$\boxed{\text{Stimatore vincolato e test F per } H_0 : R'\beta - r = 0}$

$$\underset{n\times 1}{y} = \underset{n\times K}{X}\underset{K\times 1}{\beta} + \underset{n\times 1}{\varepsilon}\ , \quad \varepsilon \mid X \sim \left(\underset{n\times 1}{0}, \sigma^2 \underset{n\times n}{I_n}\right) \qquad \text{rango}(\underset{n\times K}{X}) = K < n$$

$$H_0 : R'\beta - r = 0 \qquad \text{rango}(\underset{K\times q}{R}) = q \leq K$$

$$\hat{\beta} = (X'X)^{-1}X'y \quad \text{stimatore OLS (non vincolato)}$$

$$F_n := \frac{(R'\hat{\beta} - r)'\left[R'(X'X)^{-1}R\right]^{-1}(R'\hat{\beta} - r)/q}{\hat{\sigma}^2}\ , \quad \hat{\sigma}^2 := \frac{\hat{\varepsilon}'\hat{\varepsilon}}{n-K}\ , \quad \hat{\varepsilon} = y - X\hat{\beta}$$

$$\tilde{\beta} = \hat{\beta}_{RLS} = \hat{\beta} - (X'X)^{-1}R\left[R'(X'X)^{-1}R\right]^{-1}(R'\hat{\beta} - r)$$

$$\tilde{\varepsilon} := y - X\tilde{\beta} \quad \text{(residui modello vincolato)}$$

Proprietà dei residui $\tilde{\varepsilon}$

$M_X = I_n - X(X'X)^{-1}X'$ → proietta sul sottospazio ortogonale a X

$M_X \cdot \tilde{\varepsilon}$ = residui della regressione di $\tilde{\varepsilon}$ su X

$$M_X\tilde{\varepsilon} = M_X(y - X\tilde{\beta}) = \underbrace{M_X \cdot y} - \underbrace{M_X X}_{0}\tilde{\beta} = M_X\varepsilon = \hat{\varepsilon}$$

Consideriamo $\tilde{\varepsilon}'\tilde{\varepsilon}$ (somme quadrati residui modello vincolato):

$$\tilde{\varepsilon}'\tilde{\varepsilon} = \tilde{\varepsilon}'(I_n)\tilde{\varepsilon} = \tilde{\varepsilon}'(P_X + M_X)\tilde{\varepsilon}$$ in quanto $M_X = I_n - P_X$

$$= \tilde{\varepsilon}'P_X\tilde{\varepsilon} + \tilde{\varepsilon}'M_X\tilde{\varepsilon}$$

$$= \tilde{\varepsilon}'P_X\tilde{\varepsilon} + \hat{\varepsilon}'\hat{\varepsilon}$$ in quanto $\tilde{\varepsilon}'M_X\tilde{\varepsilon} = (M_X\tilde{\varepsilon})'(M_X\tilde{\varepsilon}) = \hat{\varepsilon}'\hat{\varepsilon}$

Quindi:

$$\tilde{\varepsilon}'\tilde{\varepsilon} - \hat{\varepsilon}'\hat{\varepsilon} = \tilde{\varepsilon}'P_X\tilde{\varepsilon}$$

Abbiamo

$$P_X\tilde{\varepsilon} = P_X(y - X\tilde{\beta}) = P_X y - P_X X\tilde{\beta} = X\underbrace{(X'X)^{-1}X'y}_{\hat{\beta}} - X\tilde{\beta}$$

$$= X\hat{\beta} - X\tilde{\beta} = X(\hat{\beta} - \tilde{\beta})$$

da cui

$$\tilde{\varepsilon}' P_X \tilde{\varepsilon} = (P_X \tilde{\varepsilon})'(P_X \tilde{\varepsilon}) = (\hat{\beta} - \tilde{\beta})' X' X (\hat{\beta} - \tilde{\beta})$$

dove

$$\hat{\beta} - \tilde{\beta} = (X'X)^{-1} R \left[R'(X'X)^{-1} R \right]^{-1} (R'\hat{\beta} - r)$$

Quindi

$$(\hat{\beta} - \tilde{\beta})' X'X (\hat{\beta} - \tilde{\beta}) = (R'\hat{\beta} - r)' \left[R'(X'X)^{-1} R \right]^{-1} R'(X'X)^{-1} \cancel{X'X} \cancel{(X'X)^{-1}} R \cdot$$

$$\cdot \left[R'(X'X)^{-1} R \right]^{-1} (R'\hat{\beta} - r)$$

$$= (R'\hat{\beta} - r)' \left[R'(X'X)^{-1} R \right]^{-1} (R'\hat{\beta} - r)$$

che corrisponde al numeratore della statistica F (molti-plicato per q)

Possiamo quindi concludere che:

Proprietà:

$$F_n = \frac{(\tilde{\varepsilon}'\tilde{\varepsilon} - \hat{\varepsilon}'\hat{\varepsilon})/q}{\hat{\varepsilon}'\hat{\varepsilon}/(n-K)}$$

Il test F valuta se l'incremento di variabilità residua che si manifesta passando dal modello <u>non</u> vincolato al modello vincolato sia significativo.

Nota:

$$F_n = \frac{(\tilde{\varepsilon}'\tilde{\varepsilon} - \hat{\varepsilon}'\hat{\varepsilon})/q}{\hat{\varepsilon}'\hat{\varepsilon}/(n-k)} = \frac{n-K}{q} \cdot \frac{\tilde{\varepsilon}'\tilde{\varepsilon} - \hat{\varepsilon}'\hat{\varepsilon}}{\hat{\varepsilon}'\hat{\varepsilon}}$$

$$\hat{R}^2_n := 1 - \frac{\hat{\varepsilon}'\hat{\varepsilon}}{D_n} \qquad D_n = \text{devianza totale} = (y - \bar{y}1_n)'(y - \bar{y}1_n)$$

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i, \quad 1_n = (1,1,\dots,1)'$$

$$\tilde{R}^2_n = 1 - \frac{\tilde{\varepsilon}'\tilde{\varepsilon}}{D_n} \qquad R^2 \text{ del modello vincolato.}$$

Quindi:

$$F_n = \frac{n-K}{q} \cdot \frac{\left(\tilde{\varepsilon}'\tilde{\varepsilon} - \hat{\varepsilon}'\hat{\varepsilon}\right)}{\hat{\varepsilon}'\hat{\varepsilon}} = \frac{n-K}{q} \cdot \frac{\frac{\tilde{\varepsilon}'\tilde{\varepsilon}}{D_n} - \frac{\hat{\varepsilon}'\hat{\varepsilon}}{D_n}}{\frac{\hat{\varepsilon}'\hat{\varepsilon}}{D_n}}$$

$$= \frac{n-K}{q} \cdot \frac{\left(1 - \frac{\hat{\varepsilon}'\hat{\varepsilon}}{D_n}\right) - \left(1 - \frac{\tilde{\varepsilon}'\tilde{\varepsilon}}{D_n}\right)}{1 - \left(1 - \frac{\hat{\varepsilon}'\hat{\varepsilon}}{D_n}\right)}$$

$$\boxed{F_n = \frac{n-K}{q} \cdot \frac{\hat{R}_n^2 - \tilde{R}_n^2}{1 - \hat{R}_n^2}}$$

Stima di massima verosimiglianza

<u>Esempio</u> [LOCATION MODEL]

$$y_i = \mu + \varepsilon_i \quad , \quad \varepsilon_i \sim iid\,(0, \sigma^2) \quad , \quad 0 < \sigma^2 < \infty$$

Lo stimatore OLS di μ è $\hat{\mu} = \bar{y}_n = \frac{1}{n} \sum_{i=1}^{n} y_i$

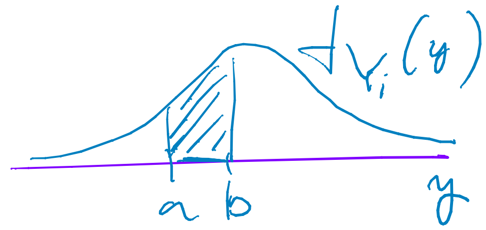
" " " " σ^2 è $\hat{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^{n} (y_i - \hat{\mu})^2$

Ipotizziamo che ε_i sia gaussiano

$$\varepsilon_i \sim \text{i.i.d. } N(0, \sigma^2)$$

Notiamo che questa ipotesi implica anche che:

$$y_i \sim \text{iid } N(\mu, \sigma^2)$$



La funzione di densità di y_i è:

$$f_{Y_i}(y;\theta) := \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}(y-\mu)^2\right\}$$

dove $\theta := (\mu, \sigma^2)'$ [vettore dei parametri da stimare]

La funzione di densità congiunta di $y_1, y_2, \dots, y_n$ è:

$$f_{Y_1,Y_2,\dots,Y_n}(y_1, y_2, \dots, y_n; \theta) = \prod_{i=1}^{n} f_{Y_i}(y_i;\theta)$$ [per l'ipotesi di indip]

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

$f: \mathbb{R}^n \to \mathbb{R}^+ \cup \{0\}$

Definiamo la funzione di verosimiglianza come la f. di densità congiunta in cui fissiamo $y_1, y_2, \dots, y_n$ ai corrispondenti valori campionari. La f.d.v. è funzione di $\theta = (\mu, \sigma^2)'$.

$$\mathcal{L}(\theta) = f_{y_1, \dots, y_n}(y_1, y_2, \dots, y_n; \theta) \qquad \mathcal{L}: \mathbb{R}^k \to \mathbb{R}^+ \cup \{0\}$$

valori campionari

Lo stimatore di massima verosimiglianza è il punto $\hat{\theta}_n$ dello spazio dei parametri che rende massima $\mathcal{L}(\theta)$:

$$\hat{\theta}_n := \underset{\theta \in \Theta}{\arg\max}\, \mathcal{L}(\theta)$$

Θ = spazio dei parametri

($\Theta = \mathbb{R} \times \mathbb{R}^+$ nel nostro caso)

Per trovare $\hat{\theta}_n$ è utile lavorare con la log-verosimiglianza:

$$\ell(\theta) := \log \mathcal{L}(\theta) = \log f_{Y_1,\dots,Y_n}(y_1 \dots y_n; \theta)$$

$$= \log \prod_{i=1}^{n} f_{Y_i}(y_i;\theta) = \sum_{i=1}^{n} \underbrace{\log f_{Y_i}(y_i;\theta)}_{\ell_i(\theta)}$$

Poiché $\log(x)$ è una funzione monotona,

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

Nel nostro esempio,

$$\ell_i(\theta) = \log f_{Y_i}(y_i) = \log\left\{\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(y_i-\mu)^2}\right\} = -\frac{1}{2}\log(2\pi) - \frac{1}{2}\log\sigma^2$$
$$-\frac{1}{2\sigma^2}(y_i-\mu)^2$$

da cui

$$\ell(\theta) = \sum_{i=1}^{n} \ell_i(\theta) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\mu)^2 \quad , \quad \theta = \begin{pmatrix}\mu \\ \sigma^2\end{pmatrix}$$

Calcoliamo $\hat{\theta}_n = (\hat{\mu}, \hat{\sigma}^2)'$ nel nostro location model.

Notiamo subito che massimizzare $\ell(\theta)$ rispetto a μ equivale a minimizzare

$$\sum_{i=1}^{n}(y_i-\mu)^2$$

che corrisponde al problema dei minimi quadrati:

$$\hat{\mu} = \bar{y}_n = \frac{1}{n}\sum_{i=1}^{n} y_i$$

$f(x) = \log x + \frac{1}{x}$ a

La log-verosimiglianza, massimizzata rispetto a μ, diventa:

$$\ell(\hat{\mu}, \sigma^2) = -\frac{n}{2}\log 2\pi - \frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\bar{y}_n)^2,$$

da massimizzare rispetto a σ^2. La derivata prima è:

$$\frac{\partial \ell(\hat{\mu}, \sigma^2)}{\partial \sigma^2} = -\frac{n}{\cancel{2}\cancel{\sigma^2}} + \frac{1}{\cancel{2}\sigma^{\cancel{4}\,2}}\sum_{i=1}^{n}(y_i-\bar{y}_n)^2 = 0$$

(in quanto $\sigma^2 \neq 0$)

$$\frac{\partial}{\partial x}\log x = \frac{1}{x}$$

$$\frac{\partial}{\partial x} x^{-1} = -\frac{1}{x^2}$$

$$\frac{1}{\sigma^2}\sum_{i=1}^{n}(y_i-\bar{y}_n)^2 = n$$

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n}(y_i-\bar{y}_n)^2}{n}$$

Riassumendo

$$\hat{\theta}_n = \begin{pmatrix}\hat{\mu}\\ \hat{\sigma}^2\end{pmatrix}, \quad \hat{\mu} = \underbrace{\bar{y}_n = \frac{1}{n}\sum y_i}_{= \text{ stimatore OLS}}, \quad \hat{\sigma}^2 = \underbrace{\frac{1}{n}\sum_{i=1}^{n}(y_i-\bar{y}_n)^2}_{\neq \text{ dallo stimatore OLS: } \frac{1}{n-1}\sum_{i=1}^{n}(y_i-\bar{y}_n)^2}$$

<u>nota</u>: lo stimatore ML (maximum likelihood) non è in generale <u>corretto</u>.

<u>nota</u> lo stimatore ML di μ coincide con lo stimatore OLS, $\bar{y}_n$.
Può avere buone proprietà (in questo caso: B.L.U.E.)
anche se ε_i non è normale

Consideriamo ora un generico modello di regressione lineare

$$\underset{n\times 1}{y} = \underset{n\times K}{X}\underset{K\times 1}{\beta} + \underset{n\times 1}{\varepsilon}, \quad \underset{n\times 1}{E[\varepsilon|X]} = \underset{n\times 1}{0} \quad \underset{n\times n}{V[\varepsilon|X]} = E[\underbrace{\underset{n\times 1}{\varepsilon}\underset{1\times n}{\varepsilon'}}_{n\times n}|X] = \sigma^2 \underset{n\times n}{I_n}$$

Stimatore OLS: $\hat{\beta} = (X'X)^{-1}X'y$, $\hat{\sigma}^2 = \dfrac{\hat{\varepsilon}'\hat{\varepsilon}}{n-K}$, $\hat{\varepsilon} := y - X\hat{\beta}$

Ipotizziamo che $\varepsilon|X$ (o, equivalentemente, $y|X$) sia gaussiano:

$$\underset{n\times 1}{\varepsilon|X} \sim N(\underset{n\times 1}{0}, \sigma^2 \underset{n\times n}{I_n})$$

ovvero:

$$y|X \sim N(X\beta, \sigma^2 I_n) \implies$$ y_i e y_j, dato X, sono indipendenti $(i \neq j)$

La funzione di densità di y (dato X) corrisponde alla f. di densità della normale multivariata:

$$f_{Y|X}(y;\beta,\sigma^2) = \prod_{i=1}^{n} f_{Y_i|X}(y_i;\beta,\sigma^2)$$

$$= \prod_{i=1}^{n} \underbrace{\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}(y_i - x_i'\beta)^2\right\}}_{y_i|X \sim N(x_i'\beta,\,\sigma^2)}$$

La log-verosimiglianza è quindi:

$$\ell(\theta) = \ell(\beta,\sigma^2) = -\frac{n}{2}\log 2\pi - \frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\underbrace{\sum_{i=1}^{n}(y_i - x_i'\beta)^2}_{(y-X\beta)'(y-X\beta)}$$

$$\underset{(K+1)\times 1}{\theta} = \begin{pmatrix}\beta\\ \sigma^2\end{pmatrix}$$ vettore dei parametri

nota: $\ell(\theta) = \sum_{i=1}^{n} \underbrace{\log f_{Y_i|X}(y_i;\theta)}_{\ell_i(\theta)} = \sum_{i=1}^{n} \ell_i(\theta)$

$$\ell_i(\theta) = -\frac{1}{2}\log 2\pi - \frac{1}{2}\log\sigma^2 - \frac{1}{2\sigma^2}(y_i - x_i'\beta)^2$$

(contributo unità i-esima alla log-verosimiglianza complessiva)

Vogliamo trovare $\hat{\theta} := \begin{pmatrix} \hat{\beta} \\ \hat{\sigma}^2 \end{pmatrix} := \underset{\theta\in\Theta}{\arg\max}\ \ell(\theta)$

Esaminando $L(\theta)$ notiamo che il β che massimizza $L(\theta)$ è il β che minimizza

$$Q(\beta) = (y - X\beta)'(y - X\beta)$$

ovvero

$$\hat{\beta}_{OLS} := (X'X)^{-1}X'y \qquad (\text{richiede rango}(X) = K = \max)$$

Sostituendo β con $\hat{\beta}$ otteniamo la log-verosimiglianza (concentrata rispetto a β):

$$\ell^c(\sigma^2) = \ell(\hat{\beta}, \sigma^2) = -\frac{n}{2}\log 2\pi - \frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}(y - X\hat{\beta})'(y - X\hat{\beta})$$

$$= -\frac{n}{2}\log 2\pi - \frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\underbrace{\hat{\varepsilon}'\hat{\varepsilon}}_{\text{devianza residua}} \qquad (\hat{\varepsilon} := y - X\hat{\beta})$$

$$-\frac{\hat{\varepsilon}'\hat{\varepsilon}}{2}\cdot\left[\frac{\partial}{\partial\sigma^2}(\sigma^2)^{-1}\right]$$

$\ell^c(\sigma^2)$ è derivabile rispetto a σ^2:

$$\frac{\partial \ell^c(\sigma^2)}{\partial\sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\hat{\varepsilon}'\hat{\varepsilon} = 0$$

da cui

$$\hat{\sigma}^2 = \frac{\hat{\varepsilon}'\hat{\varepsilon}}{n} \qquad \left(\neq \hat{\sigma}^2_{OLS} = \frac{\hat{\varepsilon}'\hat{\varepsilon}}{n-k}\right)$$

Il massimo di $\ell(\theta)$, ovvero $\ell(\hat{\theta}) = \ell(\hat{\beta}, \hat{\sigma}^2)$ è dato da:

$$\ell(\hat\theta) = -\frac{n}{2}\log 2\pi - \frac{n}{2}\log\hat\sigma^2 - \frac{1}{2\hat\sigma^2}\hat\varepsilon'\hat\varepsilon$$

$$= -\frac{n}{2}\log 2\pi - \frac{n}{2}\log\hat\sigma^2 - \frac{n}{2}$$

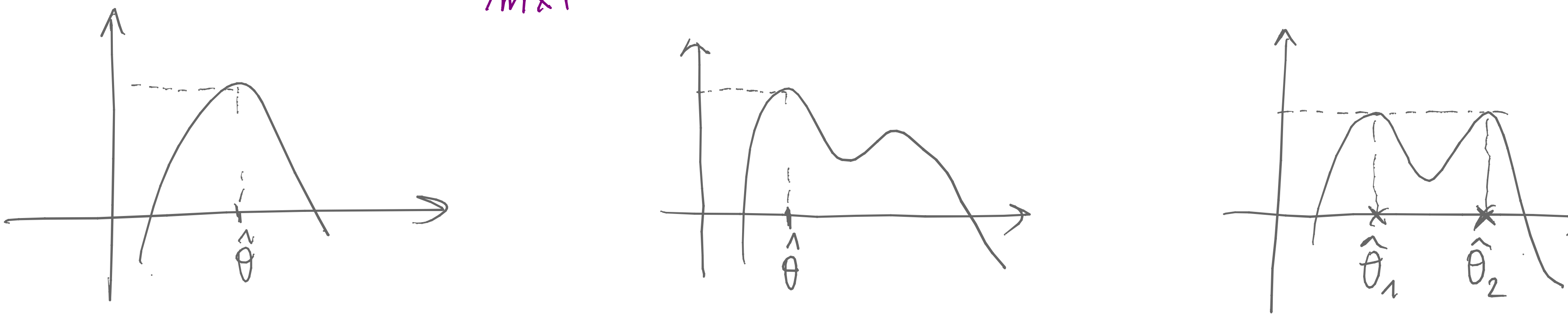
$$\frac{\hat\varepsilon'\hat\varepsilon}{\hat\sigma^2} = \frac{n\cdot\cancel{\hat\sigma^2}}{\cancel{\hat\sigma^2}} = n$$

Il massimo dipende solo da $\hat\sigma^2$

Caso generale

n osservazioni, $\ell(\theta) = \sum_{i=1}^{n} l_i(\theta)$ log-v.

$\underset{m\times 1}{\theta} \in \Theta$ (spazio dei parametri)



(a) (b) (c)

Condizione necessaria: Indicando con θ_0 il valore vero di θ, allora

$$\theta_0 \in \Theta$$

$$E\left[\ell(\theta_0)\right] > E\left[\ell(\theta)\right], \text{ per ogni } \theta \in \Theta \setminus \{\theta_0\}$$

(escludere il caso (c)) - Per n grande -

Se $\ell(\theta)$ è derivabile, possiamo considerare la derivata prima (SCORE):

$$\underset{m \times 1}{s(\theta)} := \frac{\partial \ell(\theta)}{\partial \theta} = \frac{\partial}{\partial \theta} \sum_{i=1}^{n} \ell_i(\theta) = \sum_{i=1}^{n} \frac{\partial}{\partial \theta} \ell_i(\theta) = \sum_{i=1}^{n} s_i(\theta)$$

$$s(\theta) = \frac{\partial \ell(\theta)}{\partial \theta} = \begin{bmatrix} \frac{\partial \ell(\theta)}{\partial \theta_1} \\ \vdots \\ \frac{\partial \ell(\theta)}{\partial \theta_m} \end{bmatrix}$$

lo score è $m \times 1$

Se $\ell(\theta)$ è derivabile, il punto di massimo $\hat{\theta}$ deve soddisfare:

$$s(\hat{\theta}) = 0$$

(condizione del primo ordine)

Nel nostro esempio (modello lineare)

$$s(\theta) = \sum_{i=1}^{n} s_i(\theta) = \sum_{i=1}^{n} \begin{pmatrix} \frac{\partial \ell_i(\theta)}{\partial \beta} \\ \frac{\partial \ell_i(\theta)}{\partial \sigma^2} \end{pmatrix} = \sum_{i=1}^{n} s_i(\theta)$$

$$\boxed{\frac{\partial a'x}{\partial x} = \frac{\partial x'a}{\partial x} = a}$$

$$\frac{\partial \ell_i(\theta)}{\partial \beta} = \frac{\partial}{\partial \beta}\left[-\frac{1}{2}\log 2\pi - \frac{1}{2}\log \sigma^2 - \frac{1}{2\sigma^2}(y_i - x_i'\beta)^2\right]$$

$$= -\frac{1}{2\sigma^2}\frac{\partial}{\partial \beta}(y_i - x_i'\beta)^2 = -\frac{1}{2\sigma^2}\, 2(y_i - x_i'\beta)\frac{\partial}{\partial \beta}(-x_i'\beta)$$

$$= \frac{1}{\sigma^2}(y_i - x_i'\beta)x_i = \frac{1}{\sigma^2}(y_i - \beta'x_i)x_i$$

$$\frac{\partial l_i(\theta)}{\partial \sigma^2} = \frac{\partial}{\partial \sigma^2}\left[-\frac{1}{2}\log 2\pi - \frac{1}{2}\log \sigma^2 - \frac{1}{2\sigma^2}(y_i - x_i'\beta)^2\right]$$

$$= -\frac{1}{2\sigma^2} + \frac{1}{2\sigma^4}(y_i - x_i'\beta)^2 = \frac{1}{2\sigma^2}\left(\frac{(y_i - x_i'\beta)^2}{\sigma^2} - 1\right)$$

Quindi

$$s_i(\theta) = s_i\left(\begin{pmatrix}\beta \\ \sigma^2\end{pmatrix}\right) = \frac{1}{\sigma^2}\begin{bmatrix} (y_i - \beta' x_i)x_i \\ \frac{1}{2}\left(\frac{(y_i - x_i'\beta)^2}{\sigma^2} - 1\right)\end{bmatrix}$$

e

$$s(\theta) = \sum_{i=1}^{n} s_i(\theta) = \frac{1}{\sigma^2}\begin{bmatrix} \sum_{i=1}^{n}(y_i - \beta' x_i)x_i \\ \frac{1}{2}\sum_{i=1}^{n}\left[\frac{(y_i - x_i'\beta)^2}{\sigma^2} - 1\right]\end{bmatrix}$$

Risolviamo le condizione del primo ordine $s(\theta) = 0$

Il primo blocco è $\displaystyle\sum_{i=1}^{n}(y_i-\beta' x_i)x_i = 0$, da cui

$$\sum_{i=1}^{n} x_i(y_i - x_i'\beta) = 0$$

$$\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i x_i' \beta = 0$$

$$\left(\sum_{i=1}^{n} x_i x_i'\right)\beta = \sum_{i=1}^{n} x_i y_i$$

se $\sum_{1}^{n} x_i x_i'$ ha rango K

$$\hat\beta = \left(\sum_{i=1}^{n} x_i x_i'\right)^{-1}\left(\sum_{i=1}^{n} x_i y_i\right) = (X'X)^{-1} X'y$$

Il secondo blocco è: $\displaystyle\sum_{i=1}^{n}\left[\frac{(y_i - x_i'\beta)^2}{\sigma^2} - 1\right] = 0$

$$\sum_{i=1}^{n}(y_i - x_i'\hat\beta)^2 - n\sigma^2 = 0$$

$$\hat\sigma^2 = \frac{\sum(y_i - x_i'\hat\beta)^2}{n} = \frac{\sum_{i=1}^{n}\hat\varepsilon_i^2}{n}, \quad \hat\varepsilon_i = y_i - x_i'\hat\beta$$

$$= \frac{\hat\varepsilon'\hat\varepsilon}{n}$$

<u>nota</u> allo stesso risultato si arriva usando la notazione compatta:

$$\ell(\theta) = -\frac{n}{2}\log 2\pi - \frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\underbrace{(y-X\beta)'(y-X\beta)}_{y'y-2\beta'X'y+\beta'X'X\beta}$$

$$s(\theta) = \begin{pmatrix} \frac{\partial \ell(\theta)}{\partial\beta} \\ \frac{\partial \ell(\theta)}{\partial\sigma^2} \end{pmatrix}, \text{ dove}$$

$$\frac{\partial \ell(\theta)}{\partial\beta} = -\frac{1}{2\sigma^2}\left(-2X'y + 2X'X\beta\right) = \frac{1}{\sigma^2}\left(X'y - X'X\beta\right)$$

$$\frac{\partial \ell(\theta)}{\partial\sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}(y-X\beta)'(y-X\beta)$$

Risolvendo $s(\theta)=0$ otteniamo

$$\begin{cases} X'y - X'X\beta = 0 \\ n\sigma^2 - (y-X\beta)'(y-X\beta) = 0 \end{cases} \qquad \begin{cases} \hat\beta = (X'X)^{-1}X'y \\ \hat\sigma^2 = \frac{\hat\varepsilon'\hat\varepsilon}{n} \end{cases}$$

Condizione del secondo ordine: se $\ell(\theta)$ è derivabile due volte, allora

$$H(\theta) := \frac{\partial^2}{\partial\theta\partial\theta'}\ell(\theta)$$ deve essere

negativa definita nel punto di massimo:

$H(\hat{\theta})$ negativa definita.

$x'Ax < 0$

$x'(-I_N)x = -\underbrace{x'x}_{>0} < 0$

[H si chiama matrice hessiana]

La matrice hessiana è definita come:

$$H(\theta) = \frac{\partial^2}{\partial\theta\partial\theta'} \ell(\theta) = \begin{bmatrix} \frac{\partial^2 \ell(\theta)}{(\partial\theta_1)^2} & \frac{\partial^2 \ell(\theta)}{\partial\theta_1 \partial\theta_2} & \cdots & \frac{\partial^2 \ell(\theta)}{\partial\theta_1 \partial\theta_m} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 \ell(\theta)}{\partial\theta_m \partial\theta_1} & \frac{\partial^2 \ell(\theta)}{\partial\theta_m \partial\theta_2} & \cdots & \frac{\partial^2 \ell(\theta)}{(\partial\theta_m)^2} \end{bmatrix}$$

<u>Nota</u>: una quantità importante (in statistica) è la matrice di informazione di Fisher

$$I_n(\theta) := -E[H(\theta)]$$

Se il modello è correttamente specificato

$$I_n(\theta) = -E[H(\theta)] = E[s(\theta)s(\theta)']$$

(information equality)

Se $n \to \infty$, sotto certe condizioni

$$\frac{1}{n} I_n(\theta) \longrightarrow I_\infty(\theta)$$

$I_\infty(\theta)$ è la matrice di informazione asintotica

<u>Teorema</u> Sotto alcune condizioni di regolarità, lo stimatore di massima verosimiglianza:

$$\hat{\theta} \xrightarrow{P} \theta_0 \qquad \text{(consistenza)}$$

$$\sqrt{n}\,(\hat{\theta} - \theta_0) \xrightarrow{d} N(0, V) \qquad \text{(normalità asintotica)}$$

$$V = I_\infty^{-1}(\theta_0) \qquad \text{(efficienza)}$$

Esercizio: calcolare, per il modello lineare

$$H(\theta)$$

e mostrare che $H(\hat{\theta})$ è definita negativa

Nota: se il modello è correttamente specificato, allora

$$E[S(\theta_0)] = 0$$

Esercizio: Valutare $E[S(\theta_0)]$ e $V[S(\theta_0)]$ per il modello lineare.

Modello lineare — svolto

$$s(\theta) = \sum_{i=1}^{n} s_i(\theta) = \sum_{i=1}^{n} \underbrace{\begin{bmatrix} \frac{1}{\sigma^2}(y_i - \beta' x_i)\, x_i \\ \frac{1}{2\sigma^2}\left(\frac{(y_i - x_i'\beta)^2}{\sigma^2} - 1\right) \end{bmatrix}}_{s_i(\theta)}$$

Se $\theta = \theta_0$ (valore vero), allora $y_i - \beta' x_i = \varepsilon_i$ e lo score diventa

$$s(\theta_0) = \sum_{i=1}^{n} \begin{bmatrix} \frac{1}{\sigma^2} \varepsilon_i x_i \\ \frac{1}{2\sigma^2}\left(\frac{\varepsilon_i^2}{\sigma_i^2} - 1\right) \end{bmatrix},$$

e quindi $E(s(\theta_0)) = 0$ in quanto $E[\varepsilon_i x_i] = 0$ e $E\left[\frac{\varepsilon_i^2}{\sigma^2} - 1\right] = 0$.

Consideriamo l'hessiano:

$$H(\theta) = \sum_{i=1}^{n} H_i(\theta), \quad H_i(\theta) := \frac{\partial}{\partial\theta\partial\theta'} \ell_i(\theta) = \frac{\partial}{\partial\theta'} s_i(\theta).$$

Abbiamo:

$$H_i(\theta) = \frac{\partial}{\partial\theta'} s_i(\theta) = \frac{\partial}{\partial(\beta'\sigma^2)} \begin{bmatrix} \frac{1}{\sigma^2}(y_i - \beta' x_i)\, x_i \\ \frac{1}{2\sigma^2}\left(\frac{(y_i - x_i'\beta)^2}{\sigma^2} - 1\right) \end{bmatrix} = \begin{bmatrix} -\frac{1}{\sigma^2} x_i x_i' & -\frac{1}{\sigma^4} x_i (y_i - \beta' x_i) \\ -\frac{1}{\sigma^4}(y_i - \beta' x_i) x_i' & \frac{1}{2\sigma^4} - \frac{1}{\sigma^6}(y_i - x_i'\beta)^2 \end{bmatrix}$$

da cui

$$H(\theta) = \sum_{i=1}^{n} H_i(\theta) = \begin{bmatrix} -\frac{1}{\sigma^2}\sum x_i x_i' & -\frac{1}{\sigma^4}\sum x_i (y_i - \beta' x_i) \\ -\frac{1}{\sigma^4}\sum (y_i - \beta' x_i) x_i' & \frac{n}{2\sigma^4} - \frac{1}{\sigma^6}\sum (y_i - x_i'\beta)^2 \end{bmatrix} = \begin{bmatrix} -\frac{1}{\sigma^2} X'X & -\frac{1}{\sigma^4}[X'y - X'X\beta] \\ -\frac{1}{\sigma^4}[y'X - \beta' X'X] & \frac{n}{2\sigma^4} - \frac{1}{\sigma^6}(y - X\beta)'(y - X\beta) \end{bmatrix}$$

In corrispondenza di $\theta = \hat{\theta}$, abbiamo $y_i - \beta' x_i = \hat{\varepsilon}_i$ e otteniamo:

$$H_i(\hat{\theta}) = \frac{1}{\hat{\sigma}^2} \begin{bmatrix} -x_i x_i' & -\frac{1}{\hat{\sigma}^2}\hat{\varepsilon}_i x_i \\ -\frac{1}{\hat{\sigma}^2} x_i' \hat{\varepsilon}_i & \frac{1}{2\hat{\sigma}^2} - \frac{\hat{\varepsilon}_i^2}{\hat{\sigma}^4} \end{bmatrix} = -\frac{1}{\hat{\sigma}^4} \begin{bmatrix} \hat{\sigma}^2 x_i x_i' & \hat{\varepsilon}_i x_i \\ \hat{\varepsilon}_i x_i' & \frac{1}{2} - \frac{\hat{\varepsilon}_i^2}{\hat{\sigma}^2} \end{bmatrix}$$

Sommando:

$$H(\hat{\theta}) = \sum_{i=1}^{n} H_i(\hat{\theta}) = -\frac{1}{\hat{\sigma}^4} \begin{bmatrix} \hat{\sigma}^2 \sum x_i x_i' & \sum \hat{\varepsilon}_i x_i \\ \sum \hat{\varepsilon}_i x_i' & \frac{n}{2} - \frac{\sum \hat{\varepsilon}_i^2}{\hat{\sigma}^2} \end{bmatrix} = -\frac{1}{\hat{\sigma}^4} \begin{bmatrix} \hat{\sigma}^2 X'X & 0 \\ 0 & -\frac{n}{2} \end{bmatrix}$$

$$= - \begin{bmatrix} \frac{1}{\hat\sigma^2}(X'X) & 0 \\ 0 & \frac{n}{2\hat\sigma^4} \end{bmatrix}$$

in quanto IMPORTANTE! $\sum_{i=1}^{n} \hat\varepsilon_i x_i = X'\hat\varepsilon = X' M_X \varepsilon = 0$ – Quindi,

$H(\hat\theta)$ è negativa definita, per cui $\hat\theta$ è il punto di massimo

In corrispondente di θ_0, l'hessiana diventa:

$$H(\theta_0) = \begin{bmatrix} -\frac{1}{\sigma^2} X'X & -\frac{1}{\sigma^4} X'\varepsilon \\ \frac{1}{\sigma^4}\varepsilon' X & \frac{n}{2\sigma^4} - \frac{1}{\sigma^6}\varepsilon'\varepsilon \end{bmatrix}$$

e la matrice di informazione di Fisher è quindi

$$I_n(\theta) := -E[H(\theta_0)] = \begin{bmatrix} \frac{1}{\sigma^2} X'X & 0 \\ 0 & -\frac{n}{2\sigma^4} + \frac{n}{\sigma^4} \end{bmatrix} = \begin{bmatrix} \frac{1}{\sigma^2} X'X & 0 \\ 0 & \frac{n}{2\sigma^4} \end{bmatrix}$$

Il limite inferiore di Rao-Cramer è:

$$I_n(\theta_0)^{-1} = \begin{bmatrix} \sigma^2 (X'X)^{-1} & 0 \\ 0 & \frac{2\sigma^4}{n} \end{bmatrix}$$

Varianza dello stimatore $\hat{\beta}$