TRINOMIALE

$\begin{pmatrix} X \\ Y \end{pmatrix} \sim TRIN(n, \theta_1, \theta_2)$ $\quad R_{X,Y} = \{(x,y): x \in \{0,1,\dots,n\}, y \in \{0,1,\dots,n\}, x+y \in \{0,1,\dots,n\}\}$

$$p_{X,Y}(x,y) = \frac{n!}{x!\,y!\,(n-x-y)!}\,\theta_1^x \theta_2^y (1-\theta_1-\theta_2)^{n-x-y} \cdot \mathbb{1}_{R_{X,Y}}(x,y)$$

QUANDO SI USA? $\Rightarrow$ n repliche indipendenti di un esperimento aleatorio che ammette 3 esiti: e_1, e_2, e_3

$n \to$ intero positivo: 1, 2, 3, ...

$0 \le \theta_1 \le 1$ $\quad \theta_1$ = probabilità che si verifichi e_1

$0 \le \theta_2 \le 1$ $\quad \theta_2$ = probabilità che si verifichi e_2

$0 \le \theta_1 + \theta_2 \le 1$ $\quad 1-\theta_1-\theta_2$ = probabilità che si verifichi e_3

X = nr. repliche con esito e_1
Y = nr. repliche con esito e_2
Z = nr. repliche con esito e_3
$Z = n - X - Y$

PROPRIETÀ:

1) $X \sim Bin(n, \theta_1)$ $\to$ DIM:

$$p_X(x) = \sum_{y=0}^{n-x} p_{XY}(x,y) = \sum_{y=0}^{n-x} \frac{n!}{x!\,y!\,(n-x-y)!}\,\theta_1^x \cdot \theta_2^y \cdot (1-\theta_1-\theta_2)^{n-x-y} \cdot \frac{(n-x)!}{(n-x)!}$$

$$= \underbrace{\frac{n!}{x!\,(n-x)!}}_{\binom{n}{x}} \cdot \theta_1^x \cdot \underbrace{\sum_{y=0}^{n-x} \frac{(n-x)!}{y!\,(n-x-y)!} \cdot \theta_2^y \cdot (1-\theta_1-\theta_2)^{n-x-y}}_{\text{per teo. binomiale } = (1-\theta_1)^{n-x}}$$

$$= \binom{n}{x} \cdot \theta_1^x (1-\theta_1)^{n-x}$$

2) $Y \sim Bin(n, \theta_2)$ $\to$ DIM.:

$$p_Y(y) = \sum_{x=0}^{n-y} \frac{n!}{x!\,y!\,(n-x-y)!}\,\theta_1^x \theta_2^y (1-\theta_1-\theta_2)^{n-x-y} \cdot \frac{(n-y)!}{(n-y)!}$$

$$= \underbrace{\frac{n!}{y!\,(n-y)!}}_{\binom{n}{y}}\,\theta_2^y \cdot \underbrace{\sum_{x=0}^{n-y} \frac{(n-y)!}{x!\,(n-y-x)!} \cdot \theta_1^x \cdot (1-\theta_1-\theta_2)^{n-x-y}}_{\text{per teo. binomiale } = (1-\theta_2)^{n-y}}$$

$$= \binom{n}{y} \cdot \theta_2^y \cdot (1-\theta_2)^{n-y}$$

3) X e Y <u>non</u> sono indipendenti

4) $Y|X=x \sim Bin\left(n-x, \frac{\theta_2}{1-\theta_1}\right)$ $\longrightarrow$ DIM.: fissato $x \in \{0,1,\dots,n\}$

$$p_{Y|X}(y|x) = \frac{p_{X,Y}(x,y)}{p_X(x)} = \frac{\frac{\cancel{n!}}{\cancel{x!}\,y!\,(n-x-y)!} \cdot \cancel{\theta_1^x} \theta_2^y \cdot (1-\theta_1-\theta_2)^{n-x-y}}{\frac{\cancel{n!}}{\cancel{x!}\,(n-x)!} \cdot \cancel{\theta_1^x} (1-\theta_1)^{n-x}} \cdot \frac{(1-\theta_1)^y}{(1-\theta_1)^y}$$

$$= \frac{(n-x)!}{y!\,(n-x-y)!} \cdot \left(\frac{\theta_2}{1-\theta_1}\right)^y \cdot \frac{(1-\theta_1-\theta_2)^{n-x-y}}{(1-\theta_1)^{n-x-y}} =$$

$$= \binom{n-x}{y}\left(\frac{\theta_2}{1-\theta_1}\right)^y \cdot \left(1 - \frac{\theta_2}{1-\theta_1}\right)^{n-x-y}$$

5) $X|Y=y \sim Bin\left(n-y, \frac{\theta_1}{1-\theta_2}\right)$

6) $cov(X,Y) = -n \cdot \theta_1 \cdot \theta_2$

VETTORE VALORI ATTESI $\begin{bmatrix} E(X) \\ E(Y) \end{bmatrix} = \begin{bmatrix} n\theta_1 \\ n\theta_2 \end{bmatrix}$

MATRICE VAR-COVAR $\begin{bmatrix} V(X) & cov(X,Y) \\ cov(X,Y) & V(Y) \end{bmatrix} = \begin{bmatrix} n\theta_1(1-\theta_1) & -n\theta_1\theta_2 \\ -n\theta_1\theta_2 & n\theta_2(1-\theta_2) \end{bmatrix}$

MULTINOMIALE

$\mathbf{X} = \begin{pmatrix} X_1 \\ \vdots \\ X_m \end{pmatrix} \sim MULTIN(n, \theta_1, \theta_2, \dots, \theta_m)$

$R_{\mathbf{X}} = \left\{ \mathbf{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_m \end{pmatrix} : x_i \in \{0,1,\dots,n\} \text{ per } i=1,\dots,m,\ x_1+\dots+x_m \in \{0,1,\dots,n\} \right\}$

$P_{\mathbf{X}}(\mathbf{x}) = \frac{n!}{x_1! \dots x_m! (n-x_1-\dots-x_m)!} \cdot \theta_1^{x_1} \dots \theta_m^{x_m} (1-\theta_1-\dots-\theta_m)^{n-x_1-\dots-x_m} \cdot \mathbb{1}_{R_{\mathbf{X}}}(\mathbf{x})$

QUANDO SI USA? ⇒ n repliche indipendenti dell'esperimento aleatorio che ammette gli esiti:

esiti	con probabilità:
e_1	θ_1
e_2	θ_2
$\vdots$	$\vdots$
e_m	θ_m
e_{m+1}	$1-\theta_1-\dots-\theta_m$

X_1 = nr. repliche con esito e_1
X_2 = nr. repliche con esito e_2
$\vdots$
X_m = nr. repliche con esito e_m
X_{m+1} = nr. repliche con esito e_{m+1}
$X_{m+1} = n - X_1 - \dots - X_m$

* se $m=1 \longrightarrow Bin(n, \theta_1)$
$m=2 \longrightarrow Trin(n, \theta_1, \theta_2)$

PROPRIETÀ:

1) $P_{\mathbf{X}}(\mathbf{x})$ è invariante rispetto a permutazioni delle variabili del vettore aleatorio $\mathbf{X}$
(si ricorda di cambiare ordine delle variabili nella funzione)

2) $\begin{pmatrix} X_1 \\ \vdots \\ X_{m-1} \\ \cancel{X_m} \end{pmatrix} \sim MULTIN(n, \theta_1, \dots, \theta_{m-1})$

3) $\begin{pmatrix} X_i \\ X_j \end{pmatrix} \sim TRIN(n, \theta_i, \theta_j)$

4) $X_i \sim Bin(n, \theta_i)$

5) $E(\mathbf{X}) = \begin{bmatrix} E(X_1) \\ \vdots \\ E(X_m) \end{bmatrix} = \begin{bmatrix} n \cdot \theta_1 \\ \vdots \\ n \cdot \theta_m \end{bmatrix}$

$cov(\mathbf{X}) = \begin{bmatrix} n\theta_1(1-\theta_1) & \dots & -n\theta_1\theta_m \\ \vdots & \ddots & \vdots \\ -n\theta_1\theta_m & \dots & n\theta_m(1-\theta_m) \end{bmatrix}$

NORMALE MULTIVARIATA

$X \sim N_m(\mu, \Sigma) \longrightarrow$ ha distribuzione normale multivariata m-dimensionale non singolare

$R_X = \mathbb{R}^m = \left\{ x = \begin{pmatrix} x_1 \\ \vdots \\ x_m \end{pmatrix} : x_i \in \mathbb{R}, i = 1, \dots, m \right\}$

$f_X(x) = (2\pi)^{-\frac{m}{2}} \cdot (\det \Sigma)^{-\frac{1}{2}} \cdot e^{-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)}$

dove: $\mu = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_m \end{pmatrix}$ $\quad \underset{m \times m}{\Sigma} = \begin{bmatrix} \sigma_1^2 & \dots & \sigma_{1m} \\ \vdots & \ddots & \vdots \\ \sigma_{m1} & \dots & \sigma_m^2 \end{bmatrix}$

* se $m = 1 \longrightarrow$ normale univariata

$\mu \in \mathbb{R}$

$\Sigma \in A$ dove A = insieme delle matrici $m \times m$ ad elementi reali definite positive

PROPRIETÀ:

Prop. 8.4 $E(X) = \mu$

$cov(X) = \Sigma$

Prop. 8.3 $X = \begin{pmatrix} X_1 \\ \vdots \\ X_m \end{pmatrix} \longrightarrow$ ha distribuzione multivariata m-dimensionale non necessariamente non singolare

se $R_X = \mathbb{R}^m$

e la f.g.m. è $M_X(t) = e^{t^T\mu + \frac{1}{2}t^T \Sigma t}$ dove $t = \begin{pmatrix} t_1 \\ \vdots \\ t_m \end{pmatrix} \in \mathbb{R}^m$

dove Σ è semidefinita positiva

Prop. 8.5 $X \sim N_m(\mu, \Sigma) \overset{sse}{\Longleftrightarrow} a^T X \sim N_1(a^T\mu, a^T \Sigma a) \quad \forall a = \begin{pmatrix} a_1 \\ \vdots \\ a_m \end{pmatrix} \in \mathbb{R}^m$

$a \neq \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}$

IMPLICAZIONE: $X \sim N_m(\mu, \Sigma) \Longrightarrow X_i \sim N_1(\mu_i, \sigma_i^2) \quad i = 1, \dots, m$

poiché: $X_i = a^T X$

$a = \begin{pmatrix} 0 \\ 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix} \rightarrow$ i-esima posizione

Prop. 8.6 se $\underset{q \times 1}{Y} = \underset{q \times m}{A} \underset{m \times 1}{X} + \underset{q \times 1}{b}$ allora $Y \sim N_q(A\mu + b, A \Sigma A^T)$

Corollario 8.7 Se $X \sim N_m(\mu, \Sigma)$ allora un qualunque sottovettore di X (di qualunque dimensione) ha ancora distribuzione normale

Prop. 8.8 Sia $W \sim N_m(\mu, \Sigma)$

sia $W = \begin{bmatrix} X \\ Y \end{bmatrix}$ una partizione di W

$\mu = \begin{bmatrix} \mu_X \\ \mu_Y \end{bmatrix} \quad \begin{matrix} \mu_X = E(X) \\ \mu_Y = E(Y) \end{matrix}$

$$\Sigma = \begin{bmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{bmatrix} \qquad \Sigma_{xx} = cov(X) \quad \Sigma_{yy} = cov(Y)$$

sotto queste condizioni:

[X e Y sono indipendenti $\xLeftrightarrow{SSE}$ Σ_{xy} è una matrice nulla] $\rightarrow$ condizione necessaria e sufficiente all'indipendenza tra 2 gruppi di variabili

Corollario 8.10 $X_1, \dots, X_m$ sono MUTUAMENTE INDIPENDENTI $\xLeftrightarrow{SSE}$ Σ è una matrice diagonale

$$\hookrightarrow \Sigma = \begin{bmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_m^2 \end{bmatrix}$$

Prop. 8.11 sia $W \sim N_m(\mu, \Sigma)$

e $W = \begin{bmatrix} X \\ Y \end{bmatrix}$ una partizione di W

$$\mu = \begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix} \quad e \quad \Sigma = \begin{bmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{bmatrix}$$

SE Σ_{xx} è non singolare: (matrice $r \times r$ definita positiva)

ALLORA $$Y|X=x \sim N_{m-r}\Big(\underbrace{\mu_y + \Sigma_{yx}\Sigma_{xx}^{-1}(x-\mu)}_{\substack{\text{valore atteso} \\ E(Y|X=x) \\ \text{regressione di } Y \text{ da } X}},\ \underbrace{\Sigma_{yy} - \Sigma_{yx}\Sigma_{xx}^{-1}\Sigma_{xy}}_{\substack{\text{matrice di var-covar} \\ \text{(indipendente da } x)}}\Big)$$

Proprietà se $\underset{n\times1}{X} = \begin{bmatrix} X_1 \\ \vdots \\ X_n \end{bmatrix} \sim N_n(0, I_n)$

allora $X_1, \dots, X_n$ sono n v.c. mutuamente indipendenti normali standardizzate

* Dal vettore appena descritto calcolo 2 nuove v.c.:

$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \sim N_1(0, \frac{1}{n})$ $\Rightarrow$ la media aritm. è un caso particolare di combinazione lineare: costante additiva $=0$ e costanti moltiplicative sono tutte uguali a $\frac{1}{n}$

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 \sim Gamma\left(\frac{n-1}{2}, \frac{n-1}{2}\right)$$

DIM.:

- $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i^2 + \bar{X}^2 - 2X_i\bar{X}) = \frac{1}{n-1}\left[\sum_{i=1}^{n} X_i^2 + n\bar{X}^2 - 2\bar{X}\underbrace{\sum_{i=1}^{n} X_i}_{n\cdot\bar{X}}\right] =$

$= \frac{1}{n-1}\left[\sum_{i=1}^{n} X_i + \cancel{n\bar{X}^2} - \cancel{2}n\bar{X}^2\right] = \frac{1}{n-1}\left[\sum_{i=1}^{n} X_i^2 - n\bar{X}\right]$

- mi avvalgo della MATRICE DI HELMERT:

$$\underset{n\times n}{Q} = \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 & 0 & \cdots & 0 & 0 \\ \frac{1}{\sqrt{2\cdot3}} & \frac{1}{\sqrt{2\cdot3}} & -\frac{2}{\sqrt{2\cdot3}} & 0 & \cdots & 0 & 0 \\ \vdots & & & & & & \\ \frac{1}{\sqrt{(n-1)n}} & \frac{1}{\sqrt{(n-1)n}} & \frac{1}{\sqrt{(n-1)n}} & & \cdots & \frac{1}{\sqrt{(n-1)n}} & -\frac{(n-1)}{\sqrt{(n-1)n}} \\ \frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}} & & \cdots & \frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}} \end{bmatrix}$$

Q è ORTOGONALE

$Q^T = Q^{-1}$

- Definisco nuovo vettore ricavato dal vettore di partenza:

$$\underset{n\times 1}{Y} = \underset{n\times n}{Q} \cdot \underset{n\times 1}{X} = \begin{bmatrix} Y_1 = \frac{1}{\sqrt{2}}(X_1 - X_2) \\ Y_2 = \frac{1}{\sqrt{2\cdot 3}}(X_1 + X_2 - 2X_3) \\ \vdots \\ Y_n = \frac{1}{\sqrt{n}}(X_1 + X_2 + \dots + X_n) \end{bmatrix}$$

$$Y \sim N_n(\underbrace{Q \cdot 0}_{= 0}, \underbrace{Q\, I_n Q^T}_{Q\cdot Q^T = I_n}) \longrightarrow$$ anche le v.c. all'interno di Y sono indipendenti e normali standardizzate

- $$Y = \begin{bmatrix} Y_1 \\ \vdots \\ Y_{n-1} \\ Y_n \end{bmatrix} \Rightarrow Y^{(1)} \sim N_{n-1}(0, I_{n-1}), \quad \Rightarrow Y_n \sim N_1(0,1)$$

$Y^{(1)}$ e Y_n sono indipendenti, $\longrightarrow$ poiché le covarianze tra Y_n e le v.c nel vettore $Y^{(1)}$ sono uguali a zero

- legame tra $\bar{X}$ e Y_n:

$$\bar{X} = \frac{1}{n}\sum_{i=1}^n X_i = \frac{1}{\sqrt{n}} \cdot Y_n$$

$$Y_n = \frac{1}{\sqrt{n}}\sum_{i=1}^n X_i = \sqrt{n} \cdot \frac{1}{n}\sum_{i=1}^n X_i = \sqrt{n}\,\bar{X}$$

- legame tra S^2 e $Y^{(1)}$:

$$\sum_{i=1}^{n-1} Y_i^2 = \sum_{i=1}^n Y_i^2 - Y_n^2$$
$$= Y^T Y - Y_n^2$$
$$= (QX)^T (QX) - \underbrace{Y_n^2}$$
$$= X^T Q^T Q X - (\sqrt{n}\cdot\bar{X})^2$$
$$= X^T X - n\bar{X}^2$$
$$= \sum_{i=1}^n X_i^2 - n\bar{X}^2 \rightarrow W = \text{DEVIANZA di } X$$

- $W \sim \text{Gamma}\left(\frac{n-1}{2}, \frac{1}{2}\right)$
- $S^2 = \frac{W}{(n-1)} \sim \text{Gamma}\left(\frac{n-1}{2}, \frac{n-1}{2}\right)$

$\bar{X}$ e S^2 sono indipendenti poiché sono trasformazioni di $Y^{(1)}$ e Y_n che sono 2 vettori indipendenti.

SPAZIO CAMPIONARIO S: insieme che contiene tutti i possibili campioni che si possono estrarre dalla popolazione tramite il metodo di campionamento scelto, fissata la dimensione del campione

MODELLO STATISTICO PARAMETRICO: generato dalla considerazione congiunta del modello probabilistico e del metodo di campionamento

$$\begin{cases} \mathcal{Y}_\theta = \{ f_Y(y,\theta) = \theta \in \Theta \subseteq \mathbb{R}^k \} \\ \vec{Y} = (Y_1, \dots, Y_n) \quad n \text{ v.c. provenienti da } f_Y(y,\theta) \in \mathcal{Y}_\theta \end{cases}$$

IDENTIFICABILITÀ DI UN MODELLO STATISTICO:

Un modello statistico parametrico si dice identificato se dato un punto qualunque θ nello spazio parametrico Θ, non esiste in Θ nessun punto osservazionalmente equivalente ad esso.

RIPARAMETRIZZAZIONE MODELLO PROBABILISTICO

Dato il modello probabilistico per Y

$$\mathcal{F}_\theta = \{ f_Y(y,\theta),\ \theta \in \Theta \subseteq \mathbb{R}^k \}$$

se $h(\theta) = \psi$ è una FUNZIONE BIUNIVOCA

allora $\mathcal{F}_\psi = \{ f_Y(y,\psi),\ \psi \in \Psi \subseteq \mathbb{R}^k \}$

dove $\Psi = \{\psi : h(\theta) = \psi,\ \theta \in \Theta\}$ è il "nuovo" spazio parametrico.

FUNZIONE DI VEROSIMIGLIANZA

def. Dato un modello statistico parametrico $(*)$ e osservato il campione $\vec{y} = (y_1, \dots, y_n)$

$(*) \left\{ \begin{array}{l} \mathcal{F}_\theta \\ \vec{y} = (y_1, \dots) \end{array} \right.$

$$L : \Theta \longrightarrow \mathbb{R}^+ \cup \{0\}$$

dato $\theta \in \Theta$

$$L(\theta, \vec{y}) = \underbrace{c(\vec{y})}_{\text{costante al variare di } \theta} \cdot \underbrace{f_{\vec{Y}}(\vec{y}, \theta)}_{\text{distribuzione campionaria di } \vec{Y} \text{ valutata in } \vec{y}}$$

* KERNEL di $L(\theta, \vec{y})$ = parte di $L(\theta, \vec{y})$ che dipende da θ

significato: $L(\theta, \vec{y})$ per $\theta \in \Theta$ è una misura dell'accordo tra $\vec{y}$ e ciascuno dei valori di $\theta \in \Theta$

FUNZIONE DI LOGVEROSIMIGLIANZA:

$$\ell : \Theta \longrightarrow \mathbb{R}$$

v. esempi

$\theta \in \Theta$

$$\ell(\theta, \vec{y}) = \ln(L(\theta, \vec{y})) \quad \text{per } \theta \in \Theta$$

STATISTICA:

def: Dato un modello statistico parametrico dove:

Θ = spazio parametrico

S = spazio campionario

si definisce statistica una qualunque funzione

$$T : S \longrightarrow \mathbb{R}^r$$

tale che

$\forall \vec{y} \in S \quad T(\vec{y})$ sia una funzione di $\vec{y}$ indipendente da θ

- se $r=1$ → funzione scalare
  $r>1$ → funzione vettoriale

$T(\vec{y})$: vettore di r v.c. ognuna delle quali è data da una diversa funzione scalare delle n v.c. campionarie

$T(\vec{y})=t$ → determinazione di $T(\vec{y})$ calcolata per un certo valore $\vec{y}$

utilità: poichè le statistiche utili al nostro scopo sono statistiche per cui $r<n$, la statistica T produce una sintesi delle n v.c. campionarie

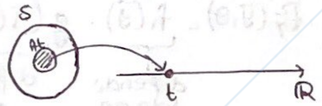
PROPRIETÀ:

1. ci possono essere statistiche differenti che hanno lo stesso supporto, e in comune la partizione dello spazio parametrico

2. Data t:

$A_t = \{\vec{y}=(y_1,\dots,y_n) \in S \text{ t.c. } T(\vec{y})=t\}$

antiimmagine di t



Poichè $T: S \longrightarrow \mathbb{R}^r$

3. Una qualunque trasformazione biunivoca di una statistica NON modifica la partizione dello spazio campionario indotta dai suoi valori campionari

4. trasformazione biunivoca di una statistica restituisce statistica con medesime caratteristiche (es. sufficiente minimale) della statistica di partenza

Def. particolare categoria di statistica:

Dato un modello statistico
$$\begin{cases} \mathcal{Y}_\theta = \{f_Y(y,\theta),\ \theta \in \Theta \subseteq \mathbb{R}^k\} \\ \vec{Y}=(Y_1,\dots,Y_n) \ \text{i.i.d.} \end{cases}$$

Definisco momento campionario di ordine h:
$$T(\vec{Y}) = \frac{1}{n}\sum_{i=1}^{n} Y_i^{\,h}$$

se $h=1$ → media aritmetica campionaria

⇒ tutti i momenti campionari sono esempi di statistiche con $r=1$

STATISTICA SUFFICIENTE

Dato un modello statistico

$$\begin{cases} \mathcal{F}_\theta = \{ f_Y(y,\theta),\ \theta \in \Theta \subseteq \mathbb{R}^k \} \\ \vec{Y} = (Y_1, \dots, Y_n) \ \text{i.i.d.} \end{cases}$$

$T(\vec{Y})$ è detta statistica sufficiente per θ

se per ogni $\vec{y} \in S$
$\vec{z} \in S$

$$T(\vec{y}) = T(\vec{z}) \implies L(\theta, \vec{y}) \underset{\substack{\text{uguali a} \\ \text{meno di una} \\ \text{costante}}}{\propto} L(\theta, \vec{z})$$

TEOREMA DI FATTORIZZAZIONE DI NEYMAN (→ ci dice se stat. è suff.)

Dato un modello statistico
una statistica $T(\vec{Y})$ è sufficiente per il parametro del modello

$$\xleftrightarrow{\text{SSE}} \quad f_{\vec{Y}}(\vec{y}, \theta) = \underbrace{h(\vec{y})}_{\substack{\text{dipende} \\ \text{solo da } \vec{y}}} \cdot \underbrace{g(T(\vec{y}); \theta)}_{\substack{\text{dipende da } T(\vec{y}) \\ \text{e } \theta}}$$

TEOREMA: (→ ci dice se stat. è suff.)

Dato un modello statistico

$$\begin{cases} \mathcal{F}_\theta = \{ f_Y(y,\theta),\ \theta \in \Theta \subseteq \mathbb{R}^k \} \\ \vec{Y} = (Y_1, \dots, Y_n) \ \text{i.i.d.} \end{cases}$$

Una statistica $T(\vec{Y})$ è sufficiente per θ $\xleftrightarrow{\text{SSE}}$ $f_{\vec{Y}|T(\vec{y})=t}(\vec{y}|t)$ NON dipende da θ

STATISTICA SUFFICIENTE MINIMALE:

OB: fare in modo che la sintesi avvenga nella massima misura possibile (r deve essere il più piccolo possibile)

Def. Dato un modello statistico

$$\begin{cases} \mathcal{F}_\theta = \{ f_Y(y,\theta),\ \theta \in \Theta \subseteq \mathbb{R}^k \} \\ \vec{Y} = (Y_1, \dots, Y_n) \ \text{i.i.d.} \end{cases}$$

$T(\vec{Y})$ è detta statistica sufficiente minimale per θ

se presi $\vec{y} \in S$
$\vec{z} \in S$

$$T(\vec{y}) = T(\vec{z}) \iff L(\theta, \vec{y}) \propto L(\theta, \vec{z}) \quad \forall \theta \in \Theta$$

RIASSUMENDO: Data una statistica sufficiente $T(\vec{y})$

presi 2 campioni: $\vec{y} = (y_1, \dots, y_n)$

$\vec{z} = (z_1, \dots, z_n)$

supponiamo siano tali che:

$$L(\theta, \vec{y}) \propto L(\theta, \vec{z})$$

ovvero tali che

$$L(\theta, T(\vec{y})) \propto L(\theta, T(\vec{z}))$$

ovvero tali che

$$\frac{L(\theta, T(\vec{y}))}{L(\theta, T(\vec{z}))} \text{ è indipendente da } \theta$$

verifichiamo:

se $T(\vec{y}) = T(\vec{z}) \implies T$ è stat. sufficiente minimale per θ

o se $T(\vec{y}) \neq T(\vec{z}) \implies T$ è stat. sufficiente ma non minimale per θ.

TEOREMA: (metodo alternativo per dire se stat. è sufficiente minimale)

Data una classe parametrica che costituisce una famiglia esponenziale ed è scritta in forma ridotta

Allora l'insieme $\{t_1(y), \dots, t_r(y)\}$ è una statistica sufficiente minimale per il parametro θ del modello.

FAMIGLIA ESPONENZIALE

def. Una classe parametrica

$$\mathcal{Y}_\theta = \{f_y(y;\theta), \theta \in \Theta \subseteq \mathbb{R}^k\}$$

costituisce una famiglia esponenziale se

$$f_y(y;\theta) = q(y) \cdot e^{\sum_{i=1}^{r} \psi_i(\theta) \cdot t_i(y) - \tau(\theta)}$$

dove: $q(y), t_1(y), \dots, t_r(y)$ sono funzioni di y indipendenti da θ

$\psi_1(\theta), \dots, \psi_r(\theta), \tau(\theta)$ sono funzioni di θ indipendenti da y

r è numero intero positivo

* $q(y) \cdot e^{\sum_{i=1}^{r} \psi_i(\theta) t_i(y) - \tau(\theta)}$ sarà pari a zero quando $q(y) = 0$

Poichè ciò che sta a destra di $f_Y(y,\theta)$ è $=0$ solo quando $q(y)=0$, le distribuzioni il cui supporto dipende dal parametro non saranno una famiglia esponenziale.

PROPRIETÀ:

1) Proprietà di riproducibilità di una famiglia esponenziale rispetto al CCS

Data una classe parametrica

$\mathcal{Y}_\theta = \{f_Y(y,\theta), \theta \in \Theta \subseteq \mathbb{R}^k\}$ che costituisce una famiglia esponenziale

allora $\{f_{\vec{Y}}(\vec{y},\theta), \theta \in \Theta \subseteq \mathbb{R}^k\}$ è ancora una famiglia esponenziale

se I.I.D $\Rightarrow$ cioè se metodo di campionamento è casuale semplice

2) proprietà matematica per invertire l'ordine dell'integrale e della derivata

Se $f_{\vec{Y}}(\vec{y},\theta)$ appartiene ad una fam. exp.
data $g(\vec{y})$ funzione di $\vec{y}$ indipendente da θ
tale che $\int_{\vec{y}\in S} g(\vec{y}) \cdot f_{\vec{Y}}(\vec{y},\theta)\, d\vec{y} < +\infty$

allora è possibile $\frac{\partial}{\partial\theta}\int_{\vec{y}\in S} g(\vec{y})\cdot f_{\vec{Y}}(\vec{y},\theta)\, d\vec{y} = \int_{\vec{y}\in S} \frac{\partial}{\partial\theta} g(\vec{y})\cdot f_{\vec{Y}}(\vec{y},\theta)\, d\vec{y}$

Se le funzioni $\{1, \psi_1(\theta), \dots, \psi_r(\theta)\}$ sono linearmente indipendenti in θ e le funzioni $\{1, t_1(y), \dots, t_r(y)\}$ sono linearmente indipendenti in y allora la famiglia esponenziale è scritta in forma ridotta e r è detto ordine della famiglia esponenziale

$\rightarrow$ anche quando $r=1$ la fam. exp. è scritta in forma ridotta

CONDIZIONI DI REGOLARITÀ DI UNA FAM. ESPONENZIALE

1) $\Theta = \{\theta$ tali che $\int_{\vec{y}\in S} f_{\vec{Y}}(\vec{y},\theta)\, d\vec{y} < +\infty\}$ e Θ così definito è intervallo aperto in $\mathbb{R}^k$

2) $r=k$ $\rightarrow$ ordine fam. exp $=$ dimensione spazio parametrico $\Theta \subseteq \mathbb{R}^k$

3) $\{\psi_1(\theta), \dots, \psi_r(\theta)\}$ è insieme di funzioni invertibili

4) le funzioni $\psi_1(\theta), \dots, \psi_r(\theta)$ ammettono derivate di ogni ordine rispetto a θ

Se 4 condizioni sono soddisfatte

$\Rightarrow$ FAMIGLIA ESPONENZIALE REGOLARE

STIMA PARAMETRICA

Dato un modello statistico parametrico

$$\begin{cases} \mathcal{Y}_\theta = \{ f_Y(y,\theta),\ \theta \in \Theta \subseteq \mathbb{R}^k \} \\ \vec{Y} = (Y_1, \dots, Y_n) \end{cases}$$ n-pla di v.c. campionarie provenienti da $f_Y(y, \underset{\text{NON NOTO}}{\theta_0}) \in \mathcal{Y}_\theta$

dato $\vec{y} = (y_1, \dots, y_n) \in S$

vogliamo stimare il valore incognito di θ_0 del parametro θ del modello statistico dato $\vec{y}$

OB. individuare, a partire dal campione estratto, per il valore incognito del parametro, un valore $\in \Theta$, che giustifichi il fatto che il campione estratto è quello e non un altro campione

STIMATORE DEL VALORE INCOGNITO DI UN PARAMETRO

$$T: S \longrightarrow \Theta \subseteq \mathbb{R}^k$$

indipendente da θ

* uno stimatore è sicuramente una statistica (codominio stat. è sottoinsieme di $\mathbb{R}^k$)
* una statistica può non essere uno stimatore (se ad es. statistica assume valori che non rientrano in Θ, non è uno stimatore)

STIMA DI MASSIMA VEROSIMIGLIANZA

def. Dato un modello statistico parametrico
nota la funzione di verosimiglianza $L(\theta, \vec{y})$, $\theta \in \Theta$
si definisce stima di massima verosimiglianza di θ

$$\hat{\theta} = \arg\sup_{\theta \in \Theta} L(\theta, \vec{y}) \qquad * \text{ oppure } \quad \hat{\theta} = \arg\sup_{\theta \in \Theta} \ell(\theta, \vec{y})$$

PROPRIETÀ $\hat{\theta}$:

① TEO.
se $T(\vec{Y})$ è una statistica sufficiente per θ
allora $\hat{\theta}$ è una funzione di $T(\vec{y})$

② PROPRIETÀ DI EQUIVARIANZA DI $\hat{\theta}$ RISPETTO A TRASFORMAZIONI BIUNIVOCHE

Dato il modello statistico $\mathcal{Y}_\theta = \{ f_Y(y,\theta), \theta \in \Theta \subseteq \mathbb{R}^k \}$

$$\hat{\theta} = \arg\sup_{\theta \in \Theta} L(\theta, \vec{y})$$

se $h: \Theta \longrightarrow \Psi$ è funzione biunivoca

allora $\mathcal{Y}_\psi = \{ f_Y(y, \psi),\ \psi = h(\theta),\ \psi \in \Psi \}$ è una riparametrizzazione di $\mathcal{Y}_\theta$

e $\hat{\psi} = h(\hat{\theta})$

trovo $\hat{\theta}$ con $\ell(\theta, \vec{y})$:

$$\boxed{\frac{\partial \ell(\theta,\vec{y})}{\partial \theta} = 0} \rightarrow$$ EQUAZIONE DI VEROSIMIGLIANZA (risolta rispetto a θ)

- se $K=1$: singola equazione
- se $K>1$: sistema di K equazioni nelle incognite $\theta_1, \dots, \theta_K$

$$\theta = \begin{pmatrix} \theta_1 \\ \vdots \\ \theta_K \end{pmatrix} \qquad \begin{cases} \frac{\partial \ell(\theta,\vec{y})}{\partial \theta_1} = 0 \\ \vdots \\ \frac{\partial \ell(\theta,\vec{y})}{\partial \theta_K} = 0 \end{cases}$$

* se classe parametrica è fam. exp. allora $\hat{\theta}$ è la soluzione all'equazione di verosimiglianza

* se calcoliamo derivata seconda di $\ell(\theta,\vec{y})$ in corrispondenza del p.to di massimo otteniamo un numero negativo.

se $K=1$ $\quad \left.\frac{\partial^2 \ell(\theta,\vec{y})}{\partial \theta^2}\right|_{\theta=\hat{\theta}} < 0$

se $K>1$ $\quad \theta = \begin{pmatrix} \theta_1 \\ \vdots \\ \theta_K \end{pmatrix} \rightarrow$ MATRICE HESSIANA $= \underset{K \times K}{H(\theta)} = \begin{bmatrix} \frac{\partial^2 \ell(\theta,\vec{y})}{\partial \theta_1 \partial \theta_1} & \dots & \frac{\partial^2 \ell(\theta,\vec{y})}{\partial \theta_1 \partial \theta_K} \\ \vdots & & \vdots \\ \frac{\partial^2 \ell(\theta,\vec{y})}{\partial \theta_K \partial \theta_1} & \dots & \frac{\partial^2 \ell(\theta,\vec{y})}{\partial \theta_K \partial \theta_K} \end{bmatrix} = \frac{\partial^2 \ell(\theta,\vec{y})}{\partial \theta \partial \theta^T}$

dove θ vettore $K \times 1$

$H(\theta)$ calcolata per $\theta = \hat{\theta}$

$\Rightarrow$ matrice definita negativa

INFORMAZIONE OSSERVATA DI FISHER

$\rightarrow$ è una misura della preferenza relativa di $\hat{\theta}$ rispetto ad altri valori del parametro

$IO(\hat{\theta}) = -H(\hat{\theta})$

se $K=1$: è uno scalare $\quad IO(\hat{\theta}) = -\left(\left.\frac{\partial^2 \ell(\theta,\vec{y})}{\partial \theta^2}\right|_{\theta=\hat{\theta}}\right)$

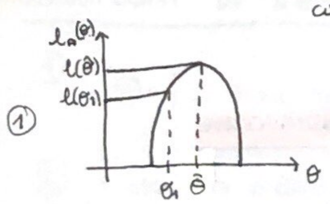
$K=1 \rightarrow IO(\hat{\theta})$ è uno scalare >0

$\downarrow$

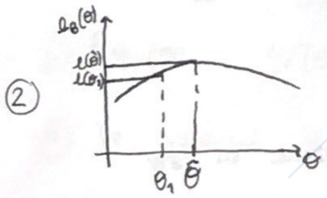
$\forall \theta \in$ ad un intorno di $\hat{\theta}$

$\ell(\theta) = \ell(\hat{\theta}) + \frac{1}{2}\ell''(\hat{\theta})(\theta - \hat{\theta})^2 + \dots$

$\ell(\theta) - \ell(\hat{\theta}) \cong \frac{1}{2}\underbrace{\ell''(\hat{\theta})}_{\text{esprime la curvatura locale}}(\theta-\hat{\theta})^2 < 0$



Nella situazione ① appena ci si discosta dal p.to di massimo $\hat{\theta}$, la variazione che la f. di logverosimiglianza ha subito è molto grande; il divario tra $\ell(\hat{\theta})$ e $\ell(\theta_1)$ è grande perché $\ell(\theta)$ ha una curvatura locale molto forte e decade velocemente se ci si sposta da $\hat{\theta}$



Nella situazione ② il divario è più piccolo.

$\Rightarrow$ Il divario dipende dalla curvatura locale di $\ell(\theta)$ in corrispondenza del p.to di massimo, cioè $\ell''(\theta)$

$\Rightarrow$ la preferenza da attribuire a $\hat{\theta}$ è maggiore nella situazione ①

- PROPRIETÀ DI CORRETTEZZA : distribuzione di probabilità dello stimatore è centrata sul valore incognito del parametro

$$E[T(\vec{Y})] = \theta_0 \implies T(\vec{Y}) \text{ è NON DISTORTO/CORRETTO per } \theta_0$$

$$E[T(\vec{Y})] \neq \theta_0 \implies T(\vec{Y}) \text{ è DISTORTO per } \theta_0$$

↓

→ DISTORSIONE DI $T(\vec{Y})$: $E[T(\vec{Y})] - \theta_0$

- PROPRIETÀ EFFICIENZA
- PROPRIETÀ DI CORRETTEZZA ASINTOTICA : (per stimatori distorti)

$$\lim_{n\to+\infty} E[T(\vec{Y})] = \theta_0 \implies T(\vec{Y}) \text{ è asintoticamente corretto per } \theta_0$$

- PROPRIETÀ DI CONSISTENZA DEBOLE

$$\boxed{\lim_{n\to+\infty} P[|T(\vec{Y}) - \theta_0| < \varepsilon] = 1 \quad \forall \varepsilon > 0} \implies T(\vec{Y}) \text{ è debolmente consistente}$$

$$T(\vec{Y}) \xrightarrow{P} \theta_0$$

$$\vec{Y} = (Y_1, \dots, Y_n) \quad n \to +\infty$$

- PROPRIETÀ DI CONSISTENZA FORTE

$$\boxed{P\left[\lim_{n\to+\infty} |T(\vec{Y}) - \theta_0| = 0\right] = 1} \implies T(\vec{Y}) \text{ è fortemente consistente}$$

$$T(\vec{Y}) \xrightarrow{q.c.} \theta_0$$

poiché converg. in q.c. implica convergenza in prob. ⇒ uno stimatore fortem. consistente sarà anche debolmente consistente

* $\lim_{n\to+\infty} E[(T(\vec{Y}) - \theta_0)^2] = 0 \implies T(\vec{Y}) \xrightarrow{M.q.} \theta_0 \implies T(\vec{Y}) \xrightarrow{P} \theta_0$

* $\left.\begin{aligned} &\lim_{n\to\infty} E[T(\vec{Y})] = \theta_0 \\ &\lim_{n\to\infty} V[T(\vec{Y})] = 0 \end{aligned}\right\}$ se sono soddisfatte $T(\vec{Y})$ è DEBOLMENTE CONSISTENTE } condizione sufficiente

TEOREMA

SE
- $K=1$
- problema regolare di stima
- $\vec{Y} = (Y_1, \dots, Y_n)$ I.I.D.
- $i(\theta_0) = IA(\theta_0, Y_i)$ $i = 1, \dots, n$ è positiva e finita
- $\hat{\theta}$ è fortemente consistente
- $\exists$ una funzione $M(y,\theta)$ tale che $\left|\frac{\partial^3}{\partial\theta^3} \log f(y,\theta)\right| < M(y,\theta)$ e $E(M(Y,\theta)) < \infty$

ALLORA $\hat{\theta} \xrightarrow{d} N(\theta_0, (n \cdot i(\theta_0))^{-1})$

E
- $\hat{\theta}$ ha distribuzione asintotica normale
- $\hat{\theta}$ è asintoticamente corretto
- $\hat{\theta}$ è asintoticamente pienamente efficiente (perché varianza della distribuzione a cui converge è uguale al limite inferiore di Rao-Cramer.)

CONDIZIONI DI REGOLARITÀ DI UN PROBLEMA DI STIMA → se sono soddisfatte, stiamo affrontando un problema regolare di stima

1. Il modello statistico è identificabile (v. identificabilità modello stat.)
2. Θ è intervallo aperto dello spazio euclideo $\mathbb{R}^k$
3. gli elementi della classe parametrica devono avere tutti lo stesso supporto (R_y non dipende da θ)
4. $$\int_{\vec{y}\in S} \frac{\partial}{\partial\theta} f_{\vec{Y}}(\vec{y},\theta)\, d\vec{y} = \frac{\partial}{\partial\theta}\int_{\vec{y}\in S} f_{\vec{Y}}(\vec{y},\theta)\, d\vec{y}$$
$$\int_{\vec{y}\in S} \frac{\partial^2}{\partial\theta\,\partial\theta^T} f_{\vec{Y}}(\vec{y},\theta)\, d\vec{y} = \frac{\partial^2}{\partial\theta\,\partial\theta^T}\int_{\vec{y}\in S} f_{\vec{Y}}(\vec{y},\theta)\, d\vec{y}$$

FUNZIONE PUNTEGGIO DI FISHER / FUNZ. SCORE

$u: \Theta \longrightarrow \mathbb{R}^k$ $\quad$ k = nr. parametri del modello esaminato

$$u(\theta,\vec{y}) = \frac{\partial \ell(\theta,\vec{y})}{\partial\theta} \qquad \theta\in\Theta$$

se $k>1$: $u(\theta,\vec{y}) = \begin{bmatrix} \frac{\partial \ell(\theta,\vec{y})}{\partial\theta_1} \\ \vdots \\ \frac{\partial \ell(\theta,\vec{y})}{\partial\theta_k} \end{bmatrix}$ $\quad$ se $k=1$ è una funzione sola

PROPRIETÀ: (nell'ottica del campionamento ripetuto)

$u(\theta;\vec{Y})$

vettore v.c. campionarie (prima di estrarre campione $\vec{y}$)

1. $$E\left[u(\theta,\vec{Y})\right] = E\left[\frac{\partial}{\partial\theta}\ell(\theta,\vec{Y})\right] = E\left[\frac{\partial}{\partial\theta}\log f_{\vec{Y}}(\vec{y},\theta)\right] =$$
$$= \int_{\vec{y}\in S}\left[\frac{\partial}{\partial\theta}\log f_{\vec{Y}}(\theta,\vec{y})\right]\cdot f_{\vec{Y}}(\vec{y},\theta)\, d\vec{y} =$$
$$= \int_{\vec{y}\in S}\left[\frac{1}{\cancel{f_{\vec{Y}}(\vec{y},\theta)}}\cdot\frac{\partial f_{\vec{Y}}(\vec{y},\theta)}{\partial\theta}\right]\cancel{f_{\vec{Y}}(\vec{y},\theta)}\, d\vec{y} =$$
$$= \int_{\vec{y}\in S}\frac{\partial f_{\vec{Y}}(\vec{y},\theta)}{\partial\theta}\, d\vec{y} =$$
* se siamo in problema regolare di stima posso scambiare integrale e derivata
$$= \frac{\partial}{\partial\theta}\underbrace{\int_{\vec{y}\in S} f_{\vec{Y}}(\vec{y},\theta)\, d\vec{y}}_{=1} =$$
$$= \frac{\partial}{\partial\theta}1 = 0 \longrightarrow \text{vettore di lunghezza } k$$

② $V\left[u(\theta,\vec{Y})\right] = E\left[\underset{K\times 1}{u(\theta,\vec{Y})}\cdot \underset{1\times K}{u(\theta,\vec{Y})^T}\right]$

se siamo in problema regolare di stima

$$= E\begin{bmatrix} \frac{\partial \ell(\theta)}{\partial \theta_1}\cdot\frac{\partial \ell(\theta)}{\partial \theta_1} & \cdots & \frac{\partial \ell(\theta)}{\partial \theta_1}\cdot\frac{\partial \ell(\theta)}{\partial \theta_K} \\ \vdots & & \vdots \\ \frac{\partial \ell(\theta)}{\partial \theta_K}\cdot\frac{\partial \ell(\theta)}{\partial \theta_1} & \cdots & \frac{\partial \ell(\theta)}{\partial \theta_K}\cdot\frac{\partial \ell(\theta)}{\partial \theta_K} \end{bmatrix}$$

!! $E\left[\frac{\partial \ell(\theta)}{\partial \theta}\cdot\left(\frac{\partial \ell(\theta)}{\partial \theta}\right)^T\right] = -E\left[\frac{\partial^2 \ell(\theta)}{\partial\theta\,\partial\theta^T}\right] = \underbrace{-E\left[H(\theta)\right]}_{V\left[u(\theta,\vec{Y})\right]}$

se siamo in problema regolare di stima

INFORMAZIONE ATTESA DI FISHER ⇒ misura dell'informazione sul parametro incognito θ contenuta in $\boxed{\text{media}}$ in un qualunque campione

$IA(\theta) = V\left[u(\theta,\vec{Y})\right]$

Ⓥⓢ $IO(\hat{\theta})$: misura dell'informazione sul parametro θ contenuta nello specifico campione utilizzato per il calcolo di $\hat{\theta}$

$= E\left[u(\theta)\cdot u(\theta)^T\right]$ $\qquad = -E\left[H(\theta)\right] = -E\left[\frac{\partial u(\theta,\vec{Y})}{\partial\theta}\right]$

PROPRIETÀ: per variabili indipendenti

$\vec{Y} = (Y_1,\dots,Y_n)$ n v.c. i.i.d.

$IA(\theta, Y_i) = i(\theta) \quad i=1,\dots,n$ → tutte le v.c. contengono la stessa informazione attesa sul parametro perché sono identicamente distribuite

Ed essendo indipendenti:

$IA(\theta,\vec{Y}) = i(\theta)+\dots+i(\theta) = n\cdot i(\theta)$

TEOREMA che fissa il limite inferiore per la varianza di $T(\vec{Y})$ di θ in un prob. regolare di stima

SE
- siamo nell'ambito di un problema regolare di stima
- $K=1$

sia $T(\vec{Y})$ stimatore di θ tale che: $E\left[T(\vec{Y})\right] = a(\theta)$ esiste e sia derivabile rispetto a θ

si ha $\boxed{V\left[T(\vec{Y})\right] \geq \boxed{\frac{\left[a'(\theta)\right]^2}{IA(\theta)}}}$ → disuguaglianza di Rao-Cramer

limite inferiore di Rao-Cramer

dove: $a'(\theta) = \frac{\partial}{\partial\theta}E\left[T(\vec{Y})\right]$

e $0 < IA(\theta) < +\infty$

* SE $V(T(\vec{Y})) = \frac{\left[a'(\theta)\right]^2}{IA(\theta)}$

ALLORA $T(\vec{Y})$ è PIENAMENTE EFFICIENTE

* SE $IA(\theta) = \frac{1}{V(T(\vec{Y}))}$

$IA(\theta)$ è misura della precisione massima della stima attesa del parametro incognito a partire da uno stimatore corretto

ERRORE QUADRATICO MEDIO DI $T(\vec{y})$:
- misura per valutare qualità stimatore
- è un indicatore della distanza attesa tra lo stimatore e il parametro incognito

$$EQM\left[T(\vec{y})\right] = E\left[\left(T(\vec{y}) - \theta_0\right)^2\right]$$

* Poiché $V[T(\vec{y})] = E\left[T(\vec{y}) - E[T(\vec{y})]^2\right]$

se $\underbrace{E[T(\vec{y})] = \theta}_{\text{stimatore è corretto}} \implies EQM[T(\vec{y})] = V[T(\vec{y})]$

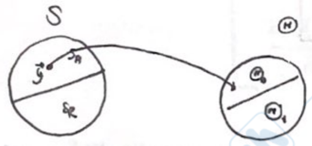
→ È una misura più appropriata dell'accuratezza delle stime ottenute da uno stimatore distorto

→ se stimatore è corretto : EQM coincide con varianza
↓
è misura più accurata per valutare accuratezza dello stimatore

$$EQM[T(\vec{y})] = V[T(\vec{y})] + [\underbrace{E(T(\vec{y})) - \theta}_{\text{distorsione di } T(\vec{y})}]^2$$

TEST STATISTICO : è una funzione

$$T: S \longrightarrow \{\Theta_0, \Theta_1\}$$



$S_A = \{\vec{y} \in S$ tali che $T(\vec{y}) = \Theta_0\}$ regione di accettazione di Ho

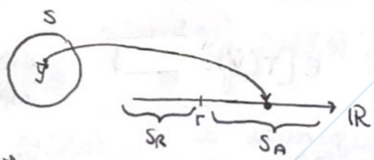
$S_R = \{\vec{y} \in S$ t.c. $T(\vec{y}) = \Theta_1\}$ regione di rifiuto di Ho

FUNZIONE TEST : strumento che serve per rendere operativa la def. di test
(statistica test) statistico

$$T: S \longrightarrow \mathbb{R} \qquad r = \text{valore critico}$$

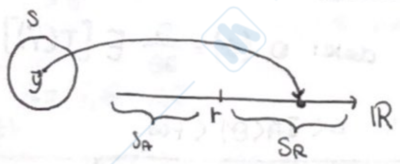
2 SITUAZIONI :

1) f. test $T(\vec{y})$ è una misura DIRETTA dell'accordo tra $\vec{y}$ e Ho :



$S_A = \{\vec{y} \in S$ t.c. $T(\vec{y}) \geq r\}$
$S_R = \{\vec{y} \in S$ t.c. $T(\vec{y}) < r\}$

2) f. test $T(\vec{y})$ è una misura INVERSA dell'accordo tra $\vec{y}$ e Ho :



$S_A = \{\vec{y} \in S$ t.c. $T(\vec{y}) \leq r\}$
$S_R = \{\vec{y} \in S$ t.c. $T(\vec{y}) > r\}$

Dato $\vec{y} \in S$

tipi di decisioni	$\theta_0 \in \Theta_0$	$\theta_0 \in \Theta_1$
$\vec{y} \in S_A$	CORRETTA	$\boxed{\text{ERRATA}}$
$\vec{y} \in S_R$	$\boxed{\text{ERRATA}}$	CORRETTA

ERRATA → errore II° tipo: accetto Ho quando è falsa (meno grave) dal p.to di vista economico ⇒ MANCATO ALLARME

ERRATA → errore di I° tipo
rifiuto Ho quando è vera
⇒ FALSO ALLARME (più grave dal p.to di vista economico)

FUNZIONE POTENZA DI UN TEST STATISTICO: bene per poter calcolare, dato un test statistico, la probabilità di commettere i 2 errori e di conseguenza cercare di individuare test statistici tali da garantirci di minimizzare la probabilità che questi 2 errori siano commessi.

$$\underset{\text{TEST}}{\gamma_T} : \Theta \longrightarrow [0,1]$$

$$\forall \theta \in \Theta \quad \gamma_T(\theta) = P\left[T(\vec{y}) = \Theta_1, \theta\right] = P\left[\vec{y} \in S_R, \theta\right]$$

POTENZA DI UN TEST:
$$\inf_{\theta \in \Theta_1} \gamma_T(\theta)$$

$\alpha_T = \gamma_T(\theta)$, $\theta \in \Theta_0$ → ampiezza errore I° tipo

$\beta_T = 1 - \gamma_T(\theta)$, $\theta \in \Theta_1$ → ampiezza errore II° tipo

LIVELLO DI UN TEST: probabilità massima di commettere errore di 1° tipo

$$\alpha_T = \sup_{\theta \in \Theta_0} \gamma_T(\theta)$$

LEMMA FONDAMENTALE DI NEYMAN-PEARSON

Dato un modello statistico

$$\begin{cases} \mathcal{F}_\Theta = \{f_Y(y,\theta),\ \theta \in \Theta = \{\Theta_H, \Theta_A\}\} \\ \vec{Y} = (Y_1, \dots, Y_n) \quad n \text{ v.c. i.i.D.} \quad \text{provenienti da } f_Y(y,\theta) \in \mathcal{F}_\Theta \end{cases}$$

- formulato il sistema di ipotesi

$$\begin{cases} H_0 : \theta_0 = \theta_H \\ H_1 : \theta_0 = \theta_A \end{cases}$$

- fissato un valore α (⇒ max probabilità di commettere errore di primo tipo)

Il test che ha potenza più alta per saggiare Ho contro H1 fra tutti i test di livello non superiore ad α, è quello che ha la seguente regione di rifiuto:

$$S_R = \{\vec{y} \in S \ \text{ t.c. } \lambda^*(\vec{y}) \le \lambda_\alpha\}$$

dove $\lambda^*(\vec{y}) = \dfrac{L(\theta_H, \vec{y})}{L(\theta_A, \vec{y})}$ → FUNZIONE TEST

λ_α = valore critico, scelto in modo tale che $P\left[\lambda^*(\vec{y}) \le \lambda_\alpha ; \theta_0 = \theta_H\right] = \alpha$

TEST DEL RAPPORTO DI VEROSIMIGLIANZA

1) H_0 e H_1 SEMPLICI :

$$\begin{cases} H_0 : \theta_0 = \theta_H \\ H_1 : \theta_0 = \theta_A \end{cases} \qquad \theta \in \Theta = \{\theta_H, \theta_A\} \qquad \lambda^*(\vec{y}) : S \longrightarrow \mathbb{R}^+ \cup \{0\}$$

$$\lambda^*(\vec{y}) = \frac{L(\theta_H, \vec{y})}{L(\theta_A, \vec{y})}$$

$\lambda^*(\vec{y}) \Rightarrow$ misura diretta dell'accordo tra $\vec{y}$ e H_0: quanto più è grande il numeratore, maggiore sarà l'accordo tra $\vec{y}$ e H_0.

$$S_R = \{\vec{y} \in S \text{ t.c. } \lambda^*(\vec{y}) \le \lambda_\alpha\}$$

λ_α valore critico associato alla f. test

↳ t.c. $P[\lambda^*(\vec{y}) \le \lambda_\alpha , \theta_0 = \theta_H] = \alpha$

2) H_0 SEMPLICE, H_1 COMPOSTA

$$\begin{cases} H_0 : \theta_0 = \theta_H \\ H_1 : \theta_0 \ne \theta_H \end{cases} \qquad \Theta_0 = \{\theta_H\} \qquad \Theta_1 = \Theta \setminus \{\theta_H\} \qquad \lambda(\vec{y}) : S \longrightarrow [0,1]$$

$$\lambda(\vec{y}) = \frac{L(\theta_H, \vec{y})}{L(\underbrace{\hat{\theta}}_{\text{stimatore MVS}}, \vec{y})}$$

$\lambda(\vec{y}) \Rightarrow$ misura diretta dell'accordo tra $\vec{y}$ e H_0: quanto più è grande il denominatore, maggiore sarà l'accordo tra $\vec{y}$ e H_0

$$S_R = \{\vec{y} \in S \text{ t.c. } \lambda(\vec{y}) \le \lambda_\alpha\}$$

dove $\lambda_\alpha \to$ valore critico associato alla f. test

↳ t.c. $P[\lambda(\vec{y}) \le \lambda_\alpha , \theta_0 = \theta_H] = \alpha$

* FORMULAZIONE EQUIVALENTE:

$$S_R = \{\vec{y} \in S \text{ t.c. } W(\vec{y}) \ge W_\alpha\}$$

dove $W(\vec{y}) = -2\ln(\lambda(\vec{y})) \longrightarrow$ F. TEST $\Rightarrow$ misura inversa dell'accordo tra $\vec{y}$ e H_0

$W_\alpha = -2\ln(\lambda_\alpha) \longrightarrow$ valore critico

* $W(\vec{y}) = 2(\ell(\hat{\theta}, \vec{y}) - \ell(\theta_H, \vec{y}))$

e $W(\vec{y}) \xrightarrow{d} \chi^2_{(\nu=1)}$ ($\nu = k-1$) per n grande

$W_\alpha = 3{,}84$ per $\alpha = 0{,}05$

$= 6{,}63$ per $\alpha = 0{,}01$

3) H_0 e H_1 COMPOSTE

$$\begin{cases} H_0 : \theta_0 \in \Theta_0 \\ H_1 : \theta_0 \in \Theta_1 \end{cases}$$

Θ_0 e Θ_1 contengono almeno 2 valori per θ_0

$\lambda(\vec{y}) \Rightarrow$ misura diretta del massimo accordo tra $\vec{y}$ e H_0

$$\lambda(\vec{y}) = \frac{L(\hat{\theta}_0, \vec{y})}{L(\hat{\theta}, \vec{y})}$$

dove: $\hat{\theta} = \arg\sup_{\theta \in \Theta} L(\theta, \vec{y})$

$\hat{\theta}_0 = \arg\sup_{\theta \in \Theta_0} L(\theta, \vec{y})$

Valore critico : λ_α t.c. $\sup_{\theta_0 \in \Theta_0} P[\lambda(\vec{y}) \le \lambda_\alpha, \theta_0 \in \Theta_0] \le \alpha$

$S_R = \{\vec{y} \in S$ t.c. $\lambda(\vec{y}) \le \lambda_\alpha\}$ o $S_R = \{\vec{y} \in S$ t.c. $W(\vec{y}) \ge W_\alpha\}$

dove $W_\alpha = -2\ln(\lambda_\alpha)$

$W(\vec{y}) = -2\ln(\lambda(\vec{y})) = 2\left(\ell(\hat{\theta},\vec{y}) - \ell(\hat{\theta}_0,\vec{y})\right)$

P-VALUE - LIVELLO DI SIGNIFICATIVITÀ OSSERVATO: metodo equivalente per definire regione di rifiuto

Def. $T(\vec{y})$ f. test per la verifica di $\begin{cases} H_0: \theta_0 \in \Theta_0 \\ H_1: \theta_0 \in \Theta_1 \end{cases}$

$\vec{y} = (y_1, \dots, y_n)$ c.c.s. estratto da $f_Y(y,\theta)$

$T(\vec{y}) = t^* \longrightarrow$ valore che f. test assume nella funzione

$$\text{P-value}(t^*) = \begin{cases} \sup_{\theta \in \Theta_0} P[T(\vec{y}) \ge t^*] & \text{se } S_R = \{\vec{y} \in S : T(\vec{y}) \ge r\} \\ \sup_{\theta \in \Theta_0} P[T(\vec{y}) \le t^*] & \text{se } S_R = \{\vec{y} \in S : T(\vec{y}) \le r\} \end{cases}$$

Estremo superiore delle probabilità di osservare per la funzione test un valore che è in disaccordo con H_0 almeno tanto quanto lo è t^*

Significato $0 \le \text{P-value} \le 1 \quad \forall t^*$

scarso accordo tra $\vec{y}$ e H_0 ... forte accordo tra $\vec{y}$ e H_0

misura diretta accordo tra $\vec{y}$ e H_0

Utilità: misurare l'accordo tra $\vec{y}$ e H_0 con un numero compreso tra 0 e 1

* Regione di rifiuto sulla base del p-value : $S_R = \{\vec{y} \in S$ t.c. p-value $(t^*) \le \alpha_T\}$

α_T: livello del test T

$S_R = \{\vec{y} \in S$ t.c. p-value $(t^*) \le \alpha\}$

α: test di livello α

* $T(\vec{y}) = \dfrac{\bar{y} - \mu_0}{s/\sqrt{n}} = t^*$

p-value $(t^*) = 2 \cdot P[T \ge |t^*|]$

dove $T \sim t\,(\nu = 1)$