Instrumente Statistice pentru Finanțe

Estimatorul de verosimilitate maximă

Alexandru Amărioarei

Faculty of Mathematics and Computer Science
University of Bucharest

Master I, Semestrul II, 2020

Outline

Introducere

- *Metoda Verosimilității Maxime* (MLE - maximum likelihood method) este una dintre metodele cel mai des întâlnite în estimarea parametrilor unui model
- Verosimilitatea sau funcția de verosimilitate (unul dintre cele mai importante concepte statistice) stabilește o relație de preferință între posibilele valori ale parametrilor modelului, dat fiind datele (observațiile)
- Metoda verosimilității maxime alege acei parametrii din spațiul parametrilor care maximizează funcția de verosimilitate
- Estimarea prin metoda verosimilității maxime are avantajul unei abordări generale pentru problema de estimare

Întrebări

În această secțiune a cursului încercăm să răspundem la întrebări precum:

- Care sunt principalele proprietăți ale estimatorului de verosimilitate maximă ?
  - Este acest estimator (asimptotic) nedeplasat ?
  - Este consistent ?
  - Este (asimptotic) eficient ? Sub ce condiții ?
  - Care este repartiția limită ?
- Cum putem aplica principiul verosimilității maxime la modelul de regresie multiplă sau la modelul Probit / Logit, etc. ?

Plan de lucru

Materialul este organizat astfel:

1. Principiul estimatorului de verosimilitate maximă
2. Funcția de verosimilitate
3. Estimatorul de verosimilitate maximă
4. Funcția de scor, matricea Hessiană și matricea Informațională a lui Fisher
5. Proprietăți ale estimatorilor de verosimilitate maximă

Principiul MLE

Principiul estimatorului de verosimilitate maximă

Obiective

În această secțiune vom prezenta un exemplu simplu care să ne permită să introducem

- introducem notațiile
- introducem noțiunea de *funcție de verosimilitate* și *logaritmul funcției de verosimilitate*
- introducem conceptul de *estimator de verosimilitate maximă*

Exemplu introductiv

Example

Să presupunem că $X_1, X_2, \ldots, X_n$ este un eșantion (i.i.d.) dintr-o populație $Pois(\theta)$ cu funcția de masă dată de

$$\mathbb{P}(X_i = x_i) = e^{-\theta}\frac{\theta^{x_i}}{x_i!}$$

unde θ este un parametru necunoscut.

Exemplu introductiv (cont.)

Intrebare: Care este probabilitatea de a observa un eșantion $\{x_1, x_2, \ldots, x_n\}$ presupunând că datele au fost generate dintr-o populație Poisson de parametru θ (necunoscut) ?

Această probabilitate este

$$\mathbb{P}(\text{observăm } x_1, x_2, \ldots, x_n) = \mathbb{P}(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n)$$

și din independență avem

$$\begin{aligned}\mathbb{P}(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n) &= \prod_{i=1}^{n} \mathbb{P}(X_i = x_i)\\ &= \prod_{i=1}^{n} e^{-\theta}\frac{\theta^{x_i}}{x_i!} = e^{-n\theta}\frac{\theta^{\sum_{i=1}^{n} x_i}}{\prod_{i=1}^{n} x_i!}\end{aligned}$$

Exemplu introductiv (cont.)

Repartiția comună a observațiilor $x_1, x_2, \ldots, x_n$ este o funcție care depinde de θ (parametrul necunoscut), poartă numele de funcția de verosimilitate a eșantionului și este notată prin

$$L_n(\theta; x_1, x_2, \ldots, x_n) = \mathbb{P}(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n)$$

În cazul exemplului nostru devine

$$L_n(\theta; x_1, x_2, \ldots, x_n) = e^{-n\theta} \frac{\theta^{\sum_{i=1}^{n} x_i}}{\prod_{i=1}^{n} x_i!}$$

Exemplu introductiv (cont.)

Example

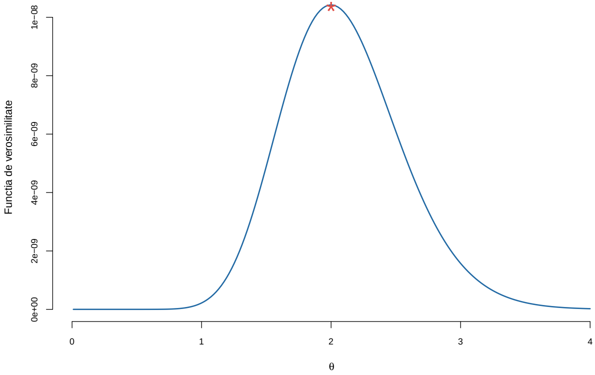
Să presupunem că $n = 10$ și că am observat următorul eșantion $\{5, 0, 1, 1, 0, 3, 2, 3, 4, 1\}$. Atunci

$$L_n(\theta; x_1, x_2, \dots, x_n) = \mathbb{P}(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n) = e^{-10\theta}\frac{\theta^{20}}{207360}.$$

Intrebare: Care este valoarea lui θ care face ca acest eșantion să fie cel mai probabil ?

Exemplu introductiv (cont.)

Figura de mai jos ilustrează funcția de verosimilitate $L_n(\theta; x)$ pentru diferite valori ale parametrului θ. Are un maxim global care se atinge pentru $\theta = 2$ și care reprezintă *estimarea* lui θ prin MLE.



Exemplu introductiv (cont.)

Considerăm problema maximizării funcției de verosimilitate $L_n(\theta; x_1, x_2, \dots, x_n)$ în raport cu θ. În general, pentru a simplifica operațiile prin trecerea de la înmulțire la sumă vom maximiza, în schimb, logartimul funcției de verosimilitate $\log L_n(\theta; x_1, x_2, \dots, x_n)$ (de ce putem face acest lucru ?). Avem

$$\log L_n(\theta; x_1, x_2, \dots, x_n) = -n\theta + \log(\theta)\sum_{i=1}^{n} x_i - \log\left(\prod_{i=1}^{n} x_i!\right)$$

$$\frac{\partial \log L_n(\theta; x_1, x_2, \dots, x_n)}{\partial \theta} = -n + \frac{1}{\theta}\sum_{i=1}^{n} x_i$$

$$\frac{\partial^2 \log L_n(\theta; x_1, x_2, \dots, x_n)}{\partial \theta^2} = -\frac{1}{\theta^2}\sum_{i=1}^{n} x_i < 0$$

Exemplu introductiv (cont.)

Estimatorul de verosimilitate maximă este definit prin

$$\hat{\theta} = \arg\max_{\theta \in \mathbb{R}_+} L_n(\theta; x_1, x_2, \dots, x_n) = \arg\max_{\theta \in \mathbb{R}_+} \log L_n(\theta; x_1, x_2, \dots, x_n)$$

În cazul exemplului nostru

$$\left.\frac{\partial \log L_n(\theta; x_1, x_2, \dots, x_n)}{\partial \theta}\right|_{\hat{\theta}} = -n + \frac{1}{\hat{\theta}} \sum_{i=1}^{n} x_i = 0 \iff \hat{\theta} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

$$\left.\frac{\partial^2 \log L_n(\theta; x_1, x_2, \dots, x_n)}{\partial \theta^2}\right|_{\hat{\theta}} = -\frac{1}{\hat{\theta}^2} \sum_{i=1}^{n} x_i < 0$$

deci $\hat{\theta}$ este maximul.

Exemplu introductiv (cont.)

- Valoarea *estimată* (număr) a parametrului θ obținută prin metoda verosimilității maxime este

$$\hat{\theta} \equiv \hat{\theta}(x) = \frac{1}{n}\sum_{i=1}^{n} x_i$$

care pentru eșantionul $\{5, 0, 1, 1, 0, 3, 2, 3, 4, 1\}$ devine $\hat{\theta}(x) = 2$

- *Estimatorul de verosimilitate maximă* (variabilă aleatoare) este

$$\hat{\theta} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

Exemplu introductiv 2

Example

Să presupunem că vrem să estimăm numărul N de porci mistreți care se află pe domeniul de la Balc prin metoda capturării și recapturării.
Această metodă presupune două etape: în prima fază sunt prinși M porci mistreți care sunt etichetați și apoi eliberați iar în a doua fază n porci mistreți sunt prinși (la întâmplare) și printre aceștia se observă că k sunt etichetați.
Considerăm $M = 25$, $n = 25$ și $k = 7$.

Exemplu introductiv 2 (cont.)

Probabilitatea să observăm k porci mistreți marcați este (repartiția hipergeometrică)

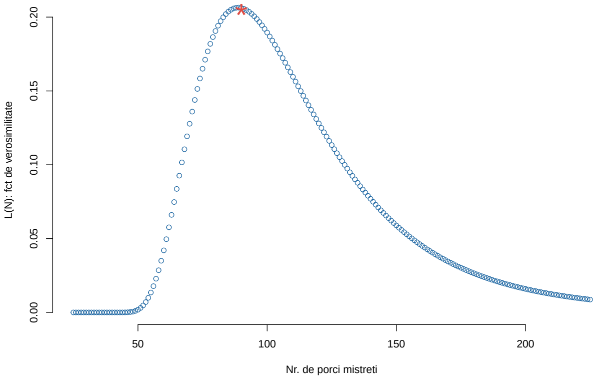
$$\mathbb{P}(X = k \mid N, M, n) = \frac{\binom{N-M}{n-k}\binom{M}{k}}{\binom{N}{n}}$$

iar funcția de verosimilitate pentru datele noastre este

$$L(N) = \frac{\binom{N-25}{25-7}\binom{25}{7}}{\binom{N}{25}}$$

Exemplu introductiv 2 (cont.)

Graficul funcției de verosimilitate este



Exemplu introductiv 2 (cont.)

Pentru a determina analitic acest maxim să observăm că

$$\frac{L(N)}{L(N-1)} = \frac{(N-M)(N-n)}{(N-M-n+k)(N)}$$

și $\frac{L(N)}{L(N-1)} < 1$ dacă $N > \frac{Mn}{k} = 89.285$ ceea ce conduce la $\hat{N} = 89$ (deoarece trebuie să fie o valoare întreagă).

Cazul variabilelor continue

În cazul variabilelor continue (absolut continue)

- probabilitatea să observăm $\{x_1, x_2, \ldots, x_n\}$ este 0
- definim funcția de verosimilitate asociată eșantionului prin densitatea comună a $(X_1, X_2, \ldots, X_n)$ evaluată în punctul $(x_1, x_2, \ldots, x_n)$

$$L_n(\theta; x_1, x_2, \ldots, x_n) = f_\theta(x_1, x_2, \ldots, x_n)$$

Cazul variabilelor continue

- Dacă variabilele aleatoare $X_1, X_2, \dots, X_n$ sunt i.i.d. atunci

$$L_n(\theta; x_1, x_2, \dots, x_n) = f_\theta(x_1, x_2, \dots, x_n) = \prod_{i=1}^{n} f_\theta(x_i)$$

unde $f_\theta(x_i)$ este densitatea marginală a lui X_i (toate sunt egale din i.d.)

- Valorile parametrilor care maximizează funcția de verosimilitate $L_n(\theta; x_1, x_2, \dots, x_n)$ sau logaritmul acesteia se numesc valorile estimate ale estimatorului de verosimilitate maximă și se notează cu $\hat{\theta}(x)$

Funcția de verosimilitate

Funcția de verosimilitate

Obiective

În această secțiune ne propunem să:

- introducem *notațiile* corespunzătoare unei probleme de estimare atât în cazul marginal cât și în cazul condiționat
- definim funcția de verosimilitate și logaritmul funcției de verosimilitate
- introducem conceptul de funcție de verosimilitate condiționată
- exemplificăm prin exemple

Notații

- Fie X o variabilă aleatoare cu densitatea de repartiție, respectiv funcția de masă, $f_\theta(x)$ (legea / repartiția mamă - cea care guvernează populația)
- Presupunem că $\theta = (\theta_1, \ldots, \theta_k)^\intercal$ este un $k \times 1$ vector de parametrii necunoscuți cu $\theta \in \Theta \subset \mathbb{R}^k$, Θ este spațiul parametrilor
- Vom considera $X_1, X_2, \ldots, X_n$ un eșantion (în general i.i.d.) de talie (volum) n din populația f_θ, notăm $X_1, X_2, \ldots, X_n \sim f_\theta$
- Realizarea eșantionului $X_1, X_2, \ldots, X_n$ o notăm cu $x_1, x_2, \ldots, x_n$ (setul de date) sau cu x

Exemplu

Example

Dacă $X \sim \mathcal{N}(\mu, \sigma^2)$ atunci

$$f_\theta(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad x \in \mathbb{R}$$

cu $k = 2$, $\Theta = \mathbb{R} \times \mathbb{R}_+$ și

$$\theta = (\theta_1, \theta_2)^\intercal = \begin{pmatrix} \mu \\ \sigma^2 \end{pmatrix}$$

Funcția de verosimilitate

Definition (Funcția de verosimilitate)

Funcția de verosimilitate asociată eșantionului $X_1, \ldots, X_n$ este definită prin $L_n : \Theta \times \mathbb{R}^n \to \mathbb{R}_+$ cu

$$(\theta; x_1, \ldots, x_n) \mapsto L_n(\theta; x_1, \ldots, x_n) = f_\theta(x_1, \ldots, x_n) \overset{iid}{=} \prod_{i=1}^{n} f_\theta(x_i)$$

Definition (Logaritmul funcției de verosimilitate)

Logaritmul funcției de verosimilitate este definit prin $\ell_n = \log L_n : \Theta \times \mathbb{R}^n \to \mathbb{R}$ cu

$$(\theta; x_1, \ldots, x_n) \mapsto \ell_n(\theta; x_1, \ldots, x_n) = \log f_\theta(x_1, \ldots, x_n) \overset{iid}{=} \sum_{i=1}^{n} \log f_\theta(x_i)$$

Funcția de verosimilitate - notații

Putem folosi următoarele notații alternative în funcție de context

$$L_n(\theta; x) \equiv L_n(\theta; x_1, \ldots, x_n) \equiv L_n(\theta)$$

$$\ell_n(\theta; x) \equiv \ell_n(\theta; x_1, \ldots, x_n) \equiv \ell_n(\theta)$$

Exemplu

Example (Populații normale)

Fie $X_1, X_2, \dots, X_n \sim \mathcal{N}(\mu, \sigma^2)$ cu $\theta = \begin{pmatrix}\mu & \sigma^2\end{pmatrix}^\intercal$. Atunci

$$L_n(\theta; x) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}} = (2\pi\sigma^2)^{-\frac{n}{2}} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2}$$

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

Funcția de verosimilitate

Definition (Funcția de verosimilitate pentru o observație)

În mod analog cu definiția (logaritmului) funcției de verosimilitate pentru eșantion putem defini și corespondentul pentru o observație x_i

$$L_i(\theta; x) = f_\theta(x_i) \quad \text{cu} \quad L_n(\theta; x) = \prod_{i=1}^{n} L_i(\theta; x)$$

$$\ell_i(\theta; x) = \log f_\theta(x_i) \quad \text{cu} \quad \ell_n(\theta; x) = \sum_{i=1}^{n} \ell_i(\theta; x)$$

Remarcă

Putem vorbi de funcția de verosimilitate și respectiv de estimatorul de verosimilitate maximă atunci când știm forma parametrică a populației din care provine eșantionul. În practică, sunt puține cazurile în care știm forma repartiției care a generat observațiile.

Funcția de verosimilitate (model)

Remarcă

Putem folosi metoda verosimilității maxime și în cazul în care vrem să estimăm parametrii unui model (cu variabile dependente și explicative sau covariabile) de forma

$$y = g(x; \theta) + \varepsilon$$

unde θ este un vector de parametrii, X este o mulțime de variabile explicative, ε este termenul eroare iar $g(\cdot)$ este o funcție de legătură (link).

În acest caz, vom considera *repartiția condiționată* a variabilei dependente Y la vectorul de variabile explicative X, care este echivalentă cu repartiția necondiționată a termenului eroare ε:

$$Y|X \sim F \iff \varepsilon \sim F$$

Funcția de verosimilitate (model) - notații

În cazul unui model:

- Fie Y o variabilă aleatoare (denumită variabilă dependentă sau variabilă răspuns) și X un vector aleator (de variabile explicative sau covariabile)
- Presupunem că repartiția condiționată a lui Y la $X = x$ admite densitatea de repartiție $f_\theta(y|x)$
- Presupunem că $\theta = (\theta_1, \ldots, \theta_k)^\intercal$ este un $k \times 1$ vector de parametrii necunoscuți cu $\theta \in \Theta \subset \mathbb{R}^k$, Θ este spațiul parametrilor
- Vom considera eșantionul $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$ i.i.d. și considerăm $(x_i, y_i)_{i=1}^n$ o realizare a sa

Funcția de verosimilitate condiționată

Definition (Funcția și logaritmul funcției de verosimilitate condiționată)

Funcția de verosimilitate (condiționată) asociată eșantionului $(X_i, Y_i)_{i=1}^{n}$ este definită prin

$$L_n(\theta; y|x) = \prod_{i=1}^{n} f_\theta(y_i|x_i).$$

Logaritmul funcției de verosimilitate (condiționată) este definit prin

$$\ell_n(\theta; y|x) = \sum_{i=1}^{n} \log f_\theta(y_i|x_i)$$

unde $f_\theta(y_i|x_i)$ reprezintă densitatea condiționată a lui Y_i la X_i.

Exemplu - regresie

Example (Modelul de regresie liniară)

Considerăm următorul model de regresie:

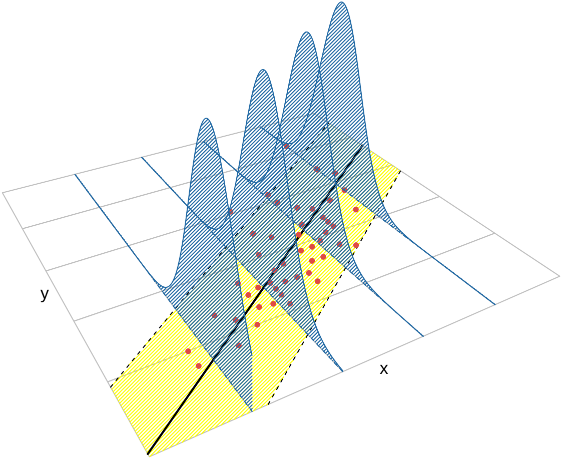
$$y_i = \boldsymbol{X}_i^\intercal \boldsymbol{\beta} + \varepsilon_i$$

unde $\boldsymbol{X}_i$ este un $k \times 1$ vector de variabile aleatoare iar $\boldsymbol{\beta} = \begin{pmatrix} \beta_1 & \cdots & \beta_k \end{pmatrix}^\intercal$ este un $k \times 1$ vector de parametrii. Presupunem că ε_i sunt variabile aleatoare i.i.d. repartizate $\mathcal{N}(0, \sigma^2)$. Atunci, repartiția condiționată a lui Y_i la $\boldsymbol{X}_i = \boldsymbol{x}_i$ este

$$Y_i | \boldsymbol{X}_i = \boldsymbol{x}_i \sim \mathcal{N}(\boldsymbol{x}_i^\intercal \boldsymbol{\beta}, \sigma^2)$$

astfel $f_{\boldsymbol{\theta}}(y_i|\boldsymbol{x}_i) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(y_i - \boldsymbol{x}_i^\intercal \beta)^2}{2\sigma^2}}$ unde vectorul de parametrii $\boldsymbol{\theta} = \begin{pmatrix} \beta & \sigma^2 \end{pmatrix}^\intercal$ este un $(k+1) \times 1$ vector.

Exemplu - regresie



Exemplu - regresie

Example (Modelul de regresie liniară)

Având dat eșantionul i.i.d. $(y_i, \boldsymbol{x}_i)_{i=1}^n$, putem construi funcția de verosimilitate condiționată

$$
\begin{aligned}
L_n(\boldsymbol{\theta}; y|\boldsymbol{x}) &= \prod_{i=1}^{n} f_{\boldsymbol{\theta}}(y_i|\boldsymbol{x}_i) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(y_i - \boldsymbol{x}_i^\intercal \beta)^2}{2\sigma^2}} \\
&= (2\pi\sigma^2)^{-\frac{n}{2}} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \boldsymbol{x}_i^\intercal \beta)^2}
\end{aligned}
$$

$$
\ell_n(\boldsymbol{\theta}; y|\boldsymbol{x}) = -\frac{n}{2}\log(\sigma^2) - \frac{n}{2}\log(2\pi) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \boldsymbol{x}_i^\intercal \beta)^2
$$

Exemplu - Modelele Probit / Logit

Example (Probit/Logit)

Fie Y_i o variabilă aleatoare cu $y_i \in \{0, 1\}$ astfel ca $Y_i = 1$ dacă firma i se află în imposibilitate de plată și 0 în caz contrar. Fie $\boldsymbol{X}_i = \begin{pmatrix} X_{i1} & \cdots & X_{ik} \end{pmatrix}$ vectorul de caracteristici ale firmei i. Presupunem că repartiția condiționată de imposibilitate de plată a firmei i dat fiind caracteristicile $\boldsymbol{X}_i$ are forma

$$\mathbb{P}(Y_i = 1 | \boldsymbol{X}_i = \boldsymbol{x}_i) = F(\boldsymbol{x}_i^{\intercal}\beta)$$

unde $\beta = \begin{pmatrix} \beta_1 & \cdots & \beta_k \end{pmatrix}$ este un vector de parametrii iar $F(\cdot)$ este o funcție de repartiție.

$$Y_i = \begin{cases} 1, & \text{cu probabilitatea } F(\boldsymbol{x}_i^{\intercal}\beta) \\ 0, & \text{cu probabilitatea } 1 - F(\boldsymbol{x}_i^{\intercal}\beta) \end{cases}$$

Exemplu - Modelele Probit / Logit

Definition (Probit)

În modelul **probit**, repartiția condiționată a evenimentului $Y_i = 1$ este

$$\mathbb{P}(Y_i = 1 | \boldsymbol{X}_i = \boldsymbol{x}_i) = \Phi(\boldsymbol{x}_i^\intercal \beta) = \int_{-\infty}^{\boldsymbol{x}_i^\intercal \beta} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}\, dt$$

unde $\Phi(\cdot)$ reprezintă funcția de repartiție a repartiției normale $\mathcal{N}(0, 1)$.

Definition (Logit)

În modelul **logit**, repartiția condiționată a evenimentului $Y_i = 1$ este

$$\mathbb{P}(Y_i = 1 | \boldsymbol{X}_i = \boldsymbol{x}_i) = \Lambda(\boldsymbol{x}_i^\intercal \beta) = \frac{1}{1 + e^{-\boldsymbol{x}_i^\intercal \beta}}$$

unde $\Lambda(\cdot)$ reprezintă funcția de repartiție a repartiției logistice $f(x) = \frac{e^{-x}}{(1+e^{-x})^2}$.

Exemplu - Modelele Probit / Logit

Example (Probit/Logit)

Fie $(\boldsymbol{x}_i, y_i)_{i=1}^n$ un eșantion de talie n. Indiferent de forma funcției $F(\cdot)$, $Y_i|\boldsymbol{X}_i = \boldsymbol{x}_i$ este repartizată Bernoulli $\mathcal{B}\left(F(\boldsymbol{x}_i^\intercal\beta)\right)$. Astfel, pentru $\boldsymbol{\theta} = \beta$, avem

$$L_n(\boldsymbol{\theta}; y|\boldsymbol{x}) = \prod_{i=1}^{n} f_{\boldsymbol{\theta}}(y_i|\boldsymbol{x}_i) = \prod_{i=1}^{n} \left[F(\boldsymbol{x}_i^\intercal\beta)\right]^{y_i} \left[1 - F(\boldsymbol{x}_i^\intercal\beta)\right]^{1-y_i}$$

$$\ell_n(\boldsymbol{\theta}; y|\boldsymbol{x}) = \sum_{i=1}^{n} y_i \log\left(F(\boldsymbol{x}_i^\intercal\beta)\right) + \sum_{i=1}^{n}(1 - y_i)\log\left(1 - F(\boldsymbol{x}_i^\intercal\beta)\right)$$

unde $f_{\boldsymbol{\theta}}(y_i|\boldsymbol{x}_i)$ este funcția de masă condiționată a lui Y_i.

Estimatorul de verosimilitate maximă

Estimatorul de verosimilitate maximă

Obiective

În această secțiune ne propunem să:

- definim *estimatorul de verosimilitate maximă*
- să obținem estimări ale parametrului θ
- să introducem *principiul de invarianță*
- exemplificăm prin exemple

Înainte de a începe studiul vom considera problema *identifiabilității*.

Indentifiabilitate

Definition (Identifiabilitate)

Spunem că vectorul de parametrii $\theta \in \Theta$ este identifiabil (estimabil) dat fiind eșantionul $x = (x_1, \dots, x_n)$ dacă pentru orice alt vector de parametrii $\theta^* \in \Theta$, $\theta^* \neq \theta$ avem

$$L_n(\theta; x) \neq L_n(\theta^*; x).$$

Exemplu - identifiabilitate

Example (Model latent)

Considerăm variabila latentă (continuă și neobservabilă) Y_i^* astfel ca

$$Y_i^* = \boldsymbol{X}_i^\intercal \beta + \varepsilon_i$$

cu $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_k)$, $\boldsymbol{X}_i = (X_{i1}, \ldots, X_{ik})^\intercal$ iar ε_i sunt variabile aleatoare i.i.d. cu $\mathbb{E}[\varepsilon_i] = 0$ și $Var(\varepsilon_i) = \sigma^2$. Presupunem că repartiția lui ε_i este simetrică în jurul lui 0. Fie $G(\cdot)$ funcția de repartiție a termenului standardizat $\frac{\varepsilon_i}{\sigma}$ și presupunem că aceasta nu depinde de σ sau $\boldsymbol{\beta}$ (e.g. $\mathcal{N}(0,1)$).

Observăm variabila Y_i definită prin

$$Y_i = \begin{cases} 1, & Y_i^* > 0 \\ 0, & \text{altfel} \end{cases}$$

Este vectorul de parametrii $\boldsymbol{\theta} = \begin{pmatrix} \boldsymbol{\beta}^\intercal & \sigma^2 \end{pmatrix}^\intercal$ identifiabil ?

Exemplu - identifiabilitate (cont.)

Pentru a verifica problema identifiabilității vom calcula funcția de verosimilitate pentru eșantionul (datele observabile) $\{y_i, \boldsymbol{x}_i\}_{i=1}^n$. Avem

$$\begin{aligned}
\mathbb{P}(Y_i = 1 | \boldsymbol{X}_i = \boldsymbol{x}_i) &= \mathbb{P}(Y_i^* > 0 | \boldsymbol{X}_i = \boldsymbol{x}_i) \\
&= \mathbb{P}(\varepsilon_i > -\boldsymbol{x}_i^\intercal \boldsymbol{\beta}) \\
&= 1 - \mathbb{P}(\varepsilon_i \leq -\boldsymbol{x}_i^\intercal \boldsymbol{\beta}) \\
&= 1 - \mathbb{P}\left(\frac{\varepsilon_i}{\sigma} \leq -\boldsymbol{x}_i^\intercal \frac{\boldsymbol{\beta}}{\sigma}\right) \\
&= 1 - G\left(-\boldsymbol{x}_i^\intercal \frac{\boldsymbol{\beta}}{\sigma}\right) \overset{simetrie}{=} G\left(\boldsymbol{x}_i^\intercal \frac{\boldsymbol{\beta}}{\sigma}\right)
\end{aligned}$$

Exemplu - identifiabilitate (cont.)

Pentru $\boldsymbol{\theta} = \begin{pmatrix} \beta^{\intercal} & \sigma^2 \end{pmatrix}^{\intercal}$ funcția de verosimilitate (log) este

$$\ell_n(\boldsymbol{\theta}; y|\boldsymbol{x}) = \sum_{i=1}^{n} y_i \log\left(G\left(\boldsymbol{x}_i^{\intercal}\frac{\beta}{\sigma}\right)\right) + \sum_{i=1}^{n}(1-y_i)\log\left(1 - G\left(\boldsymbol{x}_i^{\intercal}\frac{\beta}{\sigma}\right)\right)$$

și depinde doar de raportul $\frac{\beta}{\sigma}$.

Astfel observăm că pentru $\boldsymbol{\theta} = \begin{pmatrix} \beta^{\intercal} & \sigma^2 \end{pmatrix}^{\intercal}$ și $\boldsymbol{\theta}^* = \begin{pmatrix} u \times \beta^{\intercal} & u \times \sigma^2 \end{pmatrix}^{\intercal}$, $u \neq 1$ avem

$$\ell_n(\boldsymbol{\theta}; y|\boldsymbol{x}) = \ell_n(\boldsymbol{\theta}^*; y|\boldsymbol{x})$$

prin urmare parametrii β și σ^2 **nu pot fi indetificați**, putem estima doar raportul $\frac{\beta}{\sigma}$.

Exemplu - identifiabilitate (cont.)

Remarcă:

În cazul modelului latent considerat, doar raportul $\frac{\beta}{\sigma}$ a putut fi estimat deoarece

$$\mathbb{P}(Y_i = 1|\boldsymbol{X}_i = \boldsymbol{x}_i) = \mathbb{P}\left(\frac{\varepsilon_i}{\sigma} \leq \boldsymbol{x}_i^\intercal \frac{\boldsymbol{\beta}}{\sigma}\right) = G\left(\boldsymbol{x}_i^\intercal \frac{\boldsymbol{\beta}}{\sigma}\right)$$

În prima etapă a modelului logit/probit are loc o normalizare:

- logit/probit: $\mathbb{P}(Y_i = 1|\boldsymbol{X}_i = \boldsymbol{x}_i) = F(\boldsymbol{x}_i^\intercal \tilde{\beta})$ cu $\tilde{\beta} = \frac{\beta}{\sigma}$ iar $Var\left(\frac{\varepsilon_i}{\sigma}\right) = 1$

MLE - JUSTIFICARE

Fie $X_1, \ldots, X_n \sim f_\theta$ un eșantion de talie n dintr-o populație f_θ care verifică condițiile (de regularitate):

- (R0) Identifiabilitate: $\theta \neq \theta^* \Rightarrow f_\theta(x) \neq f_{\theta^*}(x)$
- (R1) $\mathrm{supp} f_\theta = \mathrm{supp} f_{\theta^*}$, $\forall \theta \neq \theta^*$, unde $\mathrm{supp} f_\theta = \overline{\{x \mid f_\theta(x) \neq 0\}}$

Lemma

Fie θ_0 parametrul real care a generat eșantionul. Sub condițiile de regularitate (R0) și (R1) avem

$$\lim_{n\to\infty} \mathbb{P}_{\theta_0}\left(L_n(\theta_0; \boldsymbol{X}) > L_n(\theta; \boldsymbol{X})\right) = 1, \quad \textit{pentru orice } \theta \neq \theta_0$$

- Rezultatul lemei ne spune că funcția de verosimilitate separă (asimptotic) adevăratul model în θ_0 de celelalte modele în $\theta \neq \theta_0$

MLE - justificare (cont.)

Demonstrație: Considerăm, pentru $X(\omega) = (X_1(\omega), \dots, X_n(\omega))$, evenimentul

$$A_n = \{\omega \in \Omega | L_n(\theta_0; \boldsymbol{X}(\omega)) > L_n(\theta; \boldsymbol{X}(\omega))\}$$

și avem echivalența

$$\begin{aligned}
\omega \in A_n \iff & \omega \in \left\{\omega \Big| \frac{L_n(\theta_0; \boldsymbol{X}(\omega))}{L_n(\theta; \boldsymbol{X}(\omega))} > 1\right\} \\
\iff & \omega \in \left\{\omega \Big| \log \frac{L_n(\theta_0; \boldsymbol{X}(\omega))}{L_n(\theta; \boldsymbol{X}(\omega))} > 0\right\} \\
\iff & \omega \in \left\{\omega \Big| \frac{1}{n} \sum_{i=1}^{n} \log \frac{f_{\theta_0}(X_i(\omega))}{f_{\theta}(X_i(\omega))} > 0\right\}
\end{aligned}$$

Notând cu $K_n(\theta_0, \theta) = \frac{1}{n} \sum_{i=1}^{n} \log \frac{f_{\theta_0}(X_i)}{f_{\theta}(X_i)}$ am văzut că

$$\omega \in A_n \iff \omega \in \{\omega | K_n(\theta_0, \theta) > 0\}$$

MLE - justificare (cont.)

Fie $K(\theta_0, \theta) = \mathbb{E}_{\theta_0}\left[\log \frac{f_{\theta_0}(X)}{f_\theta(X)}\right]$ (divergența Kullback-Leibler).

Aplicând inegalitatea lui Jensen pentru funcția convexă $-\log(x)$ avem

$$\mathbb{E}_{\theta_0}\left[-\log \frac{f_\theta(X)}{f_{\theta_0}(X)}\right] \geq -\log \mathbb{E}_{\theta_0}\left[\frac{f_\theta(X)}{f_{\theta_0}(X)}\right] = -\log \int \frac{f_\theta(x)}{f_{\theta_0}(x)} f_{\theta_0}(x)\, dx = 0$$

prin urmare $K(\theta_0, \theta) \geq 0$ cu egalitate dacă $\theta = \theta_0$ (vezi R0).

Din Legea Numerelor Mari avem: $K_n(\theta_0, \theta) = \frac{1}{n}\sum_{i=1}^{n} \log \frac{f_{\theta_0}(X_i)}{f_\theta(X_i)} \xrightarrow{\mathbb{P}_{\theta_0}} K(\theta_0, \theta) > 0$ astfel pentru $\varepsilon > 0$

$$\mathbb{P}_{\theta_0}\left(|K_n(\theta_0, \theta) - K(\theta_0, \theta)| \leq \varepsilon\right) \to 1$$

Alegând $\varepsilon > 0$ astfel ca $K(\theta_0, \theta) - \varepsilon > 0$ găsim

$$\{|K_n(\theta_0, \theta) - K(\theta_0, \theta)| \leq \varepsilon\} \subset \{K_n(\theta_0, \theta) \geq K(\theta_0, \theta) - \varepsilon\} \subset \{K_n(\theta_0, \theta) > 0\}$$

deci $\mathbb{P}_{\theta_0}\left(K_n(\theta_0, \theta) > 0\right) \to 1$. $\square$

MLE - definiție

Definition (Estimatorul de verosimilitate maximă)

Estimatorul de verosimilitate maximă $\hat{\theta}$ a lui $\theta \in \Theta$ este o soluție a problemei de maximizare

$$\hat{\theta} = \begin{cases} \arg\max_{\theta \in \Theta} L_n(\theta; x), & \text{cazul necondiționat} \\ \arg\max_{\theta \in \Theta} L_n(\theta; y|x), & \text{cazul condiționat} \end{cases}$$

sau echivalent

$$\hat{\theta} = \begin{cases} \arg\max_{\theta \in \Theta} \ell_n(\theta; x), & \text{cazul necondiționat} \\ \arg\max_{\theta \in \Theta} \ell_n(\theta; y|x), & \text{cazul condiționat} \end{cases}$$

MLE - ecuațiile de verosimilitate

Definition (Ecuațiile de verosimilitate)

În anumite condiții de regularitate, estimatorul de verosimilitate maximă a lui θ este definit ca fiind soluția sistemului (condiții de ordin unu)

$$\begin{cases} \left.\frac{\partial L_n(\theta;x)}{\partial\theta}\right|_{\hat{\theta}} = \underset{(k,1)}{0}, & \text{cazul necondiționat} \\ \left.\frac{\partial L_n(\theta;y|x)}{\partial\theta}\right|_{\hat{\theta}} = \underset{(k,1)}{0}, & \text{cazul condiționat} \end{cases}$$

sau echivalent

$$\begin{cases} \left.\frac{\partial \ell_n(\theta;x)}{\partial\theta}\right|_{\hat{\theta}} = \underset{(k,1)}{0}, & \text{cazul necondiționat} \\ \left.\frac{\partial \ell_n(\theta;y|x)}{\partial\theta}\right|_{\hat{\theta}} = \underset{(k,1)}{0}, & \text{cazul condiționat} \end{cases}$$

MLE - ecuațiile de verosimilitate (cont.)

În ceea ce privește notațiile avem că derivata întâi (gradientul) (logaritmului) funcției de verosimilitate (condiționată) evaluată în punctul $\hat{\theta}$ satisface

$$\left.\frac{\partial L_n(\theta; y|x)}{\partial \theta}\right|_{\hat{\theta}} = \frac{\partial L_n(\hat{\theta}; y|x)}{\partial \theta} = \nabla L_n(\hat{\theta}; y|x) = 0$$

- ecuațiile de verosimilitate corespund la un sistem liniar/neliniar de k ecuații cu k necunoscute $\theta_1, \ldots, \theta_k$

$$\nabla L_n(\theta; x)|_{\hat{\theta}} = \left.\frac{\partial L_n(\theta; x)}{\partial \theta}\right|_{\hat{\theta}} = \begin{pmatrix} \left.\frac{\partial L_n(\theta; x)}{\partial \theta_1}\right|_{\hat{\theta}} \\ \cdots \\ \left.\frac{\partial L_n(\theta; x)}{\partial \theta_k}\right|_{\hat{\theta}} \end{pmatrix} = \begin{pmatrix} 0 \\ \cdots \\ 0 \end{pmatrix}.$$

MLE - condiții de ordin doi

Definition (Condițiile de ordin doi)

În anumite condiții de regularitate, soluțiile ecuațiilor de verosimilitate ne dau punctele critice (din interiorul domeniului) iar pentru determinarea maximului este necesară verificarea condițiilor de ordin doi: matricea Hessiană evaluată în $\hat{\theta}$ trebuie să fie negativ definită

$$\begin{cases} \left.\frac{\partial^2 L_n(\theta;x)}{\partial\theta\partial\theta^\intercal}\right|_{\hat{\theta}} \text{ este negativ definită,} & \text{cazul necondiționat} \\ \left.\frac{\partial^2 L_n(\theta;y|x)}{\partial\theta\partial\theta^\intercal}\right|_{\hat{\theta}} \text{ este negativ definită,} & \text{cazul condiționat} \end{cases}$$

sau echivalent

$$\begin{cases} \left.\frac{\partial^2 \ell_n(\theta;x)}{\partial\theta\partial\theta^\intercal}\right|_{\hat{\theta}} \text{ este negativ definită,} & \text{cazul necondiționat} \\ \left.\frac{\partial^2 \ell_n(\theta;y|x)}{\partial\theta\partial\theta^\intercal}\right|_{\hat{\theta}} \text{ este negativ definită,} & \text{cazul condiționat} \end{cases}$$

MLE - condiții de ordin doi (cont.)

Reamintim că matricea Hessiană este

$$\frac{\partial^2 L_n(\theta; x)}{\partial\theta\partial\theta^\intercal} = \begin{pmatrix} \frac{\partial^2 L_n(\theta;x)}{\partial\theta_1^2} & \frac{\partial^2 L_n(\theta;x)}{\partial\theta_1\partial\theta_2} & \dots & \frac{\partial^2 L_n(\theta;x)}{\partial\theta_1\partial\theta_k} \\ \frac{\partial^2 L_n(\theta;x)}{\partial\theta_2\partial\theta_1} & \frac{\partial^2 L_n(\theta;x)}{\partial\theta_2^2} & \dots & \frac{\partial^2 L_n(\theta;x)}{\partial\theta_2\partial\theta_k} \\ \dots & \dots & \dots & \dots \\ \frac{\partial^2 L_n(\theta;x)}{\partial\theta_k\partial\theta_1} & \frac{\partial^2 L_n(\theta;x)}{\partial\theta_k\partial\theta_2} & \dots & \frac{\partial^2 L_n(\theta;x)}{\partial\theta_k^2} \end{pmatrix}$$

- O matrice simetrică A este negativ definită dacă toate valorile proprii sunt negative
- Alternativ, o matrice simetrică A este negativ definită dacă $x^\intercal A x < 0$ pentru orice $x \in \mathbb{R}^n \setminus \{0\}$

MLE - exemplul 1

Example (MLE pentru cazul univariat)

Fie $X_1, \ldots, X_n \sim f_\theta$ un eșantion de talie n dintr-o populație f_θ cu

$$f_\theta(x) = \frac{x}{\theta} e^{-\frac{x^2}{2\theta}}, \quad x \geq 0$$

Determinați estimatorul de verosimilitate maximă a lui θ.

MLE - exemplul 1 (cont.)

Soluție: Avem că $\log f_\theta(x) = -\frac{x^2}{2\theta} + \log(x) - \log(\theta)$

prin urmare logaritmul funcției de verosimilitate este

$$\ell_n(\theta; x) = \sum_{i=1}^{n} \log f_\theta(x_i) = -\frac{1}{2\theta}\sum_{i=1}^{n} x_i^2 + \sum_{i=1}^{n} \log(x_i) - n\log(\theta)$$

Estimatorul de verosimilitate maximă este o soluție a ecuației de verosimilitate

$$\frac{\partial \ell_n(\theta; x)}{\partial \theta} = \frac{1}{2\theta^2}\sum_{i=1}^{n} x_i^2 - \frac{n}{\theta} = 0$$

astfel $\left.\frac{\partial \ell_n(\theta;x)}{\partial \theta}\right|_{\hat{\theta}} = 0 \iff \hat{\theta} = \frac{1}{2n}\sum_{i=1}^{n} x_i^2$.

MLE - exemplul 1 (cont.)

Pentru a verifica că $\hat{\theta}$ este maximul avem

$$\frac{\partial^2 \ell_n(\theta; x)}{\partial \theta^2} = \frac{\partial}{\partial \theta}\left(\frac{1}{2\theta^2}\sum_{i=1}^{n} x_i^2 - \frac{n}{\theta}\right) = -\frac{1}{\theta^3}\sum_{i=1}^{n} x_i^2 + \frac{n}{\theta^2}$$

prin urmare condiția de ordin doi devine

$$\begin{aligned} \left.\frac{\partial^2 \ell_n(\theta; x)}{\partial \theta^2}\right|_{\hat{\theta}} &= -\frac{1}{\hat{\theta}^3}\sum_{i=1}^{n} x_i^2 + \frac{n}{\hat{\theta}^2} \\ &= -\frac{2n\hat{\theta}}{\hat{\theta}^3} + \frac{n}{\hat{\theta}^2} \quad \text{deoarece} \quad \hat{\theta} = \frac{1}{2n}\sum_{i=1}^{n} x_i^2 \\ &= -\frac{n}{\hat{\theta}^2} < 0 \end{aligned}$$

- Estimatorul de verosimilitate maximă este $\hat{\theta} = \frac{1}{2n}\sum_{i=1}^{n} X_i^2$

MLE - exemplul 2

Example (Populații normale)

Fie $X_1, \ldots, X_n$ un eșantion de talie n dintr-o populație $\mathcal{N}(\mu, \sigma^2)$. Care sunt estimatorii de verosimilitate maximă pentru μ și σ^2 ?

Soluție: Fie $\theta = (\mu, \sigma^2)^\intercal$

$$\hat{\theta} = \underset{\mu\in\mathbb{R},\sigma^2\in\mathbb{R}_+}{\arg\max}\ L_n(\theta; x) = \underset{\mu\in\mathbb{R},\sigma^2\in\mathbb{R}_+}{\arg\max}\ \ell_n(\theta; x)$$

- logaritmul funcției de verosimilitate este

$$\ell_n(\theta; x) = \log \prod_{i=1}^{n} f_\theta(x_i) = -\frac{n}{2}\log(\sigma^2) - \frac{n}{2}\log(2\pi) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2$$

MLE - exemplul 2 (cont.)

Condițiile de ordin unu conduc la

$$\nabla \ell_n(\theta; x)\big|_{\hat{\theta}} = \begin{pmatrix} \left.\frac{\partial \ell_n(\theta;x)}{\partial \mu}\right|_{\hat{\theta}} \\ \left.\frac{\partial \ell_n(\theta;x)}{\partial \sigma^2}\right|_{\hat{\theta}} \end{pmatrix} = \begin{pmatrix} \frac{1}{\hat{\sigma}^2}\sum_{i=1}^{n}(x_i - \hat{\mu})^2 \\ -\frac{n}{2\hat{\sigma}^2} + \frac{1}{2\hat{\sigma}^4}\sum_{i=1}^{n}(x_i - \hat{\mu})^2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

ceea ce arată că estimatorul de verosimilitate maximă este

$$\hat{\theta} = \begin{pmatrix} \hat{\mu} \\ \hat{\sigma}^2 \end{pmatrix} = \begin{pmatrix} \frac{1}{n}\sum_{i=1}^{n} X_i \\ \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2 \end{pmatrix}$$

MLE - exemplul 2 (cont.)

Matricea Hessiană este

$$\frac{\partial^2 \ell_n(\theta; x)}{\partial\theta\partial\theta^\intercal} = \begin{pmatrix} \frac{\partial^2 \ell_n(\theta;x)}{\partial\mu^2} & \frac{\partial^2 \ell_n(\theta;x)}{\partial\mu\partial\sigma^2} \\ \frac{\partial^2 \ell_n(\theta;x)}{\partial\sigma^2\partial\mu} & \frac{\partial^2 \ell_n(\theta;x)}{\partial\sigma^4} \end{pmatrix}$$

$$= \begin{pmatrix} -\frac{n}{\sigma^2} & -\frac{1}{\sigma^4}\sum_{i=1}^n (x_i - \mu) \\ -\frac{1}{\sigma^4}\sum_{i=1}^n (x_i - \mu) & \frac{n}{2\sigma^4} - \frac{1}{\sigma^6}\sum_{i=1}^n (x_i - \mu)^2 \end{pmatrix}$$

Condițiile de ordin doi conduc la

$$\left.\frac{\partial^2 \ell_n(\theta; x)}{\partial\theta\partial\theta^\intercal}\right|_{\hat\theta} = \begin{pmatrix} -\frac{n}{\hat\sigma^2} & 0 \\ 0 & \frac{n}{2\hat\sigma^4} - \frac{n\hat\sigma^2}{\hat\sigma^6} \end{pmatrix} = \begin{pmatrix} -\frac{n}{\hat\sigma^2} & 0 \\ 0 & -\frac{n}{2\hat\sigma^4} \end{pmatrix}$$

care este o matrice negativ definită

- am folosit că $n\hat\mu = \sum_{i=1}^n x_i$ și $n\hat\sigma^2 = \sum_{i=1}^n (x_i - \hat\mu)^2$

MLE - exemplul 3

Example (Modelul de regresie liniară)

Considerăm modelul de regresie liniară:

$$y_i = \boldsymbol{x}_i^\intercal \boldsymbol{\beta} + \varepsilon_i$$

unde $\boldsymbol{x}_i = \begin{pmatrix} x_{i1} & \cdots & x_{ik} \end{pmatrix}^\intercal$ și $\boldsymbol{\beta} = \begin{pmatrix} \beta_1 & \cdots & \beta_k \end{pmatrix}^\intercal$ sunt $k \times 1$ vectori. Presupunem că ε_i sunt variabile aleatoare i.i.d. repartizate $\mathcal{N}(0, \sigma^2)$. Atunci, logaritmul funcției de verosimilitate (condiționat) corespunzător eșantionului $(\boldsymbol{x}_i, y_i)$ este

$$\ell_n(\boldsymbol{\theta}; y|\boldsymbol{x}) = -\frac{n}{2}\log(\sigma^2) - \frac{n}{2}\log(2\pi) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \boldsymbol{x}_i^\intercal \beta)^2$$

unde $\theta = \begin{pmatrix} \beta^\intercal & \sigma^2 \end{pmatrix}^\intercal$ este $(k+1) \times 1$ vector. Care sunt estimatorii de verosimilitate maximă pentru β și σ^2 ?

MLE - exemplul 3 (cont.)

Reamintim: dacă $f : \mathbb{R}^k \to \mathbb{R}$ atunci

$$\frac{\partial f}{\partial \boldsymbol{x}^\intercal} = \begin{pmatrix} \frac{\partial f}{\partial x_1} & \frac{\partial f}{\partial x_2} & \cdots & \frac{\partial f}{\partial x_k} \end{pmatrix} \quad \text{și} \quad \frac{\partial f}{\partial \boldsymbol{x}} = \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \cdots \\ \frac{\partial f}{\partial x_k} \end{pmatrix}.$$

În particular, pentru $f(\boldsymbol{x}) = \boldsymbol{a}^\intercal \boldsymbol{x}$ (formă liniară) avem

$$\frac{\partial \boldsymbol{a}^\intercal \boldsymbol{x}}{\partial \boldsymbol{x}^\intercal} = \begin{pmatrix} \frac{\partial \sum_{i=1}^k a_i x_i}{\partial x_1} & \frac{\partial \sum_{i=1}^k a_i x_i}{\partial x_2} & \cdots & \frac{\partial \sum_{i=1}^k a_i x_i}{\partial x_k} \end{pmatrix} = \boldsymbol{a}^\intercal \quad \text{și} \quad \frac{\partial \boldsymbol{a}^\intercal \boldsymbol{x}}{\partial \boldsymbol{x}} = \boldsymbol{a}.$$

MLE - exemplul 3 (cont.)

În cazul în care f este o aplicație liniară, $f(\boldsymbol{x}) = A\boldsymbol{x}$ unde $A \in \mathcal{M}_{m,k}(\mathbb{R})$, atunci

$$A\boldsymbol{x} = \begin{pmatrix} \sum_{j=1}^{k} a_{1j}x_j \\ \sum_{j=1}^{k} a_{2j}x_j \\ \vdots \\ \sum_{j=1}^{k} a_{mj}x_j \end{pmatrix}, \qquad \frac{\partial A\boldsymbol{x}}{\partial x_j} = \begin{pmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{mj} \end{pmatrix}$$

$$\frac{\partial A\boldsymbol{x}}{\partial \boldsymbol{x}^\intercal} = \left(\begin{pmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{pmatrix}, \dots, \begin{pmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{mj} \end{pmatrix}, \dots, \begin{pmatrix} a_{1k} \\ a_{2k} \\ \vdots \\ a_{mk} \end{pmatrix} \right) = \begin{pmatrix} a_{11} & \cdots & a_{1j} & \cdots & a_{1k} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ a_{i1} & \cdots & a_{ij} & \cdots & a_{ik} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mj} & \cdots & a_{mk} \end{pmatrix} = A$$

și $\dfrac{\partial \boldsymbol{x}^\intercal A^\intercal}{\partial \boldsymbol{x}} = A^\intercal$.

MLE - exemplul 3 (cont.)

Dacă f este o formă pătratică $f(\boldsymbol{x}) = \boldsymbol{x}^\intercal A\boldsymbol{x}$ cu $A \in \mathcal{M}_k(\mathbb{R})$ (matrice pătratică de ordin k)

$$\frac{\partial \boldsymbol{x}^\intercal A\boldsymbol{x}}{\partial \boldsymbol{x}} = \begin{pmatrix} \frac{\partial \boldsymbol{x}^\intercal A\boldsymbol{x}}{\partial x_1} \\ \vdots \\ \frac{\partial \boldsymbol{x}^\intercal A\boldsymbol{x}}{\partial x_k} \end{pmatrix} = \begin{pmatrix} \displaystyle\sum_{j=1}^{k} a_{1j}x_j + \sum_{i=1}^{k} a_{i1}x_i \\ \vdots \\ \displaystyle\sum_{j=1}^{k} a_{kj}x_j + \sum_{i=1}^{k} a_{ik}x_i \end{pmatrix} = A\boldsymbol{x} + A^\intercal\boldsymbol{x}, \quad \frac{\partial \boldsymbol{x}^\intercal A\boldsymbol{x}}{\partial \boldsymbol{x}^\intercal} = \left(A\boldsymbol{x} + A^\intercal\boldsymbol{x}\right)^\intercal = \boldsymbol{x}^\intercal A^\intercal + \boldsymbol{x}^\intercal A.$$

În plus, dacă $A \in \mathcal{M}_k(\mathbb{R})$ este o matrice simetrică ($A^\intercal = A$) atunci

$$\frac{\partial \boldsymbol{x}^\intercal A\boldsymbol{x}}{\partial \boldsymbol{x}} = 2A\boldsymbol{x}, \quad \frac{\partial \boldsymbol{x}^\intercal A\boldsymbol{x}}{\partial \boldsymbol{x}^\intercal} = 2\boldsymbol{x}^\intercal A^\intercal.$$

MLE - exemplul 3 (sol - cont.)

Soluție: Avem estimatorul de verosimilitate maximă

$$\hat{\theta} = \underset{\beta\in\mathbb{R}^k, \sigma^2\in\mathbb{R}_+}{\arg\max} \; -\frac{n}{2}\log(\sigma^2) - \frac{n}{2}\log(2\pi) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \boldsymbol{x}_i^\intercal\beta)^2$$

Prin urmare

$$\underbrace{\frac{\partial \ell_n(\theta; y|x)}{\partial \theta}}_{(k+1)\times 1} = \begin{pmatrix} \frac{\partial \ell_n(\theta;y|x)}{\partial \beta} \\ \frac{\partial \ell_n(\theta;y|x)}{\partial \sigma^2} \end{pmatrix} = \begin{pmatrix} \frac{\partial \ell_n(\theta;y|x)}{\partial \beta_1} \\ \cdots \\ \frac{\partial \ell_n(\theta;y|x)}{\partial \beta_k} \\ \frac{\partial \ell_n(\theta;y|x)}{\partial \sigma^2} \end{pmatrix}$$

MLE - exemplul 3 (sol - cont.)

Avem

$$\underbrace{\frac{\partial \ell_n(\theta; y|x)}{\partial \beta}}_{k\times 1} = \frac{1}{\sigma^2}\sum_{i=1}^{n} \underbrace{\boldsymbol{x}_i}_{k\times 1} \underbrace{(y_i - \boldsymbol{x}_i^\intercal \beta)}_{1\times 1}$$

și

$$\underbrace{\frac{\partial \ell_n(\theta; y|x)}{\partial \sigma^2}}_{1\times 1} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n} \underbrace{(y_i - \boldsymbol{x}_i^\intercal \beta)^2}_{1\times 1}$$

MLE - exemplul 3 (sol - cont.)

Ecuația de verosimilitate devine

$$\frac{\partial \ell_n(\theta; y|x)}{\partial \theta}\bigg|_{\hat{\theta}} = \begin{pmatrix} \frac{1}{\hat{\sigma}^2}\sum_{i=1}^n \boldsymbol{x}_i(y_i - \boldsymbol{x}_i^\intercal \hat{\beta}) \\ -\frac{n}{2\hat{\sigma}^2} + \frac{1}{2\hat{\sigma}^4}\sum_{i=1}^n (y_i - \boldsymbol{x}_i^\intercal \hat{\beta})^2 \end{pmatrix} = \begin{pmatrix} 0_k \\ 0 \end{pmatrix}$$

ceea ce conduce la soluția $\hat{\theta} = \begin{pmatrix} \hat{\beta}^\intercal & \hat{\sigma}^2 \end{pmatrix}^\intercal$

$$\hat{\beta} = \left(\sum_{i=1}^n \boldsymbol{X}_i \boldsymbol{X}_i^\intercal\right)^{-1} \left(\sum_{i=1}^n \boldsymbol{X}_i Y_i\right) \quad \text{și} \quad \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^n \left(Y_i - \boldsymbol{X}_i^\intercal \hat{\beta}\right)^2$$

MLE - exemplul 3 (sol - cont.)

Matricea Hessiană este

$$\underbrace{\frac{\partial^2 \ell_n(\theta; y|x)}{\partial\theta\partial\theta^\intercal}}_{(k+1)\times(k+1)} = \begin{pmatrix} \underbrace{\frac{\partial^2 \ell_n(\theta; y|x)}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^\intercal}}_{k\times k} & \underbrace{\frac{\partial^2 \ell_n(\theta; y|x)}{\partial\boldsymbol{\beta}\partial\sigma^2}}_{k\times 1} \\ \underbrace{\frac{\partial^2 \ell_n(\theta; y|x)}{\partial\sigma^2\partial\boldsymbol{\beta}^\intercal}}_{1\times k} & \underbrace{\frac{\partial^2 \ell_n(\theta; y|x)}{\partial\sigma^4}}_{1\times 1} \end{pmatrix}$$

MLE - exemplul 3 (sol - cont.)

Cum

$$\frac{\partial \ell_n(\theta; y|x)}{\partial \beta} = \frac{1}{\sigma^2}\sum_{i=1}^{n} \mathbf{x}_i(y_i - \mathbf{x}_i^\intercal \beta)$$

$$\frac{\partial \ell_n(\theta; y|x)}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n}(y_i - \mathbf{x}_i^\intercal \beta)^2$$

avem matricea Hessiană

$$\frac{\partial^2 \ell_n(\theta; y|x)}{\partial \theta \partial \theta^\intercal} = \begin{pmatrix} -\frac{1}{\sigma^2}\sum_{i=1}^{n} \underbrace{\mathbf{x}_i}_{k\times 1} \underbrace{\mathbf{x}_i^\intercal}_{1\times k} & -\frac{1}{\sigma^4}\sum_{i=1}^{n} \underbrace{\mathbf{x}_i}_{k\times 1} \underbrace{(y_i - \mathbf{x}_i^\intercal \beta)}_{1\times 1} \\ -\frac{1}{\sigma^4}\sum_{i=1}^{n} \underbrace{\mathbf{x}_i^\intercal}_{1\times k} \underbrace{(y_i - \mathbf{x}_i^\intercal \beta)}_{1\times 1} & \frac{n}{2\sigma^4} - \frac{1}{\sigma^6}\sum_{i=1}^{n} \underbrace{(y_i - \mathbf{x}_i^\intercal \beta)^2}_{1\times 1} \end{pmatrix}$$

MLE - exemplul 3 (sol - cont.)

Condițiile de ordin doi se scriu

$$\left.\frac{\partial^2 \ell_n(\theta; y|x)}{\partial\theta\partial\theta^\intercal}\right|_{\hat{\theta}} = \begin{pmatrix} -\frac{1}{\hat{\sigma}^2}\sum_{i=1}^{n}\boldsymbol{x}_i\boldsymbol{x}_i^\intercal & -\frac{1}{\hat{\sigma}^4}\sum_{i=1}^{n}\boldsymbol{x}_i(y_i - \boldsymbol{x}_i^\intercal\hat{\beta}) \\ -\frac{1}{\hat{\sigma}^4}\sum_{i=1}^{n}\boldsymbol{x}_i^\intercal(y_i - \boldsymbol{x}_i^\intercal\hat{\beta}) & \frac{n}{2\hat{\sigma}^4} - \frac{1}{\hat{\sigma}^6}\sum_{i=1}^{n}(y_i - \boldsymbol{x}_i^\intercal\hat{\beta})^2 \end{pmatrix}$$

Și cum $\sum_{i=1}^{n}\boldsymbol{x}_i(y_i - \boldsymbol{x}_i^\intercal\hat{\beta}) = 0$ și respectiv $n\hat{\sigma}^2 = \sum_{i=1}^{n}(y_i - \boldsymbol{x}_i^\intercal\hat{\beta})^2$ avem

$$\left.\frac{\partial^2 \ell_n(\theta; y|x)}{\partial\theta\partial\theta^\intercal}\right|_{\hat{\theta}} = \begin{pmatrix} -\frac{1}{\hat{\sigma}^2}\sum_{i=1}^{n}\boldsymbol{x}_i\boldsymbol{x}_i^\intercal & 0 \\ 0 & \frac{n}{2\hat{\sigma}^4} - \frac{n\hat{\sigma}^2}{\hat{\sigma}^6} \end{pmatrix} = \begin{pmatrix} -\frac{1}{\hat{\sigma}^2}\sum_{i=1}^{n}\boldsymbol{x}_i\boldsymbol{x}_i^\intercal & 0 \\ 0 & -\frac{n}{2\hat{\sigma}^4} \end{pmatrix}$$

- este negativ definită (matricea $\sum_{i=1}^{n}\boldsymbol{x}_i\boldsymbol{x}_i^\intercal$ este pozitiv definită)

Scor și Informația Fisher

Funcția de scor, matricea Hessiană și matricea Informațională a lui Fisher

Obiective

În această secțiune ne propunem să:

- definim *funcția de scor*
- matricea Hessiană
- *matricea informațională a lui Fisher* asociată unui eșantion
- matricea informațională a lui Fisher pentru o observație pentru repartiția marginală și repartiția condiționată
- matricea informațională a lui Fisher medie

Funcția de scor

Definition (Funcția de scor)

Se numește funcție de scor (variabilă de scor) asociată eșantionului $(X_i, Y_i)_{i=1}^{n}$ vectorul aleator

$$\underset{k\times 1}{s_n(\theta; Y|x)} = \frac{\partial \ell_n(\theta; Y|x)}{\partial \theta} = \begin{pmatrix} \frac{\partial \ell_n(\theta; Y|x)}{\partial \theta_1} \\ \cdots \\ \frac{\partial \ell_n(\theta; Y|x)}{\partial \theta_k} \end{pmatrix}$$

- Vectorul de scor $s_n(\theta; Y|x)$ este un vector aleator deoarece depinde de variabilele aleatoare $Y_1, \ldots, Y_n$
- În cazul necondiționat, funcția de scor este $s_n(\theta; X) = \frac{\partial \ell_n(\theta; X)}{\partial \theta}$

Funcția de scor (proprietăți)

Condiții de regularitate:

- (R0) Identifiabilitate: $\theta \neq \theta^* \Rightarrow f_\theta(x) \neq f_{\theta^*}(x)$
- (R1) $\mathrm{supp} f_\theta = \mathrm{supp} f_{\theta^*}$, $\forall \theta \neq \theta^*$, unde $\mathrm{supp} f_\theta = \overline{\{x \mid f_\theta(x) \neq 0\}}$
- (R2) Pentru orice $y \in \mathrm{supp} f_\theta(\cdot|x)$, $\theta \in \mathring{\Omega}$ și orice i, j derivatele parțiale de ordin doi $\frac{\partial^2 f_\theta(y|x)}{\partial\theta_i \partial\theta_j}$ există și sunt finite
- (R3) Integrala $\int f_\theta(y|x)\, dy$ poate fi diferențiată (parțial) de două ori sub semnul integral ca funcție de θ (se poate interschimba $\int \longleftrightarrow \frac{\partial}{\partial\theta}$)

Lemma

În condițiile de regularitate R0-R3, vectorul de scor verifică

$$\mathbb{E}_\theta[s_n(\theta; Y|x)] = 0_k$$

- Media $\mathbb{E}_\theta$ se calculează în raport cu repartiția condiționată $Y|X = x$

Funcția de scor (cont.)

Demonstrație: Avem, pentru componenta i,

$$\begin{aligned}
\mathbb{E}_\theta\left[\frac{\partial \ell_n(\theta; Y|x)}{\partial \theta_i}\right] &= \int \frac{\partial \ell_n(\theta; y|x)}{\partial \theta_i} f_\theta(y|x)\,dy \\
&= \int \frac{\partial \log f_\theta(y|x)}{\partial \theta_i} f_\theta(y|x)\,dy \\
&= \int \frac{1}{f_\theta(y|x)} \frac{\partial f_\theta(y|x)}{\partial \theta_i} f_\theta(y|x)\,dy \\
&= \int \frac{\partial f_\theta(y|x)}{\partial \theta_i}\,dy = \frac{\partial}{\partial \theta_i} \underbrace{\int f_\theta(y|x)\,dy}_{=1} = 0
\end{aligned}$$

Astfel, $\mathbb{E}_\theta[s_n(\theta; Y|x)] = \mathbb{E}_\theta \begin{bmatrix} \frac{\partial \ell_n(\theta; Y|x)}{\partial \theta_1} \\ \cdots \\ \frac{\partial \ell_n(\theta; Y|x)}{\partial \theta_k} \end{bmatrix} = 0_k.$ $\square$

Funcția de Scor - exemplul 1

Example (Modelul de regresie liniară)

Să considerăm modelul de regresie liniară:

$$y_i = \boldsymbol{x}_i^\intercal \boldsymbol{\beta} + \varepsilon_i$$

unde $\boldsymbol{x}_i = \begin{pmatrix} x_{i1} & \cdots & x_{ik} \end{pmatrix}^\intercal$ și $\boldsymbol{\beta} = \begin{pmatrix} \beta_1 & \cdots & \beta_k \end{pmatrix}^\intercal$ sunt $k \times 1$ vectori. Presupunem că ε_i sunt variabile aleatoare i.i.d. repartizate $\mathcal{N}(0, \sigma^2)$. Determinați funcția de scor $s_n(\theta; Y|x)$ și calculați $\mathbb{E}_\theta[s_n(\theta; Y|x)]$.

Soluție: Avem vectorul de scor

$$s_n(\theta; Y|x) = \begin{pmatrix} \frac{1}{\sigma^2} \sum_{i=1}^n \boldsymbol{x}_i (Y_i - \boldsymbol{x}_i^\intercal \beta) \\ -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^n (Y_i - \boldsymbol{x}_i^\intercal \beta)^2 \end{pmatrix}$$

Funcția de Scor - exemplul 1 (cont.)

Media vectorului de scor este

$$\mathbb{E}_\theta[s_n(\theta; Y|x)] = \mathbb{E}_\theta \begin{bmatrix} \frac{1}{\sigma^2}\sum_{i=1}^n \boldsymbol{x}_i(Y_i - \boldsymbol{x}_i^\intercal \beta) \\ -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^n (Y_i - \boldsymbol{x}_i^\intercal \beta)^2 \end{bmatrix}$$

și cum $\mathbb{E}_\theta[Y_i|x] = x_i^\intercal \beta$ deducem că

$$\begin{aligned} \mathbb{E}_\theta\left[\frac{1}{\sigma^2}\sum_{i=1}^n \boldsymbol{x}_i(Y_i - \boldsymbol{x}_i^\intercal \beta)\right] &= \frac{1}{\sigma^2}\sum_{i=1}^n \boldsymbol{x}_i(\mathbb{E}_\theta[Y_i|x] - \boldsymbol{x}_i^\intercal \beta) \\ &= \frac{1}{\sigma^2}\sum_{i=1}^n \boldsymbol{x}_i(\boldsymbol{x}_i^\intercal \beta - \boldsymbol{x}_i^\intercal \beta) = 0_k \end{aligned}$$

Funcția de Scor - exemplul 1 (cont.)

Pentru a doua componentă

$$
\begin{aligned}
\mathbb{E}_\theta\left[-\frac{n}{2\sigma^2}+\frac{1}{2\sigma^4}\sum_{i=1}^{n}(Y_i-\boldsymbol{x}_i^\intercal\beta)^2\right] &= -\frac{n}{2\sigma^2}+\frac{1}{2\sigma^4}\sum_{i=1}^{n}\mathbb{E}_\theta\left[(Y_i-\boldsymbol{x}_i^\intercal\beta)^2\right]\\
&= -\frac{n}{2\sigma^2}+\frac{1}{2\sigma^4}\sum_{i=1}^{n}\mathbb{E}_\theta\left[(Y_i-\mathbb{E}_\theta[Y_i|x])^2\right]\\
&= -\frac{n}{2\sigma^2}+\frac{1}{2\sigma^4}\sum_{i=1}^{n}Var_\theta(Y_i|x)\\
&= -\frac{n}{2\sigma^2}+\frac{1}{2\sigma^4}\sum_{i=1}^{n}\sigma^2 = 0.
\end{aligned}
$$

Prin urmare $\mathbb{E}_\theta[s_n(\theta; Y|x)] = 0_{k+1}$. $\square$

Gradient

Definition (Gradient)

Gradientul asociat (logaritmului) funcției de verosimilitate este

$$\underset{k\times 1}{\nabla \ell_n(\theta; y|x)} = \frac{\partial \ell_n(\theta; y|x)}{\partial \theta} = \begin{pmatrix} \frac{\partial \ell_n(\theta; y|x)}{\partial \theta_1} \\ \cdots \\ \frac{\partial \ell_n(\theta; y|x)}{\partial \theta_k} \end{pmatrix}$$

- Vectorul $\nabla \ell_n(\theta; y|x)$ este un vector determinist deoarece depinde de valorile $y_1, \ldots, y_n$

Remarcă

Ecuația de verosimilitate conduce la

$$\nabla \ell_n(\hat{\theta}; y|x) = 0_k$$

unde $\hat{\theta} = \hat{\theta}(x)$ este estimatorul de verosimilitate maximă a lui θ.

Matricea Hessiană

Definition (Matricea Hessiană)

Matricea Hessiană asociată (logaritmul) funcției de verosimilitate este

$$H_n(\theta; y|x) = \frac{\partial^2 \ell_n(\theta; y|x)}{\partial\theta\partial\theta^\intercal}$$

- Trebuie avut grijă atunci când vrem să distingem dintre $\frac{\partial^2 \ell_n(\theta; Y|x)}{\partial\theta\partial\theta^\intercal}$ (matrice cu elemente aleatoare) și $\frac{\partial^2 \ell_n(\theta; y|x)}{\partial\theta\partial\theta^\intercal}$ (matrice deterministă)

Sumar: funcție de scor / gradient

Sumarizând obținem

Aleator	Determinist
Vector de scor $\frac{\partial \ell_n(\theta; Y\|x)}{\partial \theta}$	Vector Gradient $\frac{\partial \ell_n(\theta; y\|x)}{\partial \theta}$
Matrice Hessiană $\frac{\partial^2 \ell_n(\theta; Y\|x)}{\partial \theta \partial \theta^\intercal}$	Matrice Hessiană $\frac{\partial^2 \ell_n(\theta; y\|x)}{\partial \theta \partial \theta^\intercal}$

Matricea Informațională a lui Fisher

Definition (Matricea Informațională a lui Fisher)

Matricea (condiționată) Informațională a lui Fisher asociată eșantionului $(X_i, Y_i)_{i=1}^{n}$ este matricea de varianță-covarianță a vectorului de scor:

$$\underbrace{I_n(\theta)}_{k\times k} = Var_\theta(s_n(\theta; Y|x)) = Var_\theta\left(\frac{\partial \ell_n(\theta; Y|x)}{\partial\theta}\right)$$

unde $Var_\theta(\cdot)$ este varianța calculată în raport cu repartiția $Y|X = x$.

Remarcă

În condițiile de regularitate $\mathbb{E}_\theta[s_n(\theta; Y|x)] = 0_k$ de unde deducem că

$$\underbrace{I_n(\theta)}_{k\times k} = \mathbb{E}_\theta[\underbrace{s_n(\theta; Y|x)}_{k\times 1} \times \underbrace{s_n(\theta; Y|x)^\intercal}_{1\times k}]$$

Matricea Informațională a lui Fisher (cont.)

- Elementul de pe linia i, coloana j al matricei $I_n(\theta)$ este

$$I_n(\theta)_{i,j} = Cov\left(\frac{\partial \ell_n(\theta; Y|x)}{\partial \theta_i}, \frac{\partial \ell_n(\theta; Y|x)}{\partial \theta_j}\right) = \mathbb{E}_\theta\left[\frac{\partial \ell_n(\theta; Y|x)}{\partial \theta_i}\frac{\partial \ell_n(\theta; Y|x)}{\partial \theta_j}\right]$$

Lemma

În condițiile de regularitate R0-R3, matricea informațională a lui Fisher verifică

$$I_n(\theta) = \mathbb{E}_\theta\left[-\frac{\partial^2 \ell_n(\theta; Y|x)}{\partial \theta \partial \theta^\intercal}\right] = \mathbb{E}_\theta\left[-H_n(\theta; Y|x)\right]$$

$$I_n(\theta)_{i,j} = -\mathbb{E}_\theta\left[\frac{\partial^2 \ell_n(\theta; Y|x)}{\partial \theta_i \partial \theta_j}\right]$$

Matricea Informațională a lui Fisher (cont.)

Remarcă

Matricea Informațională a lui Fisher, în condițiile de regularitate R0-R3, are formele alternative:

$$I_n(\theta) = Var_\theta(s_n(\theta; Y|x)) = Var_\theta\left(\frac{\partial \ell_n(\theta; Y|x)}{\partial \theta}\right)$$

$$I_n(\theta) = \mathbb{E}_\theta[s_n(\theta; Y|x) \times s_n(\theta; Y|x)^\intercal] = \mathbb{E}_\theta\left[\frac{\partial \ell_n(\theta; Y|x)}{\partial \theta} \times \left(\frac{\partial \ell_n(\theta; Y|x)}{\partial \theta}\right)^\intercal\right]$$

$$I_n(\theta) = \mathbb{E}_\theta\left[-H_n(\theta; Y|x)\right] = \mathbb{E}_\theta\left[-\frac{\partial^2 \ell_n(\theta; Y|x)}{\partial \theta \partial \theta^\intercal}\right]$$

Matricea Informațională a lui Fisher (cont.)

Definition (Matricea Informațională a lui Fisher)

Matricea Informațională a lui Fisher pentru observația i, i.e. (X_i, Y_i), este definită prin:

$$I_i(\theta) = Var_\theta(s_i(\theta; Y_i|x_i)) = Var_\theta\left(\frac{\partial \ell_i(\theta; Y_i|x_i)}{\partial \theta}\right)$$

$$I_i(\theta) = \mathbb{E}_\theta[s_i(\theta; Y_i|x_i) \times s_i(\theta; Y_i|x_i)^\intercal] = \mathbb{E}_\theta\left[\frac{\partial \ell_i(\theta; Y_i|x_i)}{\partial \theta} \times \left(\frac{\partial \ell_i(\theta; Y_i|x_i)}{\partial \theta}\right)^\intercal\right]$$

$$I_i(\theta) = \mathbb{E}_\theta\left[-H_i(\theta; Y_i|x_i)\right] = \mathbb{E}_\theta\left[-\frac{\partial^2 \ell_i(\theta; Y_i|x_i)}{\partial \theta \partial \theta^\intercal}\right]$$

Matricea Informațională a lui Fisher (cont.)

Lemma

Matricea informațională a lui Fisher asociată eșantionului este suma matricelor informaționale asociate observațiilor:

$$I_n(\theta) = \sum_{i=1}^{n} I_i(\theta).$$

- În cazul necondiționat, matricea informațională a lui Fisher asociată observației i este aceeași (eșantionul este presupus i.i.d.) pentru toate valorile

$$I_i(\theta) = I_1(\theta) \quad \forall i = 1, \dots, n$$

- În cazul condiționat, matricea informațională a lui Fisher asociată variabilei Y_i dat fiind $X_i = x_i$ depinde de observația i

$$I_i(\theta) \neq I_j(\theta) \quad \forall i \neq j$$

Matricea Informațională - exemplul 1

Example (Modelul de regresie liniară)

Am văzut că

$$\frac{\partial^2 \ell_n(\theta; y|x)}{\partial\theta\partial\theta^\intercal} = \begin{pmatrix} -\frac{1}{\sigma^2}\sum_{i=1}^{n} \underbrace{\boldsymbol{x}_i}_{k\times 1}\underbrace{\boldsymbol{x}_i^\intercal}_{1\times k} & -\frac{1}{\sigma^4}\sum_{i=1}^{n} \underbrace{\boldsymbol{x}_i}_{k\times 1}\underbrace{(y_i - \boldsymbol{x}_i^\intercal\beta)}_{1\times 1} \\ -\frac{1}{\sigma^4}\sum_{i=1}^{n} \underbrace{\boldsymbol{x}_i^\intercal}_{1\times k}\underbrace{(y_i - \boldsymbol{x}_i^\intercal\beta)}_{1\times 1} & \frac{n}{2\sigma^4} - \frac{1}{\sigma^6}\sum_{i=1}^{n}\underbrace{(y_i - \boldsymbol{x}_i^\intercal\beta)^2}_{1\times 1} \end{pmatrix}$$

Care este matricea informațională a lui Fisher asociată observației i ?

Matricea informațională - exemplul 1 (cont.)

Soluție: Cum matricea informațională a lui Fisher pentru o observație este:

$$I_i(\theta) = \mathbb{E}_\theta\left[-\frac{\partial^2 \ell_i\left(\theta; Y_i|x_i\right)}{\partial\theta\partial\theta^\top}\right] = \mathbb{E}_\theta\left[-H_i\left(\theta; Y_i|x_i\right)\right]$$

unde media $\mathbb{E}_\theta$ se calculează în raport cu repartiția condiționată $Y_i|X_i = x_i$, iar

$$\frac{\partial^2 \ell_i(\theta; y|x)}{\partial\theta\partial\theta^\intercal} = \begin{pmatrix} -\frac{1}{\sigma^2}\underbrace{\boldsymbol{x}_i}_{k\times 1}\underbrace{\boldsymbol{x}_i^\intercal}_{1\times k} & -\frac{1}{\sigma^4}\underbrace{\boldsymbol{x}_i}_{k\times 1}\underbrace{(y_i - \boldsymbol{x}_i^\intercal\beta)}_{1\times 1} \\ -\frac{1}{\sigma^4}\underbrace{\boldsymbol{x}_i^\intercal}_{1\times k}\underbrace{(y_i - \boldsymbol{x}_i^\intercal\beta)}_{1\times 1} & \frac{1}{2\sigma^4} - \frac{1}{\sigma^6}\underbrace{(y_i - \boldsymbol{x}_i^\intercal\beta)^2}_{1\times 1} \end{pmatrix}$$

Matricea informațională - exemplul 1 (cont.)

Găsim că

$$I_i(\theta) = \begin{pmatrix} \frac{1}{\sigma^2}\boldsymbol{x}_i\boldsymbol{x}_i^\intercal & \frac{1}{\sigma^4}x_i\left(\mathbb{E}_\theta\left[Y_i\right] - \boldsymbol{x}_i^\intercal\beta\right) \\ \frac{1}{\sigma^4}\boldsymbol{x}_i^\intercal\left(\mathbb{E}_\theta\left[Y_i\right] - \boldsymbol{x}_i^\intercal\beta\right) & -\frac{1}{2\sigma^4} + \frac{1}{\sigma^6}\mathbb{E}_\theta\left[\left(Y_i - \boldsymbol{x}_i^\intercal\beta\right)^2\right] \end{pmatrix}$$

Cum $\mathbb{E}_\theta[Y_i] = \boldsymbol{x}_i^\intercal\beta$ și $\mathbb{E}_\theta\left[\left(Y_i - \boldsymbol{x}_i^\intercal\beta\right)^2\right] = \sigma^2$, obținem că

$$I_i(\theta) = \begin{pmatrix} \frac{1}{\sigma^2}\boldsymbol{x}_i\boldsymbol{x}_i^\intercal & 0 \\ 0 & \frac{1}{2\sigma^4} \end{pmatrix}$$

Concluzie: cum $I_i(\theta)$ depinde de $\boldsymbol{x}_i$, se constată că $I_i(\theta) \neq I_j(\theta)$ pentru $i \neq j$.

Matricea informațională medie a lui Fisher

Definition (Matricea Informațională medie a lui Fisher)

Pentru modelul condiționat, matricea informațională medie a lui Fisher asociată eșantionului $(X_i, Y_i)_{i=1}^n$ este definită prin:

$$\mathrm{I}_n(\theta) = \mathbb{E}_X[I_n(\theta)]$$

unde $\mathbb{E}_X(\cdot)$ reprezintă media calculată în raport cu repartiția lui X (variabila de condiționare).

Astfel pentru **modelul condiționat** (și doar pentru acesta) avem

$$\mathrm{I}_n(\theta) = \mathbb{E}_X\left[Var_\theta\left(\frac{\partial \ell_n(\theta; Y|X)}{\partial\theta}\right)\right] = \mathbb{E}_X\left[\mathbb{V}_\theta\left(s_n(\theta; Y|X)\right)\right]$$

$$\mathrm{I}_n(\theta) = \mathbb{E}_X\mathbb{E}_\theta\left[\frac{\partial \ell_n(\theta; Y|X)}{\partial\theta}\frac{\partial \ell_n(\theta; Y|X)^\intercal}{\partial\theta}\right] = \mathbb{E}_X\mathbb{E}_\theta\left(s_n[\theta; Y|X)\, s_n(\theta; Y|X)^\intercal\right]$$

$$\mathrm{I}_n(\theta) = \mathbb{E}_X\mathbb{E}_\theta\left[-\frac{\partial^2 \ell_n(\theta; Y|X)}{\partial\theta\partial\theta^\intercal}\right] = \mathbb{E}_X\mathbb{E}_\theta\left[-H_n(\theta; Y|X)\right]$$

Matricea informațională medie a lui Fisher

Definition (Matricea Informațională medie a lui Fisher pentru o observație)

Matricea informațională medie a lui Fisher pentru observația i, i.e. (X_i, Y_i), este definită prin:

$$\mathrm{I}(\theta) = \mathbb{E}_X[I_i(\theta)]$$

Astfel, matricea informațională medie pentru o observație devine

$$\mathrm{I}(\theta) = \mathbb{E}_X\left[Var_\theta\left(\frac{\partial \ell_i\left(\theta; Y_i|X_i\right)}{\partial\theta}\right)\right] = \mathbb{E}_X\left[Var_\theta\left(s\left(\theta; Y_i|X_i\right)\right)\right]$$

$$\mathrm{I}(\theta) = \mathbb{E}_X\mathbb{E}_\theta\left[\frac{\partial \ell_i\left(\theta; Y_i|X_i\right)}{\partial\theta}\frac{\partial \ell_i\left(\theta; Y_i|X_i\right)^\intercal}{\partial\theta}\right]$$
$$= \mathbb{E}_X\mathbb{E}_\theta\left(s_i\left[\theta; Y_i|X_i\right) s_i\left(\theta; Y_i|X_i\right)^\intercal\right]$$

$$\mathrm{I}(\theta) = \mathbb{E}_X\mathbb{E}_\theta\left[-\frac{\partial^2 \ell_i\left(\theta; Y_i|X_i\right)}{\partial\theta\partial\theta^\intercal}\right] = \mathbb{E}_X\mathbb{E}_\theta\left[-H_i\left(\theta; Y_i|X_i\right)\right]$$

Matricea informațională medie a lui Fisher

Sumarizând: pentru a calcula matricea informațională medie $\mathrm{I}(\theta)$ pentru o observație:

- Pasul 1: Calculăm matricea Hessiană sau vectorul de scor pentru o observație

$$H_i\left(\theta; Y_i|x_i\right) = \frac{\partial^2 \ell_i\left(\theta; Y_i|x_i\right)}{\partial\theta\partial\theta^\top}, \quad s_i\left(\theta; Y_i|x_i\right) = \frac{\partial \ell_i\left(\theta; Y_i|x_i\right)}{\partial\theta}$$

- Pasul 2: Calculăm media, respectiv varianța, în raport cu repartiția condiționată $Y_i|X_i = x_i$

$$I_i(\theta) = \mathbb{E}_\theta\left[-H_i\left(\theta; Y_i|x_i\right)\right] = Var_\theta\left(s_i\left(\theta; Y_i|x_i\right)\right)$$

- Pasul 3: Calculăm media în raport cu repartiția variabilei condiționate X

$$\mathrm{I}(\theta) = \mathbb{E}_X\left[I_i(\theta)\right]$$

Matricea informațională medie - exemplu

Example (Modelul de regresie liniară)

În cazul modelului de regresie, matricea informațională a lui Fisher pentru o observație este

$$I_i(\theta) = \begin{pmatrix} \frac{1}{\sigma^2}\boldsymbol{x}_i\boldsymbol{x}_i^{\intercal} & 0 \\ 0 & \frac{1}{2\sigma^4} \end{pmatrix}$$

iar matricea informațională medie pentru o observație devine

$$\mathrm{I}(\theta) = \mathbb{E}_X\left[I_i(\theta)\right] = \begin{pmatrix} \frac{1}{\sigma^2}\mathbb{E}_X\left[\boldsymbol{X}_i\boldsymbol{X}_i^{\intercal}\right] & 0 \\ 0 & \frac{1}{2\sigma^4} \end{pmatrix}$$

Sumar

Sumarizând obținem (pentru observația i)

	Model necondiționat	Model condiționat
Densitate	$f_\theta(x_i)$	$f_\theta(y_i\|x_i)$
Vector de scor	$s_i(\theta; X_i) = \frac{\partial \ell_i(\theta; X_i)}{\partial \theta}$	$s_i(\theta; Y_i\|x_i) = \frac{\partial \ell_i(\theta; Y_i\|x_i)}{\partial \theta}$
Matrice Hessiană	$H_i(\theta; X_i) = \frac{\partial^2 \ell_i(\theta; X_i)}{\partial \theta \partial \theta^\intercal}$	$H_i(\theta; Y_i\|x_i) = \frac{\partial^2 \ell_i(\theta; Y_i\|x_i)}{\partial \theta \partial \theta^\intercal}$
Matricea Informațională	$I_i(\theta) = I_1(\theta)$	$I_i(\theta)$
Matricea Informațională medie	$\mathrm{I}(\theta) = I_1(\theta)$	$\mathrm{I}(\theta) = \mathbb{E}_X\left[I_i(\theta)\right]$

unde

$$I_i(\theta) = Var_\theta\left(s_i\left(\theta; Y_i|x_i\right)\right) = \mathbb{E}_\theta[s_i(\theta; Y_i|x_i) \times s_i(\theta; Y_i|x_i)^\intercal] = \mathbb{E}_\theta\left[-H_i\left(\theta; Y_i|x_i\right)\right]$$

Estimarea matricei informaționale medii

Remarcă

Dacă $\hat{\theta}$ este un estimator consistent pentru θ_0 (valoarea adevărată a parametrului care a generat eșantionul), atunci:

$$\hat{\mathrm{I}}(\hat{\theta}) = \frac{1}{n}\sum_{i=1}^{n} \hat{I}_i(\hat{\theta})$$

$$\hat{\mathrm{I}}(\hat{\theta}) = \frac{1}{n}\sum_{i=1}^{n} \left(\left.\frac{\partial \ell_i\left(\theta; y_i|x_i\right)}{\partial\theta}\right|_{\hat{\theta}} \times \left.\frac{\partial \ell_i\left(\theta; y_i|x_i\right)}{\partial\theta}\right|_{\hat{\theta}}^{\intercal} \right)$$

$$\hat{\mathrm{I}}(\hat{\theta}) = \frac{1}{n}\sum_{i=1}^{n} \left(- \left.\frac{\partial^2 \ell_i\left(\theta; y_i|x_i\right)}{\partial\theta\partial\theta^{\intercal}}\right|_{\hat{\theta}} \right)$$

sunt trei estimatori consistenți pentru matricea informațională medie a lui Fisher.

Estimarea matricei informaționale medii

- Primul estimator este rar folosit în practică și corespunde mediei empirice a celor n matrice informaționale ale lui Fisher (pentru $Y_1, \ldots, Y_n$) evaluate în valoarea estimatorului $\hat{\theta}$
- Al doilea estimator mai este numit și estimator *OPG* (outer product of gradients) sau, în literatura econometrică, *BHHH* (de la numele autorilor care l-au propus [Berndt et al., 1974]) - cel mai utilizat în practică
- Al treilea estimator este estimatorul bazat pe matricea Hessiană - de preferat

Proprietăți ale MLE

Proprietăți ale estimatorilor de verosimilitate maximă

Obiective

În această secțiune ne propunem să:

- să introducem *principiul de invarianță*
- să prezentăm o serie de *condiții de regularitate*
- să studiem *consistența* MLE
- să stabilim în ce condiții MLE este *eficient*
- să găsim *repartiția asimptotică* a MLE

Principiul de invarianță

O proprietate utilă a MLE este proprietatea de *invarianță*. Să presupunem că o repartiție este indexată după θ dar vrem să determinăm MLE pentru parametrul $\eta = \tau(\theta)$, e.g. $\sqrt{\theta}$ sau $\frac{\theta}{1-\theta}$

- dacă τ este bijectivă atunci funcția de verosimilitate scrisă ca o funcție de η devine

$$L_n^*(\eta; x) = L_n\left(\underbrace{\tau^{-1}(\eta)}_{\theta}; x\right)$$

prin urmare $\sup_{\eta} L_n^*(\eta; x) = \sup_{\eta} L_n\left(\tau^{-1}(\eta); x\right) = \sup_{\theta} L_n(\theta; x)$ deci maximul este atins în $\hat{\eta} = \tau(\hat{\theta})$ ($\hat{\theta}$ este MLE pentru L_n)

Principiul de invarianță (cont.)

- dacă τ nu este bijectivă atunci definim *funcția de verosimilitate indusă*

$$L_n^*(\eta; x) = \sup_{\{\theta \mid \tau(\theta) = \eta\}} L_n(\theta; x)$$

Definition

Se numește estimator de verosimilitate maximă a lui $\eta = \tau(\theta)$, valoarea $\hat{\eta}$ care maximizează funcția de verosimilitate indusă $L_n^*(\eta; x)$.

Principiul de invarianță (cont.)

Theorem (Proprietatea de invarianță)

Dacă $\hat{\theta}$ este estimatorul de verosimilitate maximă a lui θ atunci pentru orice funcție τ avem că $\tau(\hat{\theta})$ este estimatorul de verosimilitate maximă a lui $\tau(\theta)$.

Demonstrație: Fie $\hat{\eta}$ valoarea care maximizează $L_n^*(\eta; x)$. Avem că

$$\begin{aligned} L_n^*\left(\hat{\eta}; x\right) &= \sup_{\eta} L_n^*(\eta; x) = \sup_{\eta} \sup_{\{\theta \mid \tau(\theta)=\eta\}} L_n(\theta; x) \\ &= \sup_{\theta} L_n(\theta; x) = L_n\left(\hat{\theta}; x\right) \end{aligned}$$

Dar, cum $\hat{\theta}$ este MLE pentru θ, găsim că

$$L_n\left(\hat{\theta}; x\right) = \sup_{\left\{\theta \mid \tau(\theta)=\tau\left(\hat{\theta}\right)\right\}} L_n(\theta; x) = L_n^*\left(\tau\left(\hat{\theta}\right); x\right)$$

Condiții de regularitate

- (R4) Există o submulțime deschisă $\Theta_0 \subset \Theta$ astfel încât $\theta_0 \in \Theta_0$ (adevăratul parametru) și derivatele parțiale de ordin trei $\frac{\partial^3 f_\theta(x)}{\partial\theta_i \partial\theta_j \partial\theta_l}$ ale lui $f_\theta(x)$ există pentru orice $\theta \in \Theta_0$
- (R5) Sunt satisfăcute condițiile (în esență putem interschimba integrarea cu derivarea) pentru $i, j = 1, 2, \ldots, k$

$$\mathbb{E}_\theta\left[\frac{\partial}{\partial\theta_i}\log f_\theta(x)\right] = 0, \quad \mathrm{I}(\theta)_{ij} = \mathbb{E}_\theta\left[-\frac{\partial^2}{\partial\theta_i\partial\theta_j}\log f_\theta(x)\right]$$

- (R6) Pentru orice $\theta \in \Theta_0$, matricea informațională (medie) $\mathrm{I}(\theta)$ este pozitiv definită
- (R7) Există, pentru $i, j, l \in \{1, 2, \ldots, k\}$, funcțiile $M_{ijl}(x)$ astfel încât $\mathbb{E}_{\theta_0}[M_{ijl}(X)] < \infty$ și

$$\left|\frac{\partial^3}{\partial\theta_i\partial\theta_j\partial\theta_l}\log f_\theta(x)\right| \leq M_{ijl}(x), \quad \forall\theta \in \Theta_0.$$

Marginea Rao-Cramer

Theorem

Fie $X_1, X_2, \ldots, X_n$ un eșantion de talie n din populația f_θ, $\theta \in \Theta \subset \mathbb{R}^k$ și fie $T(X) = T(X_1, \ldots, X_n)$ un estimator cu $\mathbb{E}_\theta[T(X)] = h(\theta)$, h o funcție diferențiabilă de θ. Presupunem că

$$\frac{\partial}{\partial\theta}\mathbb{E}_\theta[T(X)] = \int T(x)\frac{\partial}{\partial\theta}f_\theta(x)\,dx, \quad \theta \in \Theta.$$

Atunci

$$Var_\theta(T(X)) \geq \left(\frac{\partial}{\partial\theta}h(\theta)\right)^\intercal I_n(\theta)^{-1}\frac{\partial}{\partial\theta}h(\theta)$$

unde $I_n(\theta)$ este matricea informațională a lui Fisher asociată eșantionului.

Comnsistența și normalitatea asimptotică

Theorem

Fie $X_1, X_2, \ldots, X_n$ un eșantion de talie n din populația f_θ, $\theta \in \Theta \subset \mathbb{R}^k$, unde $f_\theta(x)$ satisface condițiile de regularitate R0-R7. Atunci

A) *Sistemul de ecuații de verosimilitate $\frac{\partial l_n(\theta;x)}{\partial\theta} = 0$ admite un șir de soluții $\hat{\theta}_n$ consistente, i.e. $\hat{\theta}_n \xrightarrow{\mathbb{P}} \theta_0$*

B) *Pentru fiecare astfel de șir de soluții avem*

$$\sqrt{n}\left(\hat{\theta}_n - \theta_0\right) \xrightarrow{d} \mathcal{N}_k\left(0, \mathrm{I}(\theta_0)^{-1}\right)$$

unde θ_0 este valoarea reală a parametrului și $\mathrm{I}(\theta_0)$ este matricea informațională (medie) a lui Fisher pentru o observație.

- demonstrația acestui rezultat pentru cazul multidimensional poate fi consultată în [Lehmann and Casella, 1998, pag. 463]

Comnsistența și normalitatea asimptotică

Corollary

Fie $X_1, X_2, \ldots, X_n$ un eșantion de talie n din populația f_θ, $\theta \in \Theta \subset \mathbb{R}^k$, unde $f_\theta(x)$ satisface condițiile de regularitate R0-R7 și $\hat{\theta}_n$ un șir de soluții ale ecuației de verosimilitate. Atunci $\hat{\theta}_n$ sunt estimatori asimptotic eficienți în sensul că pentru $j = 1, 2, \ldots, k$ avem

$$\sqrt{n}\left(\hat{\theta}_{n,j} - \theta_j\right) \xrightarrow{d} \mathcal{N}\left(0, \mathrm{I}(\theta_0)_{jj}^{-1}\right).$$

Proprietăți ale MLE - exemplu

Example (Modelul de regresie liniară)

Revenim la modelul de regresie liniară:

$$y_i = \boldsymbol{x}_i^\intercal \boldsymbol{\beta} + \varepsilon_i$$

unde $\boldsymbol{x}_i = \begin{pmatrix} x_{i1} & \cdots & x_{ik} \end{pmatrix}^\intercal$ și $\boldsymbol{\beta} = \begin{pmatrix} \beta_1 & \cdots & \beta_k \end{pmatrix}^\intercal$ sunt $k \times 1$ vectori. Presupunem că ε_i sunt variabile aleatoare i.i.d. repartizate $\mathcal{N}(0, \sigma^2)$. Atunci estimatorul de verosimilitate maximă a vectorului, $(k+1) \times 1$, de parametrii $\theta = \begin{pmatrix} \boldsymbol{\beta}^\intercal & \sigma^2 \end{pmatrix}^\intercal$ este $\hat{\theta} = \begin{pmatrix} \hat{\boldsymbol{\beta}}^\intercal & \hat{\sigma}^2 \end{pmatrix}^\intercal$ cu

$$\hat{\beta} = \left(\sum_{i=1}^{n} \boldsymbol{X}_i \boldsymbol{X}_i^\intercal\right)^{-1} \left(\sum_{i=1}^{n} \boldsymbol{X}_i Y_i\right) \quad \text{și} \quad \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} \left(Y_i - \boldsymbol{X}_i^\intercal \hat{\beta}\right)^2$$

Care este repartiția asimptotică a lui $\hat{\theta}$?

Proprietăți ale MLE - exemplu

Soluție: Modelul satisface condițiile de regularitate. Am văzut că matricea informațională medie a lui Fisher este

$$\mathrm{I}(\theta) = \mathbb{E}_X\left[I_i(\theta)\right] = \begin{pmatrix} \frac{1}{\sigma^2}\mathbb{E}_X\left[\boldsymbol{X}_i\boldsymbol{X}_i^{\intercal}\right] & 0 \\ 0 & \frac{1}{2\sigma^4} \end{pmatrix}$$

iar din proprietatea de normalitate asimptotică a MLE găsim că

$$\sqrt{n}\left(\hat{\theta}_n - \theta_0\right) \xrightarrow{d} \mathcal{N}_k\left(0, \mathrm{I}(\theta_0)^{-1}\right)$$

unde θ_0 este valoarea reală a parametrului care a generat observațiile.

Proprietăți ale MLE - exemplu

Matricea de varianță-covarianță asimptotică a lui $\hat{\theta}$ este $n^{-1}\mathrm{I}(\theta_0)^{-1} = \mathrm{I}_n(\theta_0)^{-1}$ unde

$$\mathrm{I}_n(\theta) = \begin{pmatrix} \frac{n}{\hat{\sigma}^2}\mathbb{E}_X\left[\boldsymbol{X}_i\boldsymbol{X}_i^\intercal\right] & 0 \\ 0 & \frac{n}{2\hat{\sigma}^4} \end{pmatrix}$$

- un estimator consistent este $\hat{\mathrm{I}}_n(\theta) = \begin{pmatrix} \frac{n}{\hat{\sigma}^2}\sum_{i=1}^n \boldsymbol{x}_i\boldsymbol{x}_i^\intercal & 0 \\ 0 & \frac{n}{2\hat{\sigma}^4} \end{pmatrix}$

Ceea ce implică: $\hat{\beta} \approx \mathcal{N}\left(\beta_0, \hat{\sigma}^2\sum_{i=1}^n \boldsymbol{x}_i\boldsymbol{x}_i^\intercal\right)$ și $\hat{\sigma}^2 \approx \mathcal{N}\left(\sigma_0^2, \frac{2\hat{\sigma}^4}{n}\right)$

Comportament la transformări

Theorem

În condițiile de regularitate R0-R7, dacă $g : \mathbb{R}^k \to \mathbb{R}^p$ este o funcție diferențiabilă cu derivatele parțiale continue și $\hat{\theta}_n$ este un șir de soluții ale ecuației de verosimilitate atunci $g\left(\hat{\theta}_n\right) \xrightarrow{\mathbb{P}} g(\theta_0)$ și

$$\sqrt{n}\left(g\left(\hat{\theta}_n\right) - g(\theta_0)\right) \xrightarrow{d} \mathcal{N}_p\left(0, G(\theta_0)\mathrm{I}(\theta_0)^{-1}G(\theta_0)^\intercal\right)$$

unde θ_0 este valoarea reală a parametrului și matricea $G(\theta_0)$ este definită prin

$$\underset{p\times k}{G(\theta_0)} = \frac{\partial g(\theta)}{\partial \theta^\intercal}.$$

Bibliografie I

E.K. Berndt, B.H. Hall, R.E. Hall, and J.A. Hausman. Estimation and inference in nonlinear structural models. *Annals of Economic and Social Measurement*, 3(4):653–665, 1974. (Citat la pagina 95.)

E.L. Lehmann and G. Casella. *Theory of Point Estimation*. Springer Verlag, 1998. ISBN 0387985026. (Citat la pagina 103.)