia unor estimatori nedeplasati?

TEOREMA RAO - CRAMER (pentru $k = 1$)

Fie modelul $P_\theta \circ X^{-1}$, avand densitatea de repartitie

$$f(x,\theta), \ \ x \in R,$$

cu $\theta \in \Theta \sqsubseteq R$.

Fie observatiile i.i.r. $X_1, ..., X_n$ si notam densitatea de repartitie a vectorului $(X_1, ..., X_n)$ cu

$$f(x_1, ..., x_n; \theta) = \prod_{i=1}^{n} f(x_i; \theta)$$

Fie $\widehat{\theta}(X_1, ..., X_n)$ un estimator nedeplasat pentru θ.

Presupunem verificate urmatoarele conditii de regularitate:

- Θ este multime deschisa;
- $f(x_1, ..., x_n; \theta)$ derivabila in raport cu θ pe Θ oricare ar fi $(x_1, ..., x_n)$, cu derivata integrabila pe R^n;
- Pentru orice θ, au loc egalitatile

$$\frac{\partial}{\partial\theta} \int_{R^n} f(x_1, ..., x_n; \theta)\, dx_1 ... dx_n = \int_{R^n} \frac{\partial f(x_1, ..., x_n; \theta)}{\partial\theta} dx_1 ... dx_n$$

$$\frac{\partial}{\partial\theta} \int_{R^n} \widehat{\theta}(x_1, ..., x_n) \cdot f(x_1, ..., x_n; \theta)\, dx_1 ... dx_n$$
$$= \int_{R^n} \widehat{\theta}(x_1, ..., x_n) \cdot \frac{\partial f(x_1, ..., x_n; \theta)}{\partial\theta} dx_1 ... dx_n$$

- Exista "informatia Fisher"

$$M_\theta \left(\frac{\partial \ln f(X_1, ..., X_n; \theta)}{\partial\theta} \right)^2 \overset{notat}{=} i_n(\theta) > 0$$

Atunci are loc inegalitatea

$$D_\theta^2 \left(\widehat{\theta} \right) \geq \frac{1}{i_n(\theta)}, \ \ \theta \in \Theta$$

Egalitatea are loc daca si numai daca exista o constanta A, independenta de $(x_1, ..., x_n)$, asa incat

$$A \cdot \left(\widehat{\theta}(x_1, ..., x_n) - \theta \right) = \frac{\partial f(x_1, ..., x_n; \theta)}{\partial\theta}, \ \ \forall (x_1, ..., x_n)$$

Demonstratie:

Notam

$$Y=\frac{\partial\ln f\left(X_1,...,X_n;\theta\right)}{\partial\theta}$$

Avem

$$M_\theta\left(Y\right)=\int\limits_{R^n}\left(\frac{1}{f\left(x_1,...,x_n;\theta\right)}\cdot\frac{\partial f\left(x_1,...,x_n;\theta\right)}{\partial\theta}\right)f\left(x_1,...,x_n;\theta\right)dx_1...dx_n$$

$$=\frac{\partial}{\partial\theta}\left(\int\limits_{R^n}f\left(x_1,...,x_n;\theta\right)dx_1...dx_n\right)=0$$

$$M_\theta\left(Y^2\right)=i_n\left(\theta\right)$$

Utilizam inegalitatea integrala a lui Schwartz,

$$\left(M\left(|UV|\right)\right)^2\leq M\left(|U|^2\right)\cdot M\left(|V|^2\right),$$

pentru $U=\widehat{\theta}-\theta$ si $V=Y-M_\theta\left(Y\right).$

Obtinem

$$\left(cov_\theta\left(\widehat{\theta},Y\right)\right)^2\leq D_\theta^2\left(\widehat{\theta}\right)\cdot i_n\left(\theta\right)$$

Dar

$$cov_\theta\left(\widehat{\theta},Y\right)=M_\theta\left(\widehat{\theta}\cdot Y\right)-M_\theta\left(\widehat{\theta}\right)\cdot M_\theta\left(Y\right)=$$

$$=\int\limits_{R^n}\left(\widehat{\theta}\left(x_1,...,x_n\right)\cdot\frac{1}{f\left(x_1,...,x_n;\theta\right)}\cdot\frac{\partial f\left(x_1,...,x_n;\theta\right)}{\partial\theta}\right)f\left(x_1,...,x_n;\theta\right)dx_1...dx_n$$

$$=\frac{\partial}{\partial\theta}\left(\int\limits_{R^n}\widehat{\theta}\left(x_1,...,x_n\right)\cdot f\left(x_1,...,x_n;\theta\right)dx_1...dx_n\right)=\frac{\partial\theta}{\partial\theta}=1$$

Rezulta

$$1\leq D_\theta^2\left(\widehat{\theta}\right)\cdot i_n\left(\theta\right).$$

O c.n.s. pentru a obtine egalitate in inegalitatea Schwartz este sa existe o constanta A,independenta de $\left(x_1,...,x_n\right)$,asa incat

$$A\cdot\left(\widehat{\theta}\left(x_1,...,x_n\right)-\theta\right)=\frac{\partial f\left(x_1,...,x_n;\theta\right)}{\partial\theta},\quad\forall\left(x_1,...,x_n\right)$$

■

Remarca:

$$i_n(\theta) = n \cdot i_1(\theta)$$

Demonstratie:

$$\frac{\partial \ln f(X_1, ..., X_n; \theta)}{\partial \theta} = \sum_{i=1}^{n} \frac{\partial \ln f(X_i; \theta)}{\partial \theta}$$

$$i_n(\theta) = M_\theta \left(\sum_{i=1}^{n} \frac{\partial \ln f(X_i; \theta)}{\partial \theta} \right)^2 =$$

$$= \sum_{i=1}^{n} M_\theta \left(\frac{\partial \ln f(X_i; \theta)}{\partial \theta} \right)^2 + 2 \sum_{i<j} M_\theta \left(\frac{\partial \ln f(X_i; \theta)}{\partial \theta} \cdot \frac{\partial \ln f(X_j; \theta)}{\partial \theta} \right) =$$

$$= n \cdot i_1(\theta) + 2 \sum_{i<j} M_\theta \left(\frac{\partial \ln f(X_i; \theta)}{\partial \theta} \right) \cdot M_\theta \left(\frac{\partial \ln f(X_j; \theta)}{\partial \theta} \right) = n \cdot i_1(\theta)$$

Definitie
Un estimator nedeplasat $\widehat{\theta}$ pentru care

$$D_\theta^2 \left(\widehat{\theta} \right) = \frac{1}{n \cdot i_1(\theta)}$$

se numeste estimator eficient.

EXEMPLU

Modelul: Repartitia Exponentiala $Expo(\theta), \theta \in (0, \infty)$

$$f(x; \theta) = \begin{cases} \frac{1}{\theta} \exp\left(-\frac{x}{\theta}\right), & x \in [0, \infty) \\ 0, & x \in (-\infty, 0) \end{cases}$$

Semnificatia parametrului

$$M_\theta(X) = \frac{1}{\theta} \int_0^\infty x \cdot \exp\left(-\frac{x}{\theta}\right) dx = \theta$$

Spatiul de selectie $n-$dimensional

$$\left([0, \infty)^n, \left(\mathcal{B}_{[0,\infty)}\right)^n, \bigotimes_{i=1}^{n} P_\theta \circ X_i^{-1} \right)$$

$$f(x_1,...,x_n;\theta)=\prod_{i=1}^{n} f(x_i;\theta)=\begin{cases} \frac{1}{\theta^n}\exp\left(-\frac{1}{\theta}\sum\limits_{i=1}^{n} x_i\right), & x_i\in[0,\infty),\ \forall i \\ 0, & \text{in rest}\end{cases}$$

Aplicam Metoda Momentelor

$$\widehat{\theta}(X_1,...,X_n)=\overline{X},$$
$$M_\theta\left(\widehat{\theta}\right)=\theta,\ \forall\theta$$

Dispersia estimatorului

$$D_\theta^2\left(\widehat{\theta}\right)=\frac{1}{n^2}\sum_{i=1}^{n} D_\theta^2(X_i)=\frac{1}{n}\cdot D_\theta^2(X)$$

$$D_\theta^2(X)=\frac{1}{\theta}\int_0^\infty x^2\cdot\exp\left(-\frac{x}{\theta}\right)dx-\theta^2=\theta^2$$

$$D_\theta^2\left(\widehat{\theta}\right)=\frac{\theta^2}{n}$$

Informatia Fisher

$$\begin{aligned} i_1(\theta)&=M_\theta\left(\frac{\partial\ln f(X;\theta)}{\partial\theta}\right)^2=M_\theta\left(\frac{1}{\theta^2}(X-\theta)\right)^2 \\ &=\frac{1}{\theta^4}\cdot D_\theta^2(X)=\frac{1}{\theta^2}\end{aligned}$$

$$i_n(\theta)=n\cdot i_1(\theta)=\frac{n}{\theta^2}$$

Marginea Rao Cramer

$$\frac{1}{i_n(\theta)}=\frac{\theta^2}{n}=D_\theta^2\left(\widehat{\theta}\right)$$

Deci $\widehat{\theta}(X_1,...,X_n)=\overline{X}$ este estimator eficient al lui θ.

TEOREMA RAO - CRAMER (pentru $k>1$)

Fie modelul $P_\theta\circ X^{-1}$, avand densitatea de repartitie

$$f(x,\theta),\ x\in R,$$

cu $\theta\in\Theta\sqsubseteq R^k$, $k>1$.

Fie observatiile i.i.r. $X_1, ..., X_n$ si notam densitatea de repartitie a vectorului $(X_1, ..., X_n)$ cu

$$f(x_1, ..., x_n; \theta) = \prod_{i=1}^{n} f(x_i; \theta)$$

Fie

$$\widehat{\theta}(X_1, ..., X_n) = \left(\widehat{\theta}_1(X_1, ..., X_n), ..., \widehat{\theta}_k(X_1, ..., X_n)\right)'$$

un estimator nedeplasat pentru $\theta = (\theta_1, ..., \theta_k)'$.

Presupunem verificate urmatoarele conditii de regularitate:

- Θ este multime deschisa;
- $f(x_1, ..., x_n; \theta)$ derivabila partial in raport cu θ_i, $i = 1, ..., k$, oricare ar fi $(x_1, ..., x_n)$, cu derivatele partiale integrabile pe R^n;
- Pentru orice θ, au loc egalitatile

$$\frac{\partial}{\partial\theta_i} \int_{R^n} f(x_1, ..., x_n; \theta)\, dx_1...dx_n = \int_{R^n} \frac{\partial f(x_1, ..., x_n; \theta)}{\partial\theta_i} dx_1...dx_n,\ i = 1, ..., k$$

$$\frac{\partial}{\partial\theta_i} \int_{R^n} \widehat{\theta}_j(x_1, ..., x_n) \cdot f(x_1, ..., x_n; \theta)\, dx_1...dx_n$$
$$= \int_{R^n} \widehat{\theta}_j(x_1, ..., x_n) \cdot \frac{\partial f(x_1, ..., x_n; \theta)}{\partial\theta_i} dx_1...dx_n,\ i, j = 1, ..., k$$

- Exista si este pozitiv definita "matricea informationala Fisher"

$$\left\| M_\theta \left(\frac{\partial \ln f(X_1, ..., X_n; \theta)}{\partial\theta_i} \cdot \frac{\partial \ln f(X_1, ..., X_n; \theta)}{\partial\theta_j} \right) \right\|_{i,j=1,...,k} \overset{notat}{=} I_n(\theta)$$

Atunci matricea

$$Cov_\theta\left(\widehat{\theta}, \widehat{\theta}\right) - I_n^{-1}(\theta)$$

este semipozitiv definita.

Remarca:

$$I_n(\theta) = n \cdot I_1(\theta)$$

METODA VEROSIMILITATII MAXIME

Fie modelul

$$P_\theta \circ X^{-1} = \begin{cases} \sum\limits_{x \in A} p(x;\theta) \cdot \delta_{\{x\}}, & \text{caz discret} \\ \text{sau} & \\ f(x;\theta) \cdot l, \quad x \in R, & \text{caz continuu} \end{cases}$$

Fie $X_1, ..., X_n$ observatii i.i.r. si $(S^n, \mathcal{S}^n)$ spatiul $n-$dimensional al valorilor de selectie.

Definitii

- Pentru datele statistice $(x_1, ..., x_n) \in S^n$, functia de verosimilitate este definita prin

$$L(x_1, ..., x_n; \theta) = \begin{cases} p(x_1, ..., x_n; \theta) = \prod\limits_{i=1}^{n} p(x_i; \theta), & \text{caz discret} \\ \text{sau} & \\ f(x_1, ..., x_n; \theta) = \prod\limits_{i=1}^{n} f(x_i; \theta), & \text{caz continuu} \end{cases}$$

- Fie functia masurabila $\widehat{\theta} : S^n \longrightarrow \Theta$. Functia $\widehat{\theta}(X_1, ..., X_n)$ se numeste estimator de verosimilitate maxima (E.V.M.) daca, pentru orice $(x_1, ..., x_n)$, valoarea $\widehat{\theta}(x_1, ..., x_n)$ este solutia problemei de optimizare

$$\sup_{\theta \in \Theta} L(x_1, ..., x_n; \theta)$$

sau a problemei echivalente

$$\sup_{\theta \in \Theta} \ln L(x_1, ..., x_n; \theta)$$

Notatie: $\widehat{\theta}_{VM}$ (Maximum Likelihood Estimator)

Comentariu:
In cazul discret,

$$L(x_1, ..., x_n; \theta) = P_\theta(X_i = x_i,\ i = 1, ..., n)$$

$\widehat{\theta}_{VM}(x_1, ..., x_n)$ este acea valoare a parametrului θ care face da datele statistice $(x_1, ..., x_n)$ sa fie cel mai verosimile.

APLICATIA 1

E.V.M. pentru parametrul θ al repartitiei $B\left(1,\theta\right)$

Modelul

$$P_{\theta}\circ X^{-1}=\sum_{x=0}^{1}\theta^{x}\left(1-\theta\right)^{1-x}\cdot\delta_{\{x\}},\quad\theta\in\left(0,1\right)$$

Datele statistice

$$\left(x_{1},...,x_{n}\right)\in\left\{0,1\right\}^{n}$$

Functia de verosimilitate

$$L\left(x_{1},...,x_{n};\theta\right)=\prod_{i=1}^{n}\theta^{x_{i}}\left(1-\theta\right)^{1-x_{i}}=\theta^{\sum_{i=1}^{n}x_{i}}\left(1-\theta\right)^{n-\sum_{i=1}^{n}x_{i}}$$

Constructia EVM

$$\ln L=\sum_{i=1}^{n}x_{i}\cdot\ln\theta+\left(n-\sum_{i=1}^{n}x_{i}\right)\cdot\ln\left(1-\theta\right)$$

$$\frac{\partial\ln L}{\partial\theta}=\frac{1}{\theta}\sum_{i=1}^{n}x_{i}-\frac{1}{1-\theta}\left(n-\sum_{i=1}^{n}x_{i}\right)$$

$$\frac{\partial^{2}\ln L}{\partial\theta^{2}}=-\frac{1}{\theta^{2}}\sum_{i=1}^{n}x_{i}-\frac{1}{\left(1-\theta\right)^{2}}\left(n-\sum_{i=1}^{n}x_{i}\right)$$

$$\frac{\partial\ln L}{\partial\theta}=0$$

$$\widehat{\theta}\left(x_{1},...,x_{n}\right)=\frac{1}{n}\sum_{i=1}^{n}x_{i}=\overline{x}$$

$$\frac{\partial^{2}\ln L}{\partial\theta^{2}}\mid_{\overline{x}}=-\frac{n}{\overline{x}\left(1-\overline{x}\right)}<0$$

$$\widehat{\theta}_{VM}\left(X_{1},...,X_{n}\right)=\frac{1}{n}\sum_{i=1}^{n}X_{i}=\overline{X}$$

Proprietatile EVM: vom stabili repartitia exacta a estimatorului, vom cerceta nedeplasarea si vom calcula eroarea medie patratica.

Repartitia lui $\widehat{\theta}_{VM}(X_1,...,X_n)$

Propozitie

Fie variabilele aleatoare independente $Y_i \sim B(r_i,\theta)$,$i = 1.2.$ Atunci $Y_1 + Y_2 \sim B(r_1 + r_2, \theta)$

Rezulta

$$n \cdot \widehat{\theta}_{VM}(X_1,...,X_n) = \sum_{i=1}^{n} X_i \sim B(n,\theta)$$

Eroarea medie patratica pentru $\widehat{\theta}_{VM}(X_1,...,X_n)$

$$M_\theta\left(n \cdot \widehat{\theta}_{VM}\right) = n\theta$$
$$D_\theta^2\left(n \cdot \widehat{\theta}_{VM}\right) = n\theta(1-\theta)$$

$$M_\theta\left(\widehat{\theta}_{VM}\right) = \theta \quad \text{(nedeplasare)}$$
$$D_\theta^2\left(\widehat{\theta}_{VM}\right) = \frac{\theta(1-\theta)}{n}$$
$$M_\theta\left(\widehat{\theta}_{VM} - \theta\right)^2 = \frac{\theta(1-\theta)}{n}$$

APLICATIA 2

E.V.M. pentru parametrul θ al repartitiei Uniforme $U(0,\theta)$

Modelul

$$P_\theta \circ X^{-1} = f(x;\theta) \cdot l$$
$$f(x;\theta) = \begin{cases} \frac{1}{\theta}, & x \in [0,\theta] \\ 0, & \text{in rest} \end{cases} \quad , \quad \theta \in (0,\infty)$$
$$F_X(y) = P_\theta(Y < y) = \begin{cases} 0, & y < 0 \\ \frac{y}{\theta}, & y \in [0,\theta] \\ 1, & y > \theta \end{cases}$$
$$M_\theta(X) = \int_0^\theta \frac{x}{\theta} dx = \frac{\theta}{2}$$

$$D^2_\theta(X)=\int_0^\theta \frac{x^2}{\theta}dx-\frac{\theta^2}{4}=\frac{\theta^2}{12}$$

Datele statistice

$$(x_1,...,x_n)\in[0,\theta]^n$$

Functia de verosimilitate

$$L(x_1,...,x_n;\theta)=\begin{cases}\frac{1}{\theta^n}, & x_i\in[0,\theta],\ i=1,...,n\\ 0, & \text{in rest}\end{cases}$$

$$L(x_1,...,x_n;\theta)=\begin{cases}\frac{1}{\theta^n}, & 0\le \max_i x_i\le\theta\\ 0, & \theta<\max_i x_i\end{cases}$$

Constructia EVM

$$\max_{\theta\in(0,\infty)} L(x_1,...,x_n;\theta)=\frac{1}{\left(\max_i x_i\right)^n}$$

se atinge pentru

$$\widehat{\theta}_{VM}(x_1,...,x_n)=\max_i x_i \overset{notat}{=} x_{(n)}$$

E.V.M. este

$$\widehat{\theta}_{VM}(X_1,...,X_n)=\max_i X_i \overset{notat}{=} X_{(n)}$$

Repartitia lui $\widehat{\theta}_{VM}(X_1,...,X_n)$

$$F_{\widehat{\theta}_{VM}}(y)=F_{X_{(n)}}(y)=P_\theta\left(X_{(n)}<y\right)=\prod_{i=1}^n P_\theta(X_i<y)=(F_X(y))^n$$

$$F_{\widehat{\theta}_{VM}}(y)=\begin{cases}0, & y<0\\ \left(\frac{y}{\theta}\right)^n, & y\in[0,\theta]\\ 1, & y>\theta\end{cases}$$

$$f_{\widehat{\theta}_{VM}}(y)=\begin{cases}\frac{n}{\theta^n}y^{n-1}, & y\in[0,\theta]\\ 0, & \text{in rest}\end{cases}$$

Eroarea medie patratica a lui $\widehat{\theta}_{VM}(X_1,...,X_n)$

$$M_\theta\left(\widehat{\theta}_{VM}\right)=\int_0^\theta y\cdot\frac{n}{\theta^n}y^{n-1}dy=\frac{n}{n+1}\cdot\theta$$

$$Bias\left(\widehat{\theta}_{VM}\right) = \frac{n}{n+1} \cdot \theta - \theta = -\frac{1}{n+1} \cdot \theta$$

$$M_{\theta}\left(\widehat{\theta}_{VM}\right)^2 = \int_0^{\theta} y^2 \cdot \frac{n}{\theta^n} y^{n-1} dy = \frac{n}{n+2} \cdot \theta^2$$

$$D_{\theta}^2\left(\widehat{\theta}_{VM}\right) = \frac{n}{n+2} \cdot \theta^2 - \left(\frac{n}{n+1}\right)^2 \cdot \theta^2 = \frac{n}{(n+2)(n+1)^2} \cdot \theta^2$$

$$M_{\theta}\left(\widehat{\theta}_{VM} - \theta\right)^2 = \frac{n}{(n+2)(n+1)^2} \cdot \theta^2 + \frac{1}{(n+1)^2} \cdot \theta^2 = \frac{2\theta^2}{(n+1)(n+2)}$$

Construim un estimator nedeplasat

$$\widehat{\theta}(X_1, ..., X_n) = \frac{n+1}{n} \cdot \widehat{\theta}_{VM}(X_1, ..., X_n)$$

$$M_{\theta}\left(\widehat{\theta}\right) = \theta$$

$$D_{\theta}^2\left(\widehat{\theta}\right) = \left(\frac{n+1}{n}\right)^2 \cdot \frac{n}{(n+2)(n+1)^2} \cdot \theta^2 = \frac{\theta^2}{n(n+2)}$$

$$M_{\theta}\left(\widehat{\theta} - \theta\right)^2 = \frac{\theta^2}{n(n+2)}$$

Comparam cei doi estimatori

$$\frac{M_{\theta}\left(\widehat{\theta}_{VM} - \theta\right)^2}{M_{\theta}\left(\widehat{\theta} - \theta\right)^2} = \frac{2n}{n+1} > 1, \quad n > 1$$

$$M_{\theta}\left(\widehat{\theta} - \theta\right)^2 < M_{\theta}\left(\widehat{\theta}_{VM} - \theta\right)^2$$

APLICATIA 3

E.V.M. pentru parametrul $\theta = (\mu, \sigma^2)$ al repartitiei Normale $N(\mu, \sigma^2)$

Modelul

$$P_\theta \circ X^{-1} = f\left(x; \mu, \sigma^2\right) \cdot l$$

$$f\left(x; \mu, \sigma^2\right) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\}$$

$$M_\theta(X) = \mu$$

$$D_\theta^2(X) = \sigma^2$$

Datele statistice

$$(x_1, ..., x_n) \in R^n$$

Functia de verosimilitate

$$L\left(x_1, ..., x_n; \mu, \sigma^2\right) = \frac{1}{\left(2\pi\sigma^2\right)^{n/2}} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right\}$$

Constructia EVM

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

$$\frac{\partial \ln L}{\partial \mu} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu)$$

$$\frac{\partial \ln L}{\partial \sigma^2} = -\frac{n}{2}\cdot\frac{1}{\sigma^2} + \frac{1}{2\left(\sigma^2\right)^2}\sum_{i=1}^{n}(x_i-\mu)^2$$

$$\frac{\partial^2 \ln L}{\partial \mu^2} = -\frac{n}{\sigma^2}$$

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

$$\frac{\partial^2 \ln L}{\partial\left(\sigma^2\right)^2} = \frac{n}{2}\cdot\frac{1}{\left(\sigma^2\right)^2} - \frac{1}{\left(\sigma^2\right)^3}\sum_{i=1}^{n}(x_i-\mu)^2$$

Sistemul de verosimilitate maxima

$$\begin{cases} \frac{\partial \ln L}{\partial \mu} = 0 \\ \frac{\partial \ln L}{\partial \sigma^2} = 0 \end{cases}$$

$$\begin{cases} \sum\limits_{i=1}^{n}(x_i-\mu)=0 \\ -n\sigma^2+\sum\limits_{i=1}^{n}(x_i-\mu)^2=0 \end{cases}$$

$$\widehat{\mu}(x_1,...,x_n)=\frac{1}{n}\sum_{i=1}^{n}x_i=\overline{x}$$
$$\widehat{\sigma^2}(x_1,...,x_n)=\frac{1}{n}\sum_{i=1}^{n}(x_i-\overline{x})^2$$

$$\frac{\partial^2\ln L}{\partial\mu^2}\mid_{\left(\widehat{\mu},\widehat{\sigma^2}\right)}=-\frac{n}{\widehat{\sigma^2}}<0$$
$$\frac{\partial^2\ln L}{\partial\mu\partial\sigma^2}\mid_{\left(\widehat{\mu},\widehat{\sigma^2}\right)}=0$$
$$\frac{\partial^2\ln L}{\partial(\sigma^2)^2}\mid_{\left(\widehat{\mu},\widehat{\sigma^2}\right)}=-\frac{n}{2}\cdot\frac{1}{\left(\widehat{\sigma^2}\right)^2}<0$$

Rezulta ca $\left(\widehat{\mu}(x_1,...,x_n),\widehat{\sigma^2}(x_1,...,x_n)\right)$ este punct de maxim pentru $\ln L$, iar EVM este

$$\left(\widehat{\mu}_{VM},\widehat{\sigma^2}_{VM}\right)(X_1,...,X_n)=\left(\overline{X},\frac{1}{n}\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2\right)$$

Pentru a stabili repartitia lui $\left(\widehat{\mu}_{VM},\widehat{\sigma^2}_{VM}\right)$ avem nevoie de "definitia constructiva" a repartitiei CHI Patrat

Repartitia $Gamma(\alpha,\theta)$

Repartitia $\chi^2(r)$

Definitie
Variabila aleatoare X are o repartitie $Gamma(\alpha,\theta),\alpha,\theta\in(0,\infty)$, daca are densitatea de repartitie

$$f(y)=\begin{cases} \frac{1}{\theta^\alpha\Gamma(\alpha)}y^{\alpha-1}\exp\left(-\frac{y}{\theta}\right), & y\geq 0 \\ 0, & y<0 \end{cases}$$

Reamintim

$$\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t} dt$$

$$\Gamma(\alpha) = (\alpha-1)\Gamma(\alpha-1)$$

$$\Gamma(r) = (r-1)!, \quad r \in N^*$$

$$M(Y) = \int_0^\infty \frac{1}{\theta^\alpha \Gamma(\alpha)} y^\alpha \exp\left(-\frac{y}{\theta}\right) dy = \frac{\theta^{\alpha+1}\Gamma(\alpha+1)}{\theta^\alpha \Gamma(\alpha)} = \theta\alpha$$

$$M\left(Y^2\right) = \int_0^\infty \frac{1}{\theta^\alpha \Gamma(\alpha)} y^{\alpha+1} \exp\left(-\frac{y}{\theta}\right) dy = \frac{\theta^{\alpha+2}\Gamma(\alpha+2)}{\theta^\alpha \Gamma(\alpha)} = \theta^2\alpha(\alpha+1)$$

$$D^2(Y) = \theta^2\alpha(\alpha+1) - \theta^2\alpha^2 = \theta^2\alpha$$

$$\varphi_Y(t) = M\left(e^{itY}\right) = \frac{1}{\theta^\alpha \Gamma(\alpha)} \left(\frac{1}{\theta} - it\right)^{-\alpha} \Gamma(\alpha) = (1 - it\theta)^{-\alpha}$$

Proprietatea 2

Fie variabilele aleatoare independente $Y_i \sim Gamma(\alpha_i, \theta), i = 1, 2$. Atunci $Y_1 + Y_2 \sim Gamma(\alpha_1 + \alpha_2, \theta)$

Demonstratie

$$\varphi_{Y_1+Y_2}(t) = \varphi_{Y_1}(t) \cdot \varphi_{Y_2}(t) = (1 - it\theta)^{-\alpha_1+\alpha_2}$$

Definitie

Repartitia $Gamma\left(\frac{r}{2}, 2\right)$, cu $r \in N^*$ se numeste repartitia CHI Patrat cu r grade de libertate, avand densitatea de repartitie

$$f(y) = \frac{1}{2^{r/2}\Gamma\left(\frac{r}{2}\right)} y^{\frac{r}{2}-1} \exp\left(-\frac{y}{2}\right), \quad y \geq 0$$

$$M(Y) = r$$

$$D^2(Y) = 2r$$

Proprietatea 3

Fie $X_1, ..., X_r$ variabile aleatoare independente, identic repartizate Normal $N(0,1)$. Atunci

$$Y = \sum_{i=1}^{r} X_i^2$$

este repartizata $\chi^2(r)$.

Demonstratie:

$$P\left(X_1^2 < z\right) = \begin{cases} 0, & z < 0 \\ P\left(|X_1| < \sqrt{z}\right), & z \geq 0 \end{cases} = \begin{cases} 0, & z < 0 \\ \frac{2}{\sqrt{2\pi}} \int\limits_0^{\sqrt{z}} e^{-x^2/2} dx, & z \geq 0 \end{cases}$$

$$f_{X_1^2}(z) = \begin{cases} 0, & z < 0 \\ \frac{2}{\sqrt{2\pi}} \cdot e^{-z/2} \cdot \frac{1}{2\sqrt{z}}, & z \geq 0 \end{cases}$$

$$f_{X_1^2}(z) = \frac{1}{2^{1/2}\Gamma\left(\frac{1}{2}\right)} \cdot z^{\frac{1}{2}-1} \cdot e^{-z/2}, \quad z \geq 0$$

Adica X_1^2 este repartizata $\chi^2(1) = Gamma\left(\frac{1}{2}, 2\right)$.

Avem $X_1^2, ..., X_r^2$ variabile aleatoare independente, identic repartizate $Gamma\left(\frac{1}{2}, 2\right)$. Rezulta

$$\sum_{i=1}^{r} X_i^2 \sim Gamma\left(\frac{r}{2}, 2\right) = \chi^2(r).$$

■

Proprietatea 4

Fie $Y_1, ..., Y_n$ variabile aleatoare independente, identic repartizate Normal $N(0,1)$ si fie

$$\overline{Y} = \frac{1}{n} \sum_{i=1}^{n} Y_i$$

$$H = \sum_{i=1}^{n} \left(Y_i - \overline{Y}\right)^2$$

Atunci $\overline{Y} \sim N\left(0, \frac{1}{n}\right)$, $H \sim \chi^2(n-1)$, iar $\overline{Y}$ si H sunt variabile aleatoare independente.

Demonstratie:

$$\sum_{i=1}^{n} Y_i \sim N(0,n) \implies \overline{Y} \sim N\left(0, \frac{1}{n}\right)$$

Notam

$$\mathbf{Y} = (Y_1, ..., Y_n)'$$

Vectorul aleator Y are (prin definitie) o repartitie normala $n-$dimensionala, $N(n; \mathbf{0}, \mathbf{I})$, cu

$$M(\mathbf{Y}) = \mathbf{0} = (0, ..., 0)'$$

$$Cov(\mathbf{Y}, \mathbf{Y}) = \|cov(Y_i, Y_j)\|_{i,j=1,...,n} = \mathbf{I}$$

Consideram transformarea liniara

$$\mathbf{Z} = A \cdot \mathbf{Y}$$

cu

$$A = \begin{pmatrix} \frac{1}{\sqrt{1\cdot 2}} & \frac{-1}{\sqrt{1\cdot 2}} & 0 & ..... & 0 \\ \frac{1}{\sqrt{2\cdot 3}} & \frac{1}{\sqrt{2\cdot 3}} & \frac{-2}{\sqrt{2\cdot 3}} & ..... & 0 \\ ... & ... & ... & ..... & ... \\ \frac{1}{\sqrt{(n-1)n}} & \frac{1}{\sqrt{(n-1)n}} & \frac{1}{\sqrt{(n-1)n}} & ..... & \frac{-(n-1)}{\sqrt{(n-1)n}} \\ \frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}} & ..... & \frac{1}{\sqrt{n}} \end{pmatrix}$$

Avem $A \cdot A' = \mathbf{I}$.

Vectorul aleator $\mathbf{Z} = (Z_1, ..., Z_n)'$ are o repartitie normala $n-$dimensionala, cu

$$M(\mathbf{Z}) = A \cdot M(\mathbf{Y}) = \mathbf{0}$$

$$Cov(\mathbf{Z}, \mathbf{Z}) = M\left(\mathbf{Z} \cdot \mathbf{Z}'\right) = \left(A \cdot \mathbf{Y} \cdot \mathbf{Y}' \cdot \mathbf{A}'\right) = A \cdot Cov(\mathbf{Y}, \mathbf{Y}) \cdot A' = A \cdot \mathbf{I} \cdot A' = \mathbf{I}$$

Componentele lui $\mathbf{Z}$ sunt variabile aleatoare independente, identic repartizate $N(0,1)$.Observam ca:

$$\sum_{i=1}^{n} Z_i^2 = \mathbf{Z}'\mathbf{Z} = \mathbf{Y}' \cdot A' \cdot A \cdot \mathbf{Y} = \mathbf{Y}'\mathbf{Y} = \sum_{i=1}^{n} Y_i^2$$

Dar

$$Z_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} Y_i = \sqrt{n} \cdot \overline{Y}.$$

$$\sum_{i=1}^{n-1} Z_i^2 = \sum_{i=1}^{n} Y_i^2 - Z_n^2 = \sum_{i=1}^{n} Y_i^2 - n\left(\overline{Y}\right)^2 = \sum_{i=1}^{n} \left(Y_i - \overline{Y}\right)^2 = H$$

Deci

$$\overline{Y} = \frac{1}{\sqrt{n}} Z_n,$$
$$H = \sum_{i=1}^{n-1} Z_i^2$$

Rezulta ca $\overline{Y}$ si H sunt variabile aleatoare independente si $H \sim \chi^2 (n-1)$.

■

Revenim la problema repartitiei E.V.M.

$$\left(\widehat{\mu}_{VM}, \widehat{\sigma^2}_{VM}\right)(X_1, ..., X_n) = \left(\overline{X}, \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2\right)$$

Proprietatea 5

Fie $X_1, ..., X_n$ variabile aleatoare independente, identic repartizate $N(\mu, \sigma^2)$ si fie $\left(\widehat{\mu}_{VM}, \widehat{\sigma^2}_{VM}\right)$ E.V.M. construit mai sus. Atunci

$$\widehat{\mu}_{VM} = \overline{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right),$$
$$\frac{n}{\sigma^2} \cdot \widehat{\sigma^2}_{VM} \sim \chi^2 (n-1)$$

si cele doua componente ale E.V.M. sunt independente.

Demonstratie:

Aplicam Proprietatea 4 pentru

$$Y_i = \frac{X_i - \mu}{\sigma} \sim N(0,1), \quad i = 1, ..., n$$
$$\overline{Y} = \frac{1}{n}\sum_{i=1}^{n} \frac{X_i - \mu}{\sigma} = \frac{\overline{X} - \mu}{\sigma}$$
$$H = \sum_{i=1}^{n}\left(\frac{X_i - \mu}{\sigma} - \frac{\overline{X} - \mu}{\sigma}\right)^2 = \frac{1}{\sigma^2}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2 = \frac{n}{\sigma^2} \cdot \widehat{\sigma^2}_{VM}$$

Rezulta ca $\frac{\overline{X}-\mu}{\sigma}$ are repartitia $N\left(0, \frac{1}{n}\right)$, adica $\overline{X}$ are repartitia $N\left(\mu, \frac{\sigma^2}{n}\right)$, iar $\frac{n}{\sigma^2} \cdot \widehat{\sigma^2}_{VM}$ are repartitia $\chi^2 (n-1)$.

Independenta celor doua componente ale E.V.M. rezulta tot din proprietatea 4.

■

EROARILE MEDII PATRATICE ALE COMPONENTELOR E.V.M. $\left(\widehat{\mu}_{VM}, \widehat{\sigma^2}_{VM}\right)$

$$
\begin{aligned}
M_\theta\left(\overline{X}\right) &= \mu \\
Bias\left(\overline{X}\right) &= 0 \\
D^2_\theta\left(\overline{X}\right) &= \frac{\sigma^2}{n} \\
M_\theta\left(\overline{X}-\mu\right)^2 &= \frac{\sigma^2}{n}
\end{aligned}
$$

$$
\begin{aligned}
M_\theta\left(\widehat{\sigma^2}_{VM}\right) &= \frac{n-1}{n}\sigma^2 \\
Bias\left(\widehat{\sigma^2}_{VM}\right) &= \frac{n-1}{n}\sigma^2 - \sigma^2 = -\frac{\sigma^2}{n} \\
D^2_\theta\left(\widehat{\sigma^2}_{VM}\right) &= \frac{2(n-1)}{n^2}\sigma^4 \\
M_\theta\left(\widehat{\sigma^2}_{VM} - \sigma^2\right)^2 &= \frac{2(n-1)}{n^2}\sigma^4 + \frac{\sigma^4}{n^2} = \frac{2n-1}{n^2}\sigma^4
\end{aligned}
$$

Putem construi un estimator nedeplasat pentru σ^2 :

$$S^2 = \frac{n}{n-1}\widehat{\sigma^2}_{VM} = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2$$

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

$$
\begin{aligned}
M_\theta\left(S^2\right) &= \sigma^2 \\
Bias\left(S^2\right) &= 0 \\
D^2_\theta\left(S^2\right) &= \frac{2\sigma^4}{n-1} \\
M_\theta\left(S^2-\sigma^2\right)^2 &= \frac{2\sigma^4}{n-1}
\end{aligned}
$$

Observam ca, desi S^2 este un estimator nedeplasat pentru σ^2, eroarea sa medie patratica este mai mare decat cea a lui $\widehat{\sigma^2}_{VM}$:

$$\frac{M_\theta\left(\widehat{\sigma^2}_{VM} - \sigma^2\right)^2}{M_\theta\left(S^2-\sigma^2\right)^2} = \frac{(2n-1)(n-1)}{2n^2} < 1$$

METODA CELOR MAI MICI PATRATE

Se adreseaza estimarii parametrilor "MODELELOR LINIARE"

MODELUL LINIAR $n-$DIMENSIONAL,

CU OBSERVATII INDEPENDENTE

Fie un sir de variabile aleatoare independente, neidentic repartizate, de forma

$$X_i = M_\theta\left(X_i\right) + Z_i, \quad i = 1, 2, ...$$

unde:

- $\{Z_i,\ i = 1, 2, ...\}$ sunt v.a. indep, identic repartizate, cu $M_\theta\left(Z_i\right) = 0$,$D_\theta^2\left(Z_i\right) = \sigma^2$,$\forall i$
- $M_\theta\left(X_i\right) = y_i'\theta = \sum_{j=1}^{k} y_{ij}\theta_j$,$i = 1, 2, ...$
- $\theta = \left(\theta_1, ..., \theta_k\right)' \in \Theta \sqsubseteq R^k$,$k \geq 1$

Observam primele n variabile ale sirului, $n > k,$ si notam

$$\mathbf{X} = \left(X_1, ..., X_n\right)'$$
$$\mathbf{Z} = \left(Z_1, ..., Z_n\right)'$$

$$\mathbf{Y} = \left\|y_{ij}\right\|_{i=1,...,n;\ j=1,...k}$$

Definitie:

Secventa de n variabile aleatoare independente, neidentic repartizate, de forma

$$X_i = \mathbf{y}_i'\theta + Z_i, \quad i = 1, 2, ...n$$

se numeste model liniar $n-$dimensional, cu observatii independente.

Are loc scrierea matriceala

$$\mathbf{X} = \mathbf{Y}\theta + \mathbf{Z}$$

Exemplu:

- $X =$ cresterea lunara in greutate la copilul de 12 - 18 luni

Cresterea in greutate depinde de regimul alimentar administrat (ratia zilnica de proteine, ratia zilnica de glucide, ratia zilnica de lipide)

- "regim alimentar" $=(y_1, y_2, y_3)'$ va fi specifiat (cunoscut) pt fiecare copil luat in studiu

$$X = y_1\theta_1 + y_2\theta_2 + y_3\theta_3 + Z$$

- parametrul necunoscut $\theta = (\theta_1, \theta_2, \theta_3)'$ exprima influenta fiecarui principiu nutritiv asupra cresterii in greutate
- n copii sunt inclusi in studiu in mod independent unul de altul si se dau $y_i = (y_{i1}, y_{i2}, y_{i3})'$,$i = 1, ..., n$
- se inregistreaza cresterile in greutate din luna in care are loc studiul, $(x_1, ..., x_n)$
- se estimeaza θ

Proprietati ale modelului

$$M_\theta(\mathbf{Z}) = (M_\theta(Z_1), ..., M_\theta(Z_n))' = (0, ..., 0)' = \mathbf{0}$$

$$Cov_\theta(\mathbf{Z}, \mathbf{Z}) = \|cov_\theta(Z_i, Z_j)\|_{i,j=1,\dots,n} = \sigma^2 \cdot \mathbf{I}$$

$$M_\theta(\mathbf{X}) = \mathbf{Y}\theta + M_\theta(\mathbf{Z}) = \mathbf{Y}\theta$$

$$Cov_\theta(\mathbf{X}, \mathbf{X}) = Cov_\theta(\mathbf{Z}, \mathbf{Z}) = \sigma^2 \cdot \mathbf{I}$$

Definitii:

Modelul liniar $n-$dimensional $\mathbf{X} = \mathbf{Y}\theta + \mathbf{Z}$ se numeste nesingular daca rangul matricii $\mathbf{Y}$ este maximal,

$$rang(\mathbf{Y}) = k$$

Modelul liniar $n-$dimensional $\mathbf{X} = \mathbf{Y}\theta + \mathbf{Z}$ se numeste ortogonal daca caloanele lui $\mathbf{Y}$ sunt vectori ortogonali din R^n.

Modelul liniar $n-$dimensional $\mathbf{X} = \mathbf{Y}\theta + \mathbf{Z}$ se numeste normal daca variabilele aleatoare indep, id. repartizate $Z_1, ..., Z_n$ au repartitie normala, $N(0, \sigma^2)$.

Fie $x = (x_1, ..., x_n)'$ datele statistice observate.
Suma abaterilor patratice (Sum of Squares)

$$SS(x_1, ..., x_n; \theta) = \sum_{i=1}^{n} (x_i - \mathbf{y}_i'\theta)^2 = (\mathbf{x} - \mathbf{Y}\theta)'(\mathbf{x} - \mathbf{Y}\theta) = \|(\mathbf{x} - \mathbf{Y}\theta)\|^2$$

Definitie
Estimatorul $\widehat{\theta}(X_1, ..., X_n)$ se numeste estimator prin metoda celor mai mici patrate (Least Squares Estimator, (L.S.E.)) daca, pentru orice $x = (x_1, ..., x_n)'$, valoarea $\widehat{\theta}(x_1, ..., x_n)$ se obtine ca solutie a problemei de optimizare

$$\inf_{\theta \in \Theta} SS(x_1, ..., x_n; \theta)$$

Estimatorul se noteaza $\widehat{\theta}_{LS}(X_1, ..., X_n)$.

Fie $SS(x_1, ..., x_n; \theta)$. Sistemul

$$\frac{\partial SS}{\partial \theta} = \mathbf{0}$$

se numeste sistemul de ecuatii normale. Explicit, sistemul liniar se scrie:

$$\mathbf{Y}'(\mathbf{x} - \mathbf{Y}\theta) = \mathbf{0}$$

sau

$$\mathbf{Y}'\mathbf{Y}\theta = \mathbf{Y}'\mathbf{x}$$

Proprietatea 6 (existenta L.S.E.)

Un estimator $\widehat{\theta}$ este L.S.E, $\widehat{\theta} = \widehat{\theta}_{LS}$, daca si numai daca, pentru orice $x = (x_1, ..., x_n)'$, valoarea $\widehat{\theta}(x_1, ..., x_n)$ este solutia sistemului de ecuatii normale $\mathbf{Y}'\mathbf{Y}\theta = \mathbf{Y}'\mathbf{x}$.

Demonstratie:
Fie $\mathbf{x} = (x_1, ..., x_n)'$ arbitrar fixat.

$$\inf_{\theta \in \Theta} SS(\mathbf{x}; \theta) \quad \Leftrightarrow \quad \inf_{\theta \in \Theta} \|(\mathbf{x} - \mathbf{Y}\theta)\|^2$$

Fie $\mathcal{L}$ spatiul liniar generat de coloanele liniar independente ale lui $\mathbf{Y}$ (subspatiu liniar al lui R^n).

Solutia problemei

$$\inf_{\mathbf{z}\in\mathcal{L}} \|(\mathbf{x}-\mathbf{z})\|^2$$

este

$$\mathbf{z}^* = pr_{\mathcal{L}}(\mathbf{x})$$

Atunci,

$$\widehat{\theta}(\mathbf{x}) = \widehat{\theta}_{LS}(\mathbf{x}) \quad \Leftrightarrow \quad \mathbf{Y}\widehat{\theta}(\mathbf{x}) = pr_{\mathcal{L}}(\mathbf{x}) \quad \Leftrightarrow$$
$$\mathbf{x}-\mathbf{Y}\widehat{\theta}(\mathbf{x}) \perp \mathcal{L} \quad \Leftrightarrow \quad \mathbf{Y}'\left(\mathbf{x}-\mathbf{Y}\widehat{\theta}(\mathbf{x})\right) = \mathbf{0}$$

■

Proprietatea 7 (L.S.E. este cel mai bun estimator liniar nedeplasat al lui θ)

Fie modelul liniar $n-$dimensional cu observatii independente $\mathbf{X} = \mathbf{Y}\theta+\mathbf{Z}$.

Presupunem modelul nesingular $(rang(\mathbf{Y}) = k < n)$.

Atunci sistemul de ecuatii normale are solutia unica

$$\widehat{\theta}_{LS}(\mathbf{x}) = (\mathbf{Y}'\mathbf{Y})^{-1}\mathbf{Y}'\mathbf{x},$$

si estimatorul $\widehat{\theta}_{LS}(\mathbf{X})$ verifica urmatoarele proprietati:

- este nedeplasat,

$$M_\theta\left(\widehat{\theta}_{LS}(\mathbf{X})\right) = \theta, \quad \forall\theta\in\Theta,$$

- pentru orice estimator $\mathbf{g}$ liniar, nedeplasat al lui θ,matricea

$$Cov_\theta(\mathbf{g},\mathbf{g}) - Cov_\theta\left(\widehat{\theta}_{LS},\widehat{\theta}_{LS}\right)$$

  este semipozitiv definita, $\forall\theta\in\Theta$.

Demonstratie:

Cum $rang(\mathbf{Y}) = k$, rezulta $rang(\mathbf{Y}'\mathbf{Y}) = k$, deci $\mathbf{Y}'\mathbf{Y}\theta= \mathbf{Y}'\mathbf{x}$ este sistem Cramer, cu solutia unica $\widehat{\theta}_{LS}(\mathbf{x}) = (\mathbf{Y}'\mathbf{Y})^{-1}\mathbf{Y}'\mathbf{x}$.

$$M_\theta\left(\widehat{\theta}_{LS}\right) = \left(\mathbf{Y}'\mathbf{Y}\right)^{-1}\mathbf{Y}'M_\theta\left(\mathbf{X}\right) = \left(\mathbf{Y}'\mathbf{Y}\right)^{-1}\mathbf{Y}'\mathbf{Y}\theta = \theta, \quad \forall\theta\in\Theta$$

$$\begin{aligned} Cov_\theta\left(\widehat{\theta}_{LS},\widehat{\theta}_{LS}\right) &= \left(\mathbf{Y}'\mathbf{Y}\right)^{-1}\mathbf{Y}'Cov_\theta\left(\mathbf{X},\mathbf{X}\right)\mathbf{Y}\left(\mathbf{Y}'\mathbf{Y}\right)^{-1} = \\ &= \left(\mathbf{Y}'\mathbf{Y}\right)^{-1}\mathbf{Y}'\cdot\sigma^2\mathbf{I}\cdot\mathbf{Y}\left(\mathbf{Y}'\mathbf{Y}\right)^{-1} = \sigma^2\left(\mathbf{Y}'\mathbf{Y}\right)^{-1} \end{aligned}$$

Fie $\mathbf{g}(\mathbf{X}) = R\mathbf{X}$ un estimator liniar, nedeplasat pentru θ. Conditia de nedeplasare revine la

$$M_\theta(\mathbf{g}) = \theta, \ \forall\theta\in\Theta,$$

respectiv la

$$R\mathbf{Y}\theta = \theta, \ \forall\theta\in\Theta,$$

adica $R\mathbf{Y} = \mathbf{I}$.

$$Cov_\theta(\mathbf{g},\mathbf{g}) = R\cdot Cov_\theta(\mathbf{X},\mathbf{X})\cdot R' = \sigma^2 RR'$$

$$Cov_\theta(\mathbf{g},\mathbf{g}) - Cov_\theta\left(\widehat{\theta}_{LS},\widehat{\theta}_{LS}\right) = \sigma^2 RR' - \sigma^2\left(\mathbf{Y}'\mathbf{Y}\right)^{-1}$$

Folosind relatia $R\mathbf{Y} = \mathbf{I}$ obtinem

$$Cov_\theta(\mathbf{g},\mathbf{g}) - Cov_\theta\left(\widehat{\theta}_{LS},\widehat{\theta}_{LS}\right) = \sigma^2\left(R - \left(\mathbf{Y}'\mathbf{Y}\right)^{-1}\mathbf{Y}'\right)\left(R - \left(\mathbf{Y}'\mathbf{Y}\right)^{-1}\mathbf{Y}'\right)'$$

Notam $\Gamma = R - \left(\mathbf{Y}'\mathbf{Y}\right)^{-1}\mathbf{Y}'$ si obtinem

$$\mathbf{z}'\left(Cov_\theta(\mathbf{g},\mathbf{g}) - Cov_\theta\left(\widehat{\theta}_{LS},\widehat{\theta}_{LS}\right)\right)\mathbf{z} = \sigma^2\mathbf{z}'\Gamma\Gamma'z = \sigma^2\left(\Gamma'\mathbf{z}\right)'\left(\Gamma'\mathbf{z}\right) \geq 0$$

■

- valorile observate: x_i, $i = 1, ..., n$
- predictorii (fitted values): $\widehat{x}_i = \mathbf{y}_i'\widehat{\theta}_{LS}, \quad i = 1, ..., n$
- reziduuri (residuals) $x_i - \widehat{x}_i$, $i = 1, ..., n$

Definim variabila aleatoare "Suma reziduurilor patratice"

$$SS_{rezid} = \sum_{i=1}^{n}\left(X_i - \mathbf{y}_i'\widehat{\theta}_{LS}\right)^2 = \left\|\mathbf{X} - \mathbf{Y}\widehat{\theta}_{LS}\right\|^2$$

Proprietatea 8

Fie modelul liniar $n-$dimensional cu observatii independente $\mathbf{X} = \mathbf{Y}\theta + \mathbf{Z}$.

Presupunem modelul nesingular si normal. Atunci

$$\frac{1}{\sigma^2} \cdot SS_{rezid} \quad \sim \quad \chi^2 (n-k)$$

Demonstratie:

Fie $\mathcal{L}$ spatiul liniar generat de coloanele liniar independente ale lui $\mathbf{Y}$.

$$\dim \mathcal{L} = rang\ \mathbf{Y} \ = k$$
$$\dim \mathcal{L}^{\perp} = n-k$$

Fie $\{\mathbf{u}_{k+1}, ..., \mathbf{u}_n\}$ o baza ortonormata pentru $\mathcal{L}^{\perp}$.

Pentru $\mathbf{x} \in R^n$, avem $\mathbf{Y}\widehat{\theta}_{LS}(\mathbf{x}) \in \mathcal{L}$, $\mathbf{x} - \mathbf{Y}\widehat{\theta}_{LS}(\mathbf{x}) \in \mathcal{L}^{\perp}$. Putem scrie

$$\mathbf{x} - \mathbf{Y}\widehat{\theta}_{LS}(\mathbf{x}) = \sum_{i=k+1}^{n} \mathbf{u}_i'\mathbf{x}$$

$$\frac{1}{\sigma^2} \cdot SS_{rezid} = \sum_{i=k+1}^{n} \left(\frac{\mathbf{u}_i'\mathbf{x}}{\sigma}\right)^2$$

Dar $\{\frac{1}{\sigma}\mathbf{u}_i'\mathbf{X},\ i = k+1, ..., n\}$ sunt var. al. independente, identic repartizate $N(0,1)$ cāci:

- sunt combinatii liniare de componentele normal repartizate ale lui $\mathbf{X} = (X_1, ..., X_n)'$ si

$$M_\theta \left(\frac{1}{\sigma}\mathbf{u}_i'\mathbf{X}\right) = \frac{1}{\sigma}\mathbf{u}_i'\mathbf{Y}\theta = 0, \quad i = k+1, ..., n$$

$$cov_\theta \left(\frac{1}{\sigma}\mathbf{u}_i'\mathbf{X}, \frac{1}{\sigma}\mathbf{u}_j'\mathbf{X}\right) = \frac{1}{\sigma^2}\mathbf{u}_i' Cov_\theta (\mathbf{X}, \mathbf{X})\, \mathbf{u}_j = \frac{1}{\sigma^2}\mathbf{u}_i' \left(\sigma^2 \mathbf{I}\right) \mathbf{u}_j = \mathbf{u}_i'\mathbf{u}_j = \delta_{ij},$$
$$i, j = k+1, ..., n$$

- fiind var al necorelate, identic repartizate normal, $N(0,1)$,sunt si independente.

Rezulta

$$\sum_{i=k+1}^{n} \left(\frac{\mathbf{u}_i'\mathbf{x}}{\sigma}\right)^2 \quad \sim \quad \chi^2 (n-k)$$

■