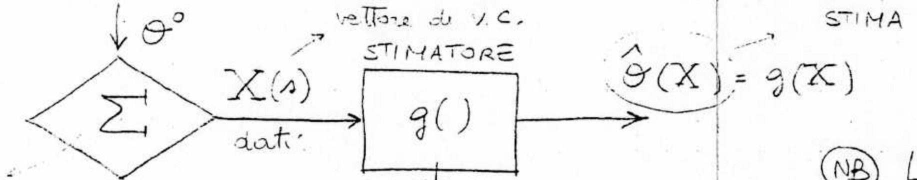
# TEORIA DELLA STIMA

In alcuni casi le ddp sono note, ma nella maggior parte dei casi non lo sono e ho a disposizione solo dati sperimentali.

**es** Errore di misura: lo descrivo come una V.C. gaussiana $X$, ma non conosco $E[X]$ né $Var[X]$.
Come <u>stimare</u> $E[X]$ e $Var[X]$ ?

**es** Stima di $\lambda$ per eventi di Poisson

Il problema della stima può essere schematizzato come segue




esperimento casuale che dipende da un parametro o da un vettore di parametri $\rightarrow \theta^\circ$

funzione
$\downarrow$
elaborazione dei dati

(NB) Lo stimatore è da PROGETTARE

$\hat{\theta}(X)$ è una V.C. !

es $X = \begin{bmatrix} 1 \\ 1 \\ 0 \\ 1 \\ 0 \\ 0 \\ 1 \end{bmatrix}$ Testa o croce $\longrightarrow$ $\quad g() = \dfrac{\sum x_i}{N} \quad \hat{\theta} = \dfrac{\sum x_i}{N}$ = probabilità di avere Testa

<u>L'esperimento $\Sigma$ è caratterizzato da una densità congiunta $f_X^{\theta^\circ}(x)$</u>

Dato che il risultato di $g(\cdot)$ è una V.C. non è detto che la stima sia sempre giusta.

**es** Stima di media ($m$) e deviazione standard ($\sigma$) di una V.C. gaussiana

- $\theta^\circ = \begin{bmatrix} \theta_1^\circ \\ \theta_2^\circ \end{bmatrix} = \begin{bmatrix} m \\ \sigma \end{bmatrix}$

- $\Sigma$ estrae $N$ valori indipendenti della V.C. gaussiana

- $X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix}$ $\quad f_X^{(\theta^\circ)}(x) = \dfrac{1}{\sqrt{(2\pi)^N}(\theta_2^\circ)^N} \cdot e^{-\frac{1}{2}\sum_1^N \left(\frac{x_i - \theta_1^\circ}{\theta_2^\circ}\right)^2}$

la ddp è fz di $\theta^\circ$

Stimatore abbastanza buono xché obbedisce alla legge forte dei grandi numeri

- $g(\cdot)$: posso stimare $m$ con la media campionaria

$$\hat{\theta} = \begin{bmatrix} \hat{\theta}_1 \\ \hat{\theta}_2 \end{bmatrix} = \begin{bmatrix} g_1(x_1, x_2 \dots, x_N) \\ g_2(x_1, x_2 \dots, x_N) \end{bmatrix} = \begin{bmatrix} \frac{1}{N}\sum_{i=1}^N x_i \\ ? \end{bmatrix}$$

lo vedremo in seguito

## <u>STIMA DEI MOMENTI DI UNA V.C.</u> : <u>MOMENTI CAMPIONARI</u>

Consideriamo un esperimento casuale che fornisce $N$ valori $x_1, x_2 \dots x_N$ iid con $f_{x_i}(x) = f_{x_j}(x) = f(x) \Rightarrow E[x_i] = m$, $Var[x_i] = \sigma^2$, $\forall i$

(ovviamente $f_X(x) = f_{x_1}(x_1) f_{x_2}(x_2) \dots f_{x_N}(x_N) = \prod_{i=1}^N f(x_i)$ )

PROBLEMA: Stimare i momenti di $X_i$

Introduciamo una classe di possibili stimatori:

MOMENTO CAMPIONARIO DI ORDINE K:

$$X = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_N \end{bmatrix}$$

$$M_k(X) := \frac{1}{N}\sum_{i=1}^{N} X_i^k \rightarrow \text{questa è una V.C.}$$

$$\left(\text{vs}: \underset{\substack{\downarrow \\ \text{questo è un numero}}}{m_k} = \int_{-\infty}^{\infty} x^k f(x)\,dx\right)$$

(NB) La media teorica e la media campionaria sono 2 cose MOLTO ≠ !

MOMENTO CENTRALE CAMPIONARIO DI ORDINE K

$$S_k(X) := \frac{1}{N}\sum_{i=1}^{N}\left(X_i - M_1\right)^k$$

$$\left(\text{vs}: \mu_k := \int_{-\infty}^{\infty}(x - m_1)^k f(x)\,dx\right)$$

Esaminiamo le proprietà di alcuni momenti campionari

MEDIA CAMPIONARIA $\quad M_1 := \frac{1}{N}\sum_{i=1}^{N} x_i$

- $E[M_1] = m$ (media teorica o media d'insieme)

$$\left(E\left[\frac{1}{N}\sum_1^N x_i\right] = \frac{1}{N}E\left[\sum x_i\right] = \frac{Nm}{N}\right)$$

- $Var[M_1] = \frac{\sigma^2}{N} \quad \left(Var\left[\frac{1}{N}\sum_1^N x_i\right] = \frac{1}{N^2}\cancel{N}\sigma^2 = \frac{\sigma^2}{N}\right)$

- $f_{x_i}$ gaussiana $\Rightarrow$ $M_1$ gaussiana (perché la somma di gaussiane è gaussiana)
- se le $f_{x_i}$ non sono gaussiane per $N \to \infty \Rightarrow M_1 \to$ gaussiana per il teo. centrale del limite

→ Se ho abbastanza dati posso "fregarmene" della ddp delle $x_i$ → già $N=10$ è suff.

OSSERVAZIONE: $\quad Z := \frac{M_1 - E[M_1]}{\sqrt{Var[M_1]}} = \frac{M_1 - m}{\sigma/\sqrt{N}} \Rightarrow M_1 = m + \underbrace{\frac{\sigma}{\sqrt{N}} \cdot Z}_{\text{componente casuale}} \rightarrow$ ERRORE

Media campionaria = media teorica + componente casuale → Termine di RUMORE

(NB) La media d'insieme non è mai raggiungibile pur essendo esistente.

Z è un termine FISSO ⇒ l'intensità del rumore dipende da σ e da N ⇒ per N grande il rumore diminuisce ⇒ → 0 per N → ∞

↓

Dà l'idea del perché funziona la legge dei grandi numeri

MOMENTO DI ORDINE 2 (VALOR QUADRATICO MEDIO CAMPIONARIO)

$$M_2 = \frac{1}{N}\sum_1^N x_i^2$$

- $E[M_2] = m_2 = E[X_i^2] \quad \left(E\left[\frac{1}{N}\sum x_i^2\right] = \frac{1}{N}\sum E[x_i^2] = E[x_i^2]\right)$

PARENTESI: LA V.C. $\chi^2$

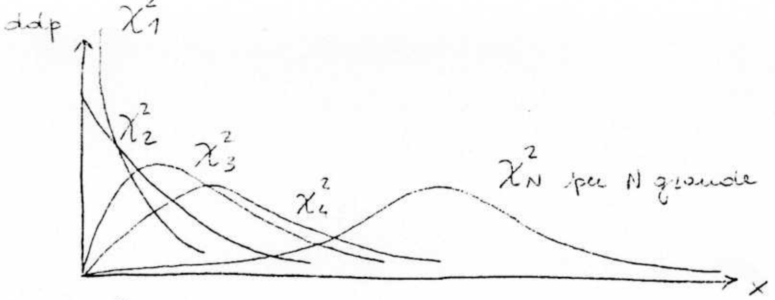
Siano $Z_i$, $i=1..N$ delle v.c. gaussiane standard ($E[Z_i]=0$, $Var[Z_i]=1$) indip.

Allora costruisco una nuova V.C:

$$\chi_N^2 := \sum_{i=1}^{N} Z_i^2$$

che prende il nome di "chi quadro" a N gradi di libertà

↓ ∃ un'intera famiglia di v.c. $\chi^2$




Dovrebbe esserci una tabella del $\chi^2$ ∀ valore di N

- $E[\chi^2_N] = N$
- $Var[\chi^2_N] = 2N$
- per $N \to \infty$ $\quad \frac{\chi^2_N - N}{\sqrt{2N}} \xrightarrow{\text{IN DISTRIBUZIONE}} Z$ (gaussiana standard) $\quad$ (Per il Teo. Centrale del limite)

FINE PARENTESI

- se $X_i$ gaussiane con $E[x_i] = 0$ si dimostra che $\frac{N M_2}{\sigma^2} = \chi^2_N$

VARIANZA CAMPIONARIA NOTO IL VALOR MEDIO

$$S^2_m := \frac{1}{N} \sum_1^N (x_i - m)^2$$

se non conosco il valor medio giusto m uso la stima $M_1$

- $E[S^2_m] = \sigma^2 \quad \left( E\left[\frac{1}{N} \sum (x_i - m)^2\right] = \frac{1}{N} \sum E(x_i - m)^2 = \frac{\cancel{N}\sigma^2}{\cancel{N}} \right)$

VARIANZA CAMPIONARIA

$$S^2 := S_2 := \frac{1}{N} \sum_{i=1}^N (x_i - M_1)^2$$

- $E[S^2] = \frac{N-1}{N} \sigma^2$ per dimostrarlo si usa:

PROPRIETÀ : $S^2 = M_2 - M_1^2$ (infatti $\sigma^2 = E[x^2] - m^2 = m_2 - m_1^2$)

$$\Rightarrow E[S^2] = E[M_2] - E[M_1^2] = m_2 - (Var[M_1] + E[M_1]^2) =$$
$$= m_2 - \frac{\sigma^2}{N} - m^2 = (m_2 - m^2) - \frac{\sigma^2}{N} = \sigma^2 - \frac{\sigma^2}{N}$$

cvd

(NB) la varianza della media campionaria ≠ media della varianza campionaria

- TEOREMA (FISHER): Se $x_i$ gaussiane iid, $S^2$ è una $\chi^2_{N-1}$ ed è indip da $M_1$

Più precisamente: $\frac{NS^2}{\sigma^2} = \sum_{i=1}^N \left(\underbrace{\frac{x_i - M_1}{\sigma}}_{\text{pseudonormalizzazione}}\right)^2 = \chi^2_{(N-1)}$ $\quad$ N-1 e NON N gradi di libertà perché i loro addendi condividono $M_1$ ⇒ perdono 1 g.d.l.

- per $N \to \infty$ ⇒ $S^2 \to$ (gaussiana & indip. da $M_1$) anche se $f_{x_i}$ non è gaus.

convergenza meno robusta della media campionaria ⇒ N deve essere grande

## PROPRIETÀ DEGLI STIMATORI

NON POLARIZZAZIONE : $\hat{\theta}(x)$ è detto non polarizzato (corretto, non deviato, unbiased) se

$$E[\hat{\theta}(x)] = \theta^0$$

*Esempio* La media campionaria è uno stimatore consistente della media (legge dei grandi numeri).

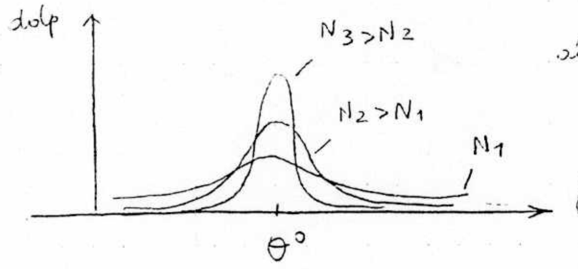
Cosa posso dire per $M_2$ ?

$Y := X^2 \Rightarrow M_2$ è la media campionaria di $Y \Rightarrow$ per la legge dei grandi numeri $M_2$ converge a $E[Y] = E[X^2]$

Allora la consistenza si può estendere anche a $M_3, M_4 \dots$

Tutti i momenti campionari sono consistenti

- La CONSISTENZA è una proprietà fortemente desiderabile




al crescere di N vorrei una delta di Dirac in $\theta^0$ $\Rightarrow$ per $N \to \infty$ voglio essere certo (probabilità 1) di essere in $\theta^0$

PROPRIETÀ : I momenti centrali campionari sono stimatori consistenti

ASINTOTICA NORMALITÀ : $\hat{\theta}(X)$ è asintoticamente normale se per $N \to \infty$ converge in distribuzione ad una gaussiana

*Esempio*: I momenti campionari sono asint. normali perchè sono medie di v.c. i.i.d. (si dimostra con $Y = X^k$)
lo stesso vale per i momenti centrali campionari (dimostrazione meno semplice)

DISUGUAGLIANZA DI CRAMER - RAO (1945)

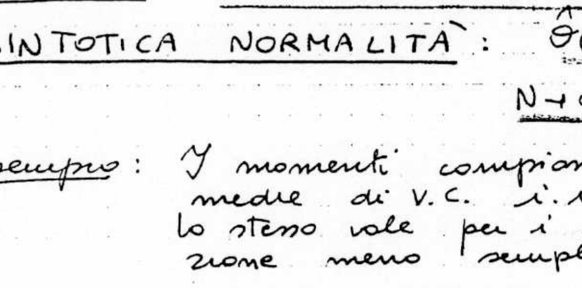
Considero uno stimatore non polarizzato ($E[\hat{\theta}] = \theta_0$) scalare. Allora, sotto condizioni di regolarità, vale

$\to \exists$ una legge analoga nel caso polarizzato

$$\mathrm{Var}[\hat{\theta}] \geq \frac{1}{-E\left[\frac{\partial^2}{\partial\theta^2} \ln f_X^{(\theta)}(X)\Big|_{\theta=\theta^0}\right]}$$

per lo stimatore NON polarizzato la varianza è SINONIMO di PRECISIONE

è una v.c. che la costruisco con i dati sperimentali




$\mathrm{Var}[\hat{\theta}]$ misura la precisione

La disuguaglianza di Cramer-Rao mi dà un limite al di sotto del quale non posso + migliorare

Interpretazione: qualunque stimatore io prenda, non riuscirò a portare la sua varianza al di sotto del limite di C.R.

Sono sicuro che se costruisco uno stimatore che ha var = a quella di CR è il migliore possibile

Il limite $\exists$ perchè dai dati sperimentali non posso "spremere" + di tanto

OSSERVAZIONI

- $S = -E\left[\frac{\partial^2}{\partial\theta^2} \ln f_X^{(\theta)}(X)\Big|_{\theta=\theta^0}\right]$ è detta quantità di informazione di Fisher

è l'inverso della varianza

S piccolo $\Rightarrow$ varianza grande

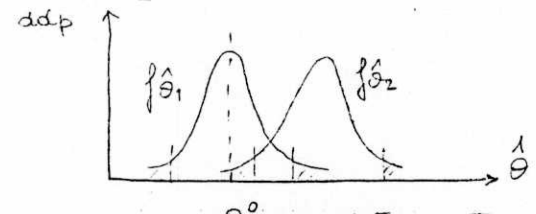
*Esempi*

- $M_1$ : non polarizzato (infatti $E[M_1] = m_1$)
- $M_2$ : " " ( " $E[M_2] = m_2$)
- $S^2_m$ : " " ( " $E[S^2_m] = \sigma^2$)
- $S^2$ : polarizzato ($E[S^2] = \frac{N-1}{N}\sigma^2$)

11-4-2001

## NON POLARIZZAZIONE

$$E[\hat{\theta}(X)] = \theta^0$$




$\theta^0$ = parametro esatto

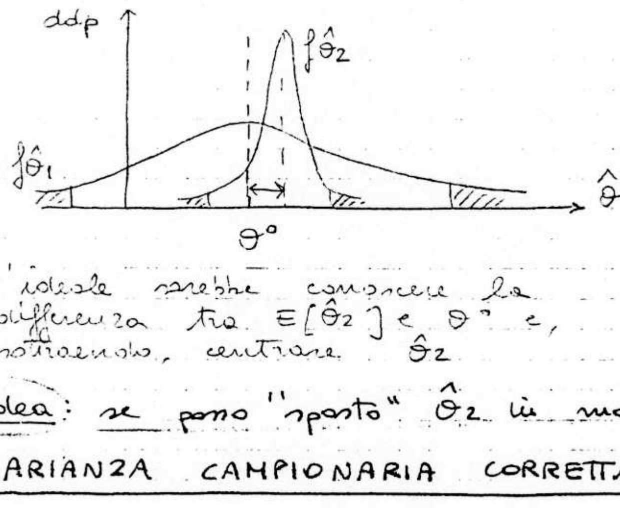
$f\hat{\theta}_1$ è NON polarizzato o CENTRATO perchè il valor medio cade sul parametro vero

$f\hat{\theta}_2$ è polarizzato

In questa situazione è preferibile $f\hat{\theta}_1$ perchè dà errori + piccoli

Sul singolo caso può capitare che $\hat{\theta}_2$ sia + vicino a $\theta^0$ di $\hat{\theta}_1$ ma in generale dà una stima migliore $\hat{\theta}_1$

(NB) Gli stimatori sono V.C. ⟹ andrebbero scritti in maiuscolo $\hat{\Theta}$




Per lo stimatore $\hat{\theta}_1$ posso commettere anche errori consistenti mentre $\hat{\theta}_2$ è + vicino al valore vero.

Anche se polarizzato $\hat{\theta}_2$ potrebbe essere preferibile !!

(NB) Posso accettare un po' di polarizzazione se guadagno in errore

L'ideale sarebbe conoscere la differenza tra $E[\hat{\theta}_2]$ e $\theta^0$ e, sottraendola, centrare $\hat{\theta}_2$

Idea: se posso "sposto" $\hat{\theta}_2$ in modo da eliminare bias

## VARIANZA CAMPIONARIA CORRETTA

$$S^2_c := \frac{N}{N-1} S^2 = \frac{1}{N-1} \sum_{i=1}^{N} (X_i - M_1)^2$$

Si potrebbe anche dividere per N xchè per N grande è la stessa cosa

dividendo per N-1 abbiamo centrato $\hat{\theta}$

Polarizzato ?

$$E[S^2_c] = E\left[\frac{N}{N-1} S^2\right] = \frac{N}{N-1} E[S^2] = \frac{\cancel{N}}{\cancel{N-1}} \cdot \frac{\cancel{N-1}}{\cancel{N}} \sigma^2 = \sigma^2 \Rightarrow \text{NON POLARIZZATO}$$

(NB) Per centrarlo abbiamo moltiplicato $S^2$ per una quantità $\left(\frac{N}{N-1}\right) > 1$ ⟹ la $Var[S^2_c] = \left(\frac{N}{N-1}\right)^2 Var[S^2]$ è + grande → pago la NON polarizzazione

**CONSISTENZA**: $\hat{\Theta}(X)$ è detto consistente se per $N \to \infty$, $\hat{\theta} \to \theta^0$ in probabilità

UNO STIMATORE che NON gode di questa proprietà è SOSPETTO

I momenti campionari sono tutti consistenti x la legge dei grandi n°

- Estensione al caso vettoriale

$S = \{ S_{ij} \}$ Matrice di INFORMAZIONE di FISHER

$$S_{ij} := -E\left[ \frac{\partial^2}{\partial\theta_i \partial\theta_j} \ln f_x^{(\theta)}(X) \Big|_{\theta=\theta^\circ} \right]$$

$$\rightarrow \boxed{Var[\hat{\theta}] \geq S^{-1}}$$

(significa che $\underbrace{Var[\hat{\theta}] - S^{-1}}_{A} \geq 0 \Rightarrow x^T A x \geq 0 \; \forall x$ cioè la differenza è semidef. positiva)

- Nel caso di $\hat{\theta}$ polarizzato NON ha senso dare un limite inferiore alla varianza

ES: Se pongo $\hat{\theta}(X) = 10 \; \forall X$ (stimatore che ignora i dati) $Var[\hat{\theta}(X)] = 0$

$\Rightarrow$ varianza nulla ma $\hat{\theta}$ è un pessimo stimatore

$\rightarrow$ Come misurare la precisione di uno stimatore polarizzato ?

Risposta: Valuto $\boxed{E[(\hat{\theta} - \theta^\circ)^2]}$ (non $E[\hat{\theta} - \theta^\circ] \cong 0$ perchè potrei ottenere questo risultato compensando errori grandi ma $>0$ con errori grandi ma $<0$).

$\hat{\theta} - \theta^\circ$ è l'errore

MEAN SQUARE ERROR

o

ERRORE QUADRATICO MEDIO

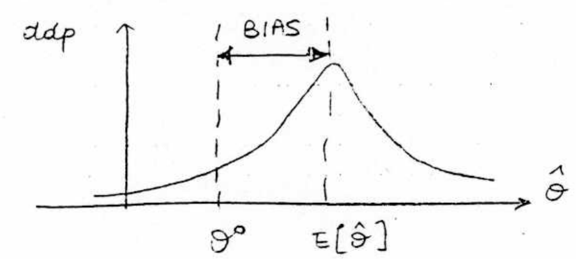
Si noti che $E[(\hat{\theta} - \theta^\circ)^2] = E[((\hat{\theta} - E[\hat{\theta}]) + (E[\hat{\theta}] - \theta^\circ))^2] =$

$$= Var[\hat{\theta}] + \underbrace{2E[\hat{\theta} - E[\hat{\theta}]]}_{=0} \cdot (\underbrace{E[\hat{\theta}] - \theta^\circ}_{\text{deterministico}}) + (E[\hat{\theta}] - \theta^\circ)^2 =$$

$$= Var[\hat{\theta}] + (\underbrace{E[\hat{\theta}] - \theta^\circ}_{\text{BIAS}})^2$$

BIAS $\rightarrow$ "scentratura" dello stimatore rispetto al valore vero

Possiamo accettare una "scentratura" se guadagnamo abbastanza in varianza o viceversa.




Def: Uno stimatore non polarizzato $\hat{\theta}^m$ si dice a MINIMA VARIANZA se

$$Var[\hat{\theta}^m] \leq Var[\hat{\theta}] \; , \; \forall \hat{\theta} \text{ non polarizzato}$$

($\hat{\theta}$ raggiunge il limite di C.R. $\Rightarrow \hat{\theta}$ è a MINIMA VARIANZA)

NON è garantito che ogni stimatore possa raggiungere il limite di C.R

$\Rightarrow$ MIN. VAR e C.R NON sono sinonimi

Esempio Calcolare la quantità di info di Fisher per la stima della media di una V.C. gaussiana.

$$X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix} \quad \text{con } x_i \text{ iid}$$

ddp che [illegible] in comune (sono tutte =)

$\boxed{\text{NOTA}}$ $f_x(x) = f_{x_1}(x_1) \cdot f_{x_2}(x_2) \cdots f_{x_N}(x_N) = \prod_{i=1}^{N} f(x_i)$

$$\Rightarrow \quad -E\left[ \frac{\partial^2}{\partial \theta^2} \ln f_X^{(\theta)}(X) \Big|_{\theta=\theta^\circ} \right] = -E\left[ \frac{\partial^2}{\partial \theta^2} \sum_{i=1}^{N} \ln f^{(\theta)}(x_i) \Big|_{\theta=\theta^\circ} \right] =$$

$$= -\sum_{i=1}^{N} E\left[ \frac{\partial^2}{\partial \theta^2} \ln f^{(\theta)}(x_i) \Big|_{\theta=\theta^\circ} \right]$$

$\theta^\circ = m \qquad f^{(\theta)}(X_i) = \dfrac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{1}{2}\left(\frac{x_i - \theta}{\sigma}\right)^2}$ generica gaussiana

$$\ln f^{(\theta)}(x_i) = -\ln\left(\sqrt{2\pi}\,\sigma\right) - \frac{1}{2}\left(\frac{x_i - \theta}{\sigma}\right)^2$$

$$\frac{\partial}{\partial \theta}\left(\ln f^{(\theta)}(x_i)\right) = \frac{x_i - \theta}{\sigma^2}$$

per la $\sum_1^N \dots$

$$\frac{\partial^2}{\partial \theta^2}\left(\ln f^{(\theta)}(x_i)\right) = -\frac{1}{\sigma^2} \quad \Rightarrow \quad S = \frac{N}{\sigma^2} \quad \Rightarrow \quad Var[\hat{\theta}] \geq \frac{\sigma^2}{N} = Var[M_1]$$

$\Rightarrow$ la media campionaria è il miglior stimatore possibile perché raggiunge il limite di C.R.

24-4-2001

$(*)$ $\quad f \xrightarrow{X} \boxed{g} \longrightarrow \hat{\theta}$ $\qquad \theta$ = Momento $\Rightarrow$ $g$ = Momenti campionari

$\theta = ? \Rightarrow g(\cdot) = ?$

## CRITERI DI STIMA

### CRITERIO DELLA MAX VEROSIMIGLIANZA (Maximum Likelihood, Lord Fisher '20)

Faccio riferimento alla formulazione generale del problema della stima. $(*)$

In particolare suppongo di conoscere $f_X^{(\theta)}(x)$ (CONGIUNTA)

Qualche volta, si scrive $f_X(x|\theta)$ $\rightarrow$ NON ha senso perché $\theta$ non è una v.c.

Per un dato valore $\bar{x}$ (vettore di numeri), dato che $\bar{x}$ è fisso, il parametro è $\theta$

$L(\theta, \bar{x}) := f_X^{(\theta)}(\bar{x}) = f_X(\bar{x}|\theta)$ è detta **VEROSIMIGLIANZA** di $\theta$ sulla base dell'osservazione $\bar{x}$

Per v.c. discrete $\quad L(\theta, \bar{x}) := P^{(\theta)}(X = \bar{x})$

*Esempio*: Moneta $\quad P(\text{Testa}) = \theta$

prove di Bernoulli (prob. di avere 4 successi)

Lanciando la moneta 4 volte ottengo 4 Teste $\Rightarrow L(\theta, \{T,T,T,T\}) = \theta^4$

$\begin{cases} \text{Se la moneta è truccata e } \theta = 1 \Rightarrow L = 1 \\ \text{" " è onesta e } \theta = \frac{1}{2} \Rightarrow L = \frac{1}{16} \end{cases}$ $\rightarrow$ è come se dessi un voto alle mie hp

$\Rightarrow$ il valore $\theta = 1$ è più verosimile di $\theta = 0{,}5$

Se invece ottengo 2 Teste e 2 Croci $\Rightarrow L(\theta, \{2T, 2C\}) = \binom{4}{2}\theta^2(1-\theta)^2$

Se $\theta = 1 \Rightarrow L = 0$ $\longrightarrow$ effettivamente NON posso pensare che la moneta abbia 2 Teste se è uscito croce

Se $\theta = 0{,}5 \Rightarrow L = 6 \cdot \frac{1}{16} = 0{,}375$

Se $\theta = 0{,}1 \Rightarrow L = 6 \cdot 0{,}01 \cdot 0{,}81 = 0{,}0486$

il valore $\theta = 0{,}5$ è il + verosimile

$\boxed{\text{NOTA}}$: $L(\cdot, x)$ NON è una ddp perché ciò che varia è $\theta$, NON $x$

parametro fisso

Pertanto, in generale, $\displaystyle\int_{D_\theta} L(\theta, x)\, d\theta \neq 1$ dove $D_\theta$ = insieme dei valori ammissibili di $\theta$ =

$\rightarrow$ non ho usato $\pm\infty$

perché potrebbe essere un integrale doppio

e perché potrebbe non $\exists$ per valori $< 0$

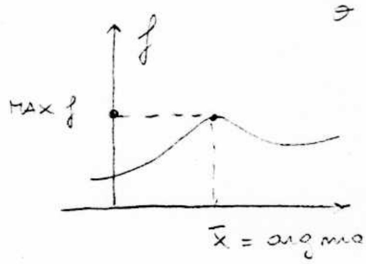
# CRITERIO DELLA MAX VEROSIMIGLIANZA → Scuola dei FREQUENTISTI

$$\theta^{ML}(X) = \underbrace{\arg\max_{\theta \in D_\theta}}_{\text{cerco il valore di } \theta \text{ che massimizza } L(\cdot, X)} L(\theta, X)$$

MAXIMUM likelihood

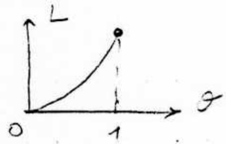
dovrei usare la maiuscola $\Theta^{ML}(X)$ perché è una V.C.

se, x es, non ammettessi valori < 0 non potrei accettare un $\bar{x} < 0$



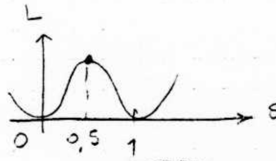

*Esempio*: Moneta, $k = n°$ di teste, $N$ lanci. $L(\theta, k) = \binom{N}{k} \theta^k (1-\theta)^{N-k}$ $\quad 0 \le \theta \le 1$ ($D_\theta$)

$\bar{k} = 4 \Rightarrow L(\theta, 4) = \theta^4$




$\Rightarrow \theta^{ML} = 1$

$\bar{k} = 2 \Rightarrow L(\theta, 2) = 6\theta^2(1-\theta)^2$




$\Rightarrow \theta^{ML} = 0{,}5$

In generale, se ho k teste su N lanci $\Rightarrow \theta^{ML} = \frac{k}{N}$

NB $\Bigg[$ Spesso è + comodo massimizzare $S = \ln L$ ( $S$ detto supporto o log-likelihood
$$\theta^{ML}(X) = \arg\max_{\theta \in D_\theta} \ln L(\theta, X)$$

Uno dei vantaggi è che, se ho N campioni indipendenti,

$$S(\theta, x) = \ln L(\theta, x) = \ln \prod_{i=1}^{N} f^{(\theta)}_{X_i}(X_i) = \sum_{i=1}^{N} \ln f^{(\theta)}_{X_i}(X_i) = \sum_{i=1}^{N} S_i(\theta, X_i)$$

Massimizzare S o L è del tutto equivalente ai fini del risultato finale
( $x_1 > x_2 \iff \ln x_1 > \ln x_2$, $x_1, x_2 \ge 0$)
$L_1 > L_2 \qquad S_1 > S_2$

*Esempio*: Dati N campioni indipendenti di una v.c. gaussiana, determina re lo stimatore congiunto ML di $m$ e $\sigma^2$.

perché li voglio stimare entrambi in un colpo solo

$$\theta = \begin{bmatrix} m \\ \sigma^2 \end{bmatrix} = \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix} \qquad S(\theta, X) = \ln\left( \frac{1}{\sqrt{(2\pi\sigma^2)^N}} e^{-\frac{\sum (x_i - m)^2}{2\sigma^2}} \right) =$$

è ancora una v.c. perché non ho ancora fatto l'esperimento

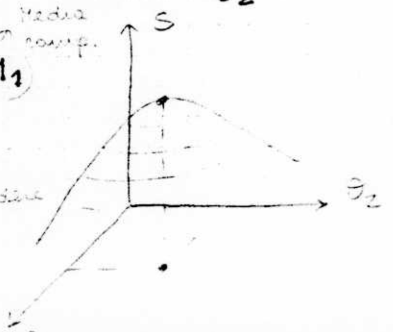
ddp congiunta dei dati sperimentali

$$= \ln\left( \frac{1}{\sqrt{(2\pi\theta_2)^N}} e^{-\frac{1}{2}\frac{\sum (x_i - \theta_1)^2}{\theta_2}} \right) = -\ln\sqrt{(2\pi)^N} - \frac{N}{2}\ln\theta_2 - \frac{1}{2}\sum_{i=1}^{N} \frac{(x_i - \theta_1)^2}{\theta_2}$$

$$\frac{\partial S(\theta, x)}{\partial \theta_1} = 0 \Rightarrow \sum_{i=1}^{N} \frac{x_i - \theta_1}{\cancel{\theta_2}} = 0 \Rightarrow \theta_1^{ML} = \frac{1}{N}\sum_{i=1}^{N} x_i = M_1$$

media campionaria

dovrei verificare l'hessiano e vedere se è un max



$$\frac{\partial S(\theta,x)}{\partial\theta_2} = 0 \Rightarrow -\frac{N}{2\theta_2} + \frac{1}{2}\sum_{i=1}^{N}\frac{(x_i-\theta_1)^2}{\theta_2^2} = 0 \Rightarrow \theta_2^{ML} = \frac{1}{N}\sum_{i=1}^{N}(x_i - M_1)^2 = S^2$$

varianza campionaria

(si verifica che è un pto di max)

<u>NB</u>: $S^2$ è uno stimatore polarizzato, $E[S^2] \neq \sigma^2$

<u>Esercizio</u>: Siano noti $N$ Tempi di attesa $W_i$ tra eventi di Poisson. Trovare lo stimatore ML di $\lambda$ e di $m := E[W_i]$ $\quad \lambda = \frac{N}{\sum W_i} \quad m = \frac{\sum W_i}{N}$

## PROPRIETÀ DEGLI STIMATORI ML

- Sotto ipotesi di regolarità, se $\eta = h(\theta)$ è un parametro funzione di $\theta$ (es: $\eta = \theta^3$) allora:

  funzione

  $$\eta^{ML} = h(\theta^{ML}) \qquad (\text{es: } \eta^{ML} = (\theta^{ML})^3)$$

  Nell' ESERCIZIO sopra $m = E[W_i] = \frac{1}{\lambda} \Rightarrow m^{ML} = \frac{1}{\lambda^{ML}}$ (ci sono altre tecniche di stima per cui non vale questa proprietà).

- Sotto hp di regolarità $\theta^{ML}$ per campioni indipendenti è uno stimatore asintoticamente:
  iid
  n° osservazioni $\to \infty$
  - consistente
  - non polarizzato (*)
  - raggiunge il limite di C.R. → tende a diventare imbattibile
  - gaussiano

  (*) NON è in contraddizione con l'es: $S^2$ perché ciò è vero ASINTOTICAMENTE, cioè x un gran n° di dati

  limite del metodo: x pochi dati tutti questi vantaggi NON valgono.

<u>PREGIO</u>: ML è una tecnica generale (pur di avere a disposizione la ddp)

<u>DIFETTO</u>: Presuppone la conoscenza di $f_X^{(\theta)}(x)$ (hp: forte)

## STIMA "A POSTERIORI" → scuola BAYESIANA

Stima ML: $\theta$ è un parametro DETERMINISTICO ignoto

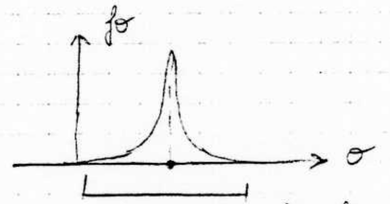
NON è una V.C.

Stima a posteriori: descrivo $\theta$ come una V.C. ⇒ <u>Teorema di Bayes</u>

26-4-2001

Idea: <u>Trattare $\theta$ come una V.C.</u>

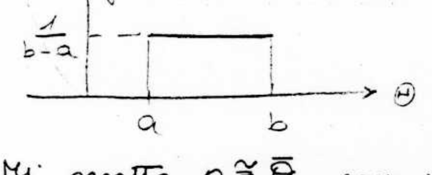
es:




mi aspetto che la cost. di gravitazione universale stia in questo intervallo

<u>Esempi</u> (come rappresentare le conoscenze a priori mediante ddp).

So che $a \le \Theta \le b$ ma non ho motivi per preferire determinati valori

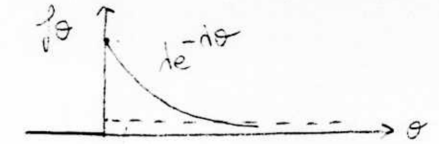



Mi aspetto $\theta \cong \bar{\theta}$ con un po' di incertezza


Se mi fidassi completamente


ma non ha + senso impostare un pb di stima

So che $\theta \geq 0$ e che valori troppo grandi sono improbabili

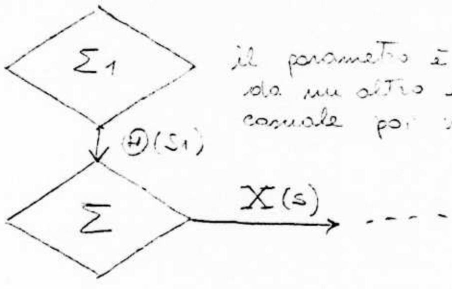



se conosco la media $\Rightarrow \lambda = \frac{1}{m}$

unico parametro dell' exp

caso di MAX incertezza $\Rightarrow \lambda$ molto piccolo $\Rightarrow$ è quasi piatta

## SCHEMA MODIFICATO DEL PB DELLA STIMA




il parametro è generato da un altro esperimento casuale poi viene usato x un esperimento casuale

$\Sigma_1$ : sorgente casuale caratterizzata da $f_\Theta(\theta)$

$\Sigma$ : sorgente casuale caratterizzata da $f_{x|\theta}(x \mid \Theta = \theta)$

Idea: se conosco $f_\Theta$ e $f_{x|\Theta}$, posso calcolare la ddp a posteriori condizionata dal vettore x delle osservazioni

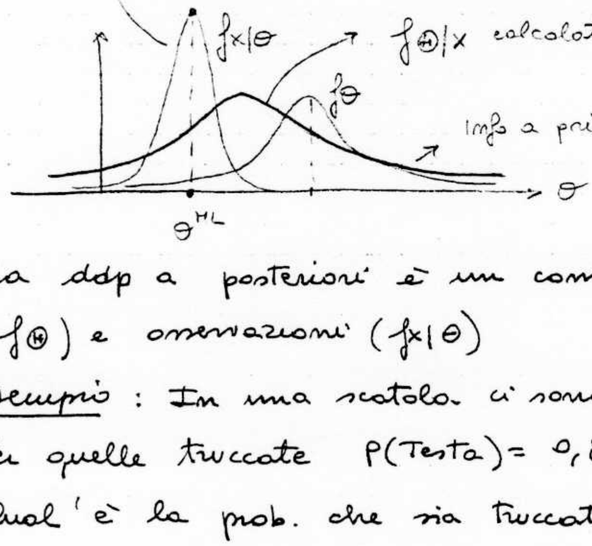
$$\underbrace{f_{\Theta|X}(\theta|X)}_{\text{Densità a posteriori}} \overset{\text{BAYES}}{=} \frac{\overbrace{f_{x|\Theta}(X|\theta)}^{\text{VEROSIMIGLIANZA}}\overbrace{f_\Theta(\theta)}^{\text{ddp a priori}}}{\underbrace{\int_{D\theta} f_{x|\Theta}(X|\theta) f_\Theta(\theta)\, d\theta}} \rightarrow \text{Teo della Prob. TOT}$$

il denominatore è una cost. di normalizzazione $\Rightarrow$ si può calcolare imponendo che l'area sia unitaria (una volta noto il numeratore)

NON dipende da $\theta$

evidenza sperimentale

es:




$f_{\Theta|x}$ calcolata con BAYES

NOTO $f_\theta$ faccio l'esperimento e calcolo la verosimiglianza

Info a priori

la ddp a posteriori è un compromesso tra informazioni a priori ($f_\Theta$) e osservazioni ($f_{x|\theta}$)

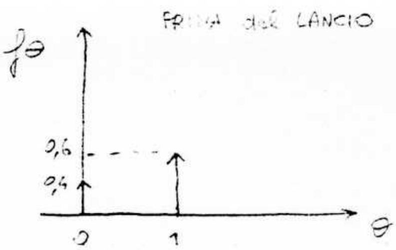
*Esempio*: In una scatola ci sono monete oneste (40%) e Truccate (60%). Per quelle truccate $P(\text{Testa}) = 0,8$. Estraggo una moneta a caso. Qual'è la prob. che sia truccata? Se dopo averla lanciata esce croce, quale è la prob. a posteriori che sia truccata

$$\Theta = \begin{cases} 0 \ (\text{onesta}) & P(\Theta = 0) = 0,4 \\ 1 \ (\text{Truccata}) & P(\Theta = 1) = 0,6 \end{cases}$$

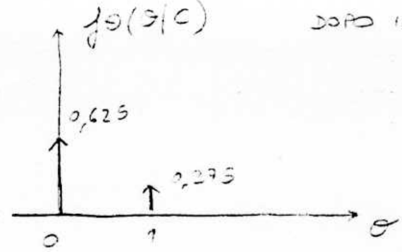
$P(c|\theta=0) = 0,5 \qquad P(c|\theta=1) = 0,2$

$$P(\Theta=1|c) = \frac{P(c|\Theta=1)P(\Theta=1)}{\underbrace{P(c|\Theta=1)P(\Theta=1) + P(c|\theta=0)P(\theta=0)}_{P(c)}} = 0,375$$

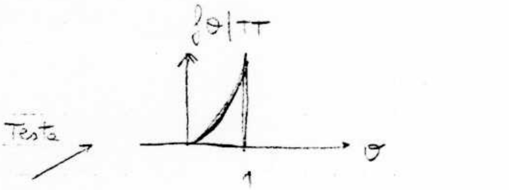
PRIMA del LANCIO

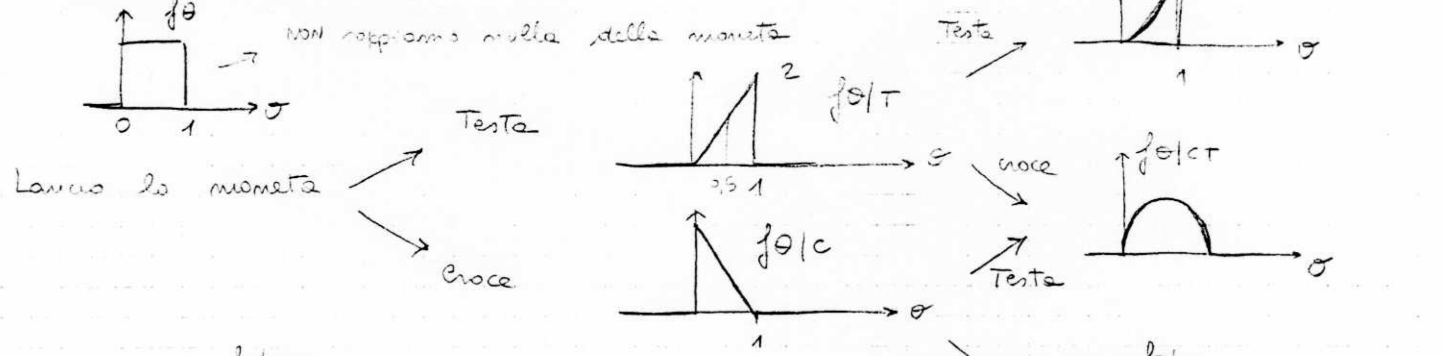


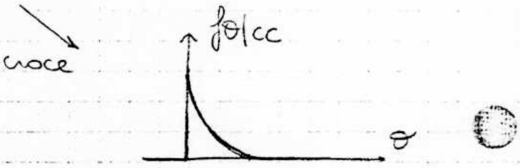

DOPO IL LANCIO (cambio opinione)



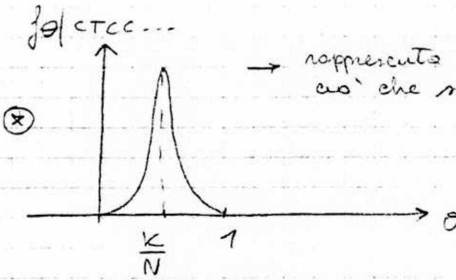

**Esempio**: moneta. Voglio stimare $p = P(T)$



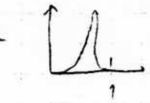

Iterando




→ rappresenta tutto ciò che so del pb.

$K =$ n° Teste su N lanci

6T su 10 lanci è $\neq$ da 600T su 1000 lanci $\Rightarrow$ se N è molto grande si stringe

Se parto hp che la moneta sia truccata  e itero, comunque ottengo la ddp giusta (*)

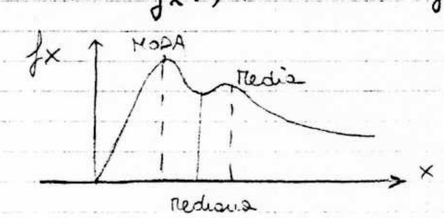
Solo se parto con 0 o 1 NON posso + correggere i miei pregiudizi

Come ottenere un pronostico dalla densità a posteriori? → pronostico

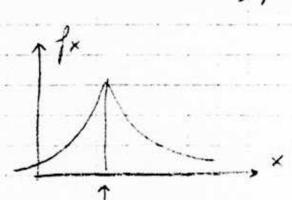
**PROBLEMA**: Una volta che conosco $f_{\theta|X}(\theta|X)$ come scelgo $\hat{\theta}$?

PARENTESI: STIMA DI UNA V.C. DI CUI È NOTA LA DDP

Conosco $f_X(x)$. Dovendo fare una scelta (un pronostico, x es), cosa scelgo?



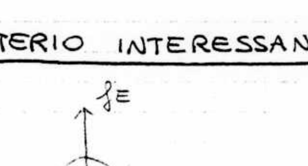

vengono usate tutte e 3
↓
posso scegliere una qualunque ma la prob. di avere esattamente un valore è $\emptyset$
$\Rightarrow$ meglio usare l'errore (che è una v.c)




CRITERIO INTERESSANTE:

$$\min_{\hat{x}} E\left[(X-\hat{x})^2\right] \qquad e := X - \hat{x}$$

↓ valore quadratico medio dell'errore
↓ piccolo solo se la densità è schiacciata intorno all'origine




errore in $|\cdot|$ piccolo $\Rightarrow$ prob. elevata che e cada in un intervallo piccolo

(minimizzo l'errore quadratico medio)

$\boxed{\text{TEOREMA}}$: $E\left[(X-\hat{x})^2\right]$ è minimo per $\hat{x} = E[X]$

Dimostrazione: $J = E\left[(X-\hat{x})^2\right] = \int_{-\infty}^{+\infty} (X-\hat{x})^2 f_X(x)\,dx$

minuscola

$$\frac{\partial J}{\partial \hat{x}} = 0 \Rightarrow \int_{-\infty}^{+\infty} \cancel{2}(X-\hat{x}) f_X(x)\, dx = 0 \Rightarrow \hat{x} = \int_{-\infty}^{\infty} x f_X(x)\, dx = E[X]$$

$$\frac{\partial^2 J}{\partial \hat{x}^2} = \int_{-\infty}^{\infty} f_X(x)\, dx = 1 > 0 \qquad \text{cvd}$$

E se voglio minimizzare $E[|X-\hat{x}|] = ?$ $\Rightarrow$ $\hat{x}$ = mediana

c'è una certa preferenza per il valor medio (xchè il quadrato è + facilmente derivabile)

STIMA DI UNA V.C. IN FUNZIONE DI UN'ALTRA V.C.

$X, Y$ v.c. congiunte con $f_{XY}(x,y)$ nota → Se NON conosco la ddp congiunta la stimo

PROBLEMA: Trovare $\hat{X}(Y)$ che stimi $X$ in base alla sola misura di $Y$

Suppongo di conoscere $f_{X|Y}$. Ragiono per $Y=y$

$\boxed{\text{TEOREMA}}$: $E[(X-\hat{X}(y))^2 \mid Y=y]$ è minimo per $\hat{X}^{MS}(Y) = E[X \mid Y=y]$

HS: Mean Square $\boxed{\text{STIMATORE IN MEDIA QUADRATICA}}$

bisogna conoscere la ddp condizionata (eventualmente la stim

Vd. fotocopie

# STIMA DI UNA V.C. IN FUNZIONE DI UN'ALTRA V.C.

Trovare $\hat{X}(Y)$ con $X, Y$ v.c. congiunte

(es conosco peso, voglio indovinare altezza) ⇒ cerco criterio che dica è metodo migliore

Criterio: voglio che errore che commetto sia piccolo

Teorema

Si vede che:

$$E[(X-\hat{X}(y))^2 \mid Y=y] \text{ è minimo per } \hat{X}^{MS}(y) = E[X \mid Y=y]$$

(stessa cosa che abbiamo visto per la singola v.c. ⇒ era il valor med

(Trovo che questo è il miglior stimatore possibile)

Questo che abbiamo risolto è il Problema di stima in media quadratica - Mean Square estimation.

**OSS:**

- Se $X$ e $Y$ sono indip. (voto an.1 e altezza) ⇒ $E[X|Y=y] = E[X]$ (perchè se le 2 var. sono indip ⇒ la ddp di $x|y$ è uguale alla ddp di $x$ ⇒ anche valor medio) ⇒ miglior stimatore che posso prendere è proprio il valor medi
- $E[(X-\hat{X}(y))^2 \mid Y=y] = E[(X-E[X])^2] = Var[X]$
- Se $X$ e $Y$ sono congiuntam. gaussiane (anche vettoriali: cioè $X$ e $Y$ possono essere vett ⇒ posso cercare di indovinare voto di geom da quello di an1 e an2 ⇒ devo considerare ddp congiunte e stimatore mig re resta il valor atteso condizionato).

→ $E[X|Y=y] = E[X] + V_{XY} V_Y^{-1} (y - E[Y])$

$$\left( Var\left[\begin{bmatrix} X \\ Y \end{bmatrix}\right] = \begin{bmatrix} V_X & V_{XY} \\ V_{XY} & V_Y \end{bmatrix} \right)$$

an1 an2

fis1 geom

Espressione dello stimatore, errore quadr. medio:

$$E[(X - E[X|Y=y])(X - E[X|Y=y])^T] =$$

$$= V_X - V_{XY} V_Y^{-1} V_{XY}^T.$$

Per v.a. scalari :

$$\underline{E[X|Y=y] = E[X] + \frac{r\,\sigma_x}{\sigma_y}\left(Y - E[Y]\right)} \quad *$$

Nel caso della gauss. è semplice:

$$\underbrace{\frac{E[X|Y=y] - \overset{\text{valor atteso a priori}}{E[X]}}{\sigma_x}}_{\text{peso normalizzato della mia previsione}} = r\,\frac{Y - E[Y]}{\sigma_y}$$

L'err. quadr medio che si commette è:

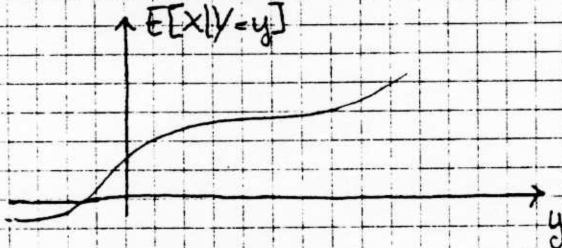
$$\rightarrow \underline{E\left[\left(X - E[X|Y=y]\right)^2 \,|\, Y=y\right] = \sigma_x^2\left(1 - r^2\right)}$$

se trovo 2 var. fortissimam correlate fra loro ($r \cong \pm 1$) $\Rightarrow$ ~~idealmente~~ ~~allora~~ l'err. quadr. medio sarebbe molto piccolo (anz condizionato da an1) ${}^{\text{r. attenersi su origine}}$ $\Rightarrow$ quando uso questo stimatore me ne accorgo, lo smaschero

Nel caso gaussiano lo stimatore $\overset{MS}{V}$ è una funzione lineare delle osservazioni.

(Nella $*$ l'unica var. è $Y$, tutti altri sono $n^i$ $\Rightarrow$ devo inserire solo $Y$ e conoscerò $E[X|Y=y]$) — nel pg che genera monostici

In generale, $E[X|Y=y]$ è una funzione non lineare di $y$.




Inoltre nel caso gaussiano, l'errore quadratico medio non dipende da $Y$. $E = \sigma_x^2(1-r^2)$ conosco già le proprietà statistiche prima che arrivino dati.

$\Rightarrow$ err. quadr medio nel caso gauss. non dipende dai dati $\Rightarrow$ posso verificarlo prima di vedere se lo stimatore funziona. — se è grande o piccolo

Io so che MS è lo stimatore migliore, però è costoso.
C'è qualcosa di + veloce anche se non migliore?

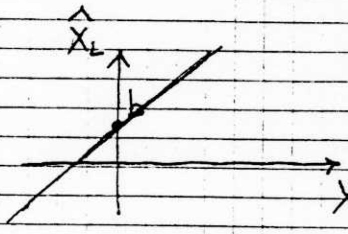
## STIMATORE MS LINEARE

Caso non gaussiano: cerco una soluzione subottima per non dover ricostruire $E[X|Y=y]$ (funz. non lineare di $y$).

Idea. restringo l'attenzione alla classe degli STIMATORI LINEARI (x' ha pochi parametri e facili da calcolare).

Problema: Considero la classe degli stimatori lineari $\hat{X}_L(Y) = aY +$
(vale anche nel caso di X e Y vettori → a e b matrici).
Trovare a e b che minimizzano:

$$J = E[(X-\hat{X}_L(Y))^2] = E[(X - aY + b)^2]$$




← è ciò che sto cercando!

(faccio derivate rispetto alle 2 (a, b), impongo =0 e verifico che hess sia >0).

Soluzione: $\hat{X}_L(Y) = E[X] + V_{XY} V_Y^{-1} (Y - E[Y])$

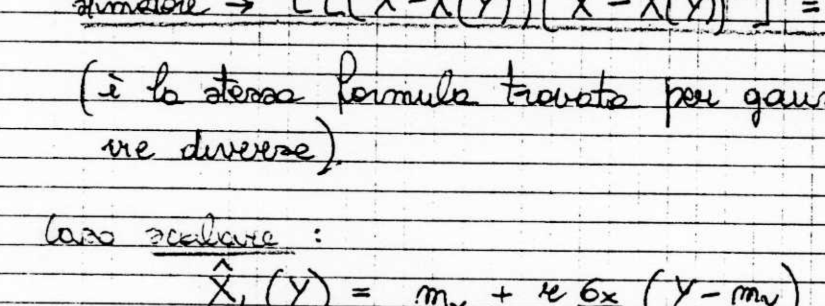
stimatore → $E[(X-\hat{X}(Y))(X-\hat{X}(Y))^T] = V_X - V_{XY} V_Y^{-1} V_{XY}^T$

(è la stessa formula trovata per gauss., ma ci siamo arrivati per vie diverse).

Caso scalare:

$$\hat{X}_L(Y) = m_x + \rho \frac{\sigma_x}{\sigma_y}(Y - m_y)$$

$$\rightarrow E[(X-\hat{X}_L(Y))^2] = \sigma_x^2 (1-\rho^2)$$




↳ faccio meglio ad usarlo, ma + difficile!

STIMATORI OTTIMI E STIMATORI LINEARI:
Lo stimatore che nel caso gaussiano è ottimo in assoluto, nel caso non gaussiano risulta ottimo limitatamente alla classe degli stimatori lineari.

(⇒ 2 strade diverse per trovare stimatore: o ipotizzo gaussianità, per devo dimostrarla in qualche modo, non solo da istogrammi; oppure sono + prudente e cerco uno stim. lineare e così trovo però uno stimatore che potrebbe non essere il migliore di tutti).

Torniamo alla stima a posteriori. Supponiamo di avere calcolato:

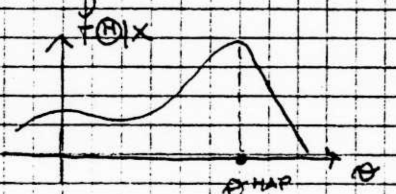
$$f_{\Theta|X}(\theta|X) \quad . \quad \text{Come generare } \hat{\theta}\,?$$

($\theta$: parametri incogniti; $X$: osservazioni)

(Ci sono $\neq$ alternative)

1ª strada: $\boxed{\text{STIMA MAP}}$ (Maximum A Posteriori)

$$\underline{\theta^{MAP}(X) = \arg\max_{\theta} f_{\Theta|X}(\theta|X)}$$




cioè scelgo la moda della ddp a posteriori come pronostico

Commenti:

- Dato che (per il teo. di Bayes)

$$f_{\Theta|X}(\theta|X) = \frac{f_{X|\Theta}(X|\theta)\, f_{\Theta}(\theta)}{(\ \ldots\ )}$$

non dipende da $\theta$

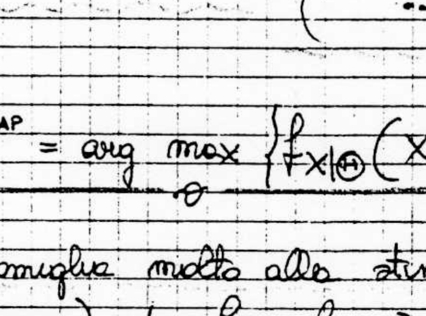
$$\Rightarrow \underline{\theta^{MAP} = \arg\max_{\theta} \left\{ f_{X|\Theta}(X|\theta)\, f_{\Theta}(\theta) \right\}}$$

Assomiglia molto alla stima ML (infatti senza $\Theta$, sembra la verosimiglianza). La formula è identica alla stima ML se $f_{\Theta}$ è uniforme (cioè se è cost).

Per calcolare max $\Rightarrow$ devo cercare pti di stazionarietà $\Rightarrow$ meglio usare LOGARITMO da max.

- Le tecniche numeriche sono analoghe a quelle già viste per la stima ML.

- Limite: in certi casi la moda è poco significativa




supp ddp a post. di tipo expon. $\Rightarrow \theta^{MAP} = 0$.
è info. sviante x' c'è 100% di probab. che sia > di 0 la probab.

2ª strada: **STIMA DI BAYES**

$$\vartheta^B(X) = E[\Theta|X]$$ (è lo stimatore in media quadr.)

($\vartheta^B$ • : stima migliore in quel caso)

Commenti

per gauss E[X] ≡ con MAP

- Se $\Theta$ e X sono cong. gaussiane,

$$\vartheta^B(X) = \vartheta^{MAP}(X) = E[\Theta] + V_{\Theta X} V_X^{-1} (X - E[X])$$

$$\left( \mathrm{Var}\left[\begin{bmatrix} X \\ \Theta \end{bmatrix}\right] = \begin{bmatrix} V_X & V_{X\Theta} \\ V_{X\Theta}^T & V_\Theta \end{bmatrix} \right)$$

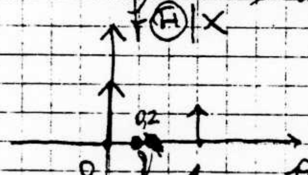
- Negli altri casi è raro trovare formule esplicite.

quando var. non sono cong. gauss.

[⇒ questo spiega successo della stima MAP è più semplice che calcolo l'$\int$ che richiede stima di B. (se ho ddp ricavo E[ ] con integr. ⇒ difficile da calcolare se ddp è brutta)]
Però ci sono tecniche di simulazione stocastica (richiedono generazione di n° casuali).

- Punto debole: in certi casi non ha senso prendere la media.

(ESEMPIO: supp. di avere param. $\Theta = 0 \to$ moneta onesta
$\Theta = 1 \to$ moneta truccata con p=0.8




media di 2 asticelle = 0,2 ma non ha senso, se moneta può avere ø od 1.

3ª strada: **STIMA MS LINEARE**

$$\vartheta^{MS}(X) = E[\Theta] + V_{\Theta X} V_X^{-1} (X - E[X])$$

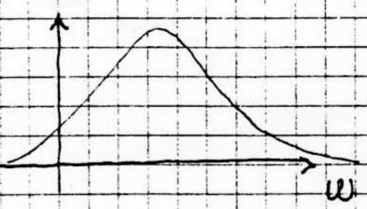
Commenti:

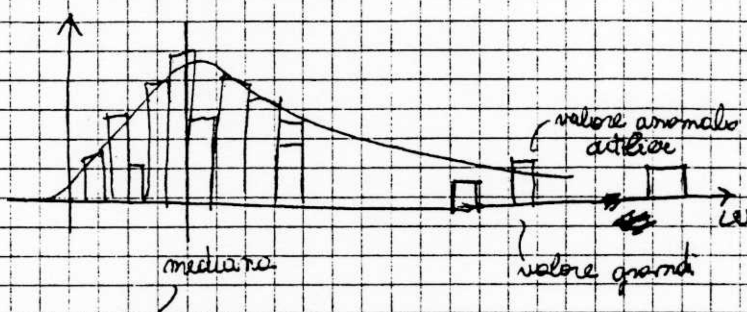
- Se $\Theta$ ed X sono congiuntam. gaussiane, $\vartheta^{MS}$ coincide con $\vartheta^B$ e $\vartheta^{MAP}$ (nel caso gaussiano, tutti gli stimatori sono equivalenti).

si può anche porre $\hat{\vartheta}$ = mediana$(f_{\Theta|x})$
(non ha un nome particolare questo stimatore)

Se ho var. cong gauss $\Rightarrow \hat{\vartheta} = \vartheta^{MS} = \vartheta^{B} = \vartheta^{MAP}$

MEDIAN in Matlab calcola la mediana
Può avere dei vantaggi rispetto allo stimatore con media x' se c'è un valore molto alto ⇒ media sballa, mentre mediana no.




anche se aggiungo valore □ estremo
mediana segna il 50% da una parte e
il 50% dall'altra ⇒ + robusto ad inserimenti
di valori estremi.

## STIMA A POSTERIORI: COMMENTI CONCLUSIVI

- Quando usarla:
  1) caso se $\theta$ è una v.c. (ex. scatola con monete oneste e truccate)
  2) Stima sequenziale (dati che arrivano a spezzoni nel tempo, consecutivamente ma ad intervalli: all'inizio ho insieme di dati e costruisco mia stima, poi arrivano altri dati ⇒ devo ricostruire stima. Posso procedere in 2 modi: o metto insieme sempre tutti i dati e costruisco valor atteso; oppure propago l'info usando la prima stima come stima a priori ed elaboro solo ultimo insieme di dati usandoli come stima a posteriori ⇒ non rischio che n° di dati esploda con questa 2ª strada).
  3) Uso $f_{\Theta}$ per rappresentare le mie conoscenze a priori (ex conoscenza di un esperto).
- Vantaggi: più generale di ML
  Possibilità di trattare situazioni con pochi dati (AUC).
- Svantaggi: Ipotesi forti (anche di + di quelle di ML: dovevo conoscere generaz. casuale). : Problema della scelta del "priore" $f_{\Theta}$. (non è facile tradurre in info a priori)

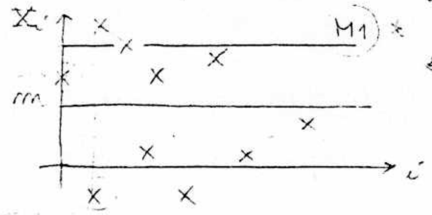
3-5-2001 

## INTERVALLI DI CONFIDENZA

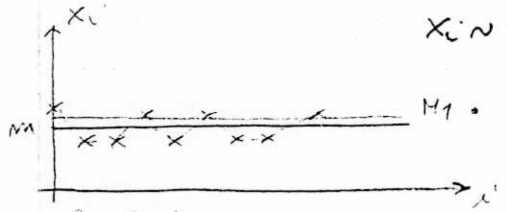
Fornire delle stime senza indicarne la precisione può essere del tutto fuorviante.

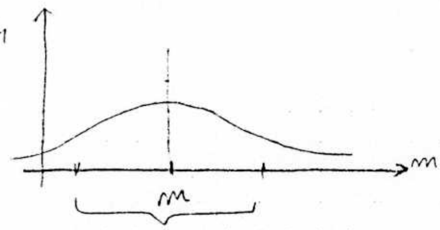
Esempio: $X_i$ iid $i = 1 \dots N$ $E[X_i] = m$ Considero come stimatore $M_1$ (media campionaria) N fissato

$X_i \sim G(m, 100)$ ← distr. gaussiana

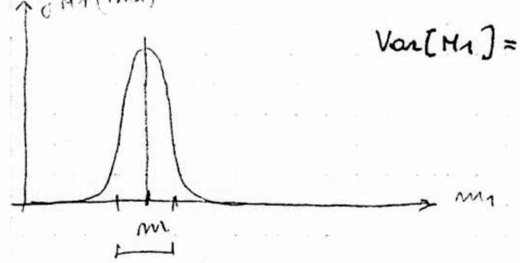



$X_i \sim G(m,1)$




$Var[M_1] = \frac{\sigma^2}{N} = \frac{100}{N}$



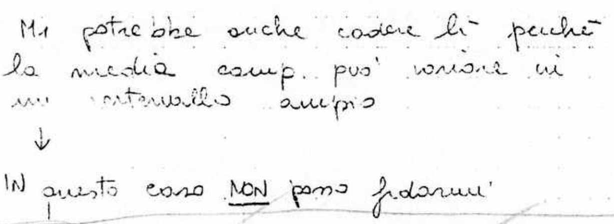

$Var[M_1] = \frac{1}{N}$

* $M_1$ potrebbe anche cadere lì perché la media camp. può variare in un intervallo ampio
↓
IN questo caso NON posso fidarmi

• $M_1$ può variare poco → intervallo + stretto

Devo sempre fornire il riferimento (→ intervalli di confidenza) prima di fornire un n°

Esempio: STIMA A POSTERIORI




ddp unimodale simmetrica
↓
non ho pb a scegliere la stima

$\theta^B \pm \delta_{0,9}$

5%

$\theta^B - \delta_{0,9}$

$\theta^{MAP} = \theta^B$ → oltre a questo n° bisogna dire quanto ci si può fidare (e sembra che NON ci si possa fidare).

Per trovare la forchetta taglio fuori 2 code (ciascuna del 5%) → c'è solo il 10% di prob di cadere fuori dallo intervallo $[+\delta \div -\delta]$

Cerco un valore $\delta_{0,9}$ tale che la prob. $P(\theta^B - \delta_{0,9} \le \Theta \le \theta^B + \delta_{0,9}) = 0,9$

Ho il 90% di prob che $\Theta \in \underbrace{I_{0,9}}_{\downarrow} := [\theta^B - \delta_{0,9}, \theta^B + \delta_{0,9}]$ potrei usare anche 0,95 o qualcos'altro

Intervallo Bayesiano (di confidenza) al 90%

$\gamma = 0,9$ è il LIVELLO DI CONFIDENZA

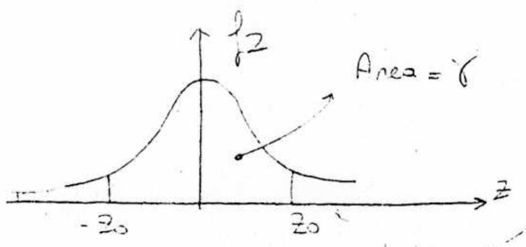
Se voglio sbagliare meno devo allungare l'intervallo (di solito si usano intervalli al 5% → 1 prob su 20 di sbagliare)

Come fare se non uso la stima a posteriori?

Esempio: Sia $\hat{\Theta}$ uno stimatore gaussiano non polarizzato ($E[\hat{\Theta}] = \theta$) avente varianza $\sigma^2_{\hat{\Theta}} := Var[\hat{\Theta}]$ NOTA. Trovare $I_\gamma$ per $\gamma$ fissato

Idea: standardizzazione $Z_1 = \frac{\hat{\theta} - \theta}{\sigma_{\hat{\theta}}} \sim G(0,1)$ (gauss. standard)

Cerco sulle Tabelle $z_0$ t.c. $P(-z_0 \le Z_1 \le z_0) = \gamma \Rightarrow P(|Z_1| \le z_0) = \gamma$




$$F_{Z_1}(z_0) - \underbrace{F_{Z_1}(-z_0)}_{\text{area}} = \gamma =$$

$$= F_{Z_1}(z_0) - (1 - F_Z(z_0)) =$$

$$= \boxed{2 F_Z(z_0) - 1 = \gamma}$$

$\Rightarrow F_Z(z_0) = \frac{1+\gamma}{2} = 0{,}5 + \frac{\gamma}{2}$

$\Rightarrow P\left(\left|\frac{\hat{\theta} - \theta}{\sigma_{\hat{\theta}}}\right| \le z_0\right) = \gamma \Rightarrow P(|\hat{\theta} - \theta| \le z_0 \sigma_{\hat{\theta}}) = \gamma \Rightarrow$

modulo dell'errore $\Rightarrow$ 1/2 forchetta

$P(\hat{\theta} - z_0 \sigma_{\hat{\theta}} \le \theta \le \hat{\theta} + z_0 \sigma_{\hat{\theta}}) = \gamma$

$\boxed{I_\gamma = [\hat{\theta} - z_0 \sigma_{\hat{\theta}},\ \hat{\theta} + z_0 \sigma_{\hat{\theta}}]}$

livello di prob.

In pratica se $\hat{\theta} = 7{,}3$, $z_0 \sigma_{\hat{\theta}} = 0{,}4 \rightarrow \theta = 7{,}3 \pm 0{,}4$ ($p = 0{,}95$)

↳ se NON è specificato si mette 95%.

Il Bayesiano dice che tutto ciò è un imbroglio perché $\hat{\theta}$ NON è una v.c. xché $\gamma$ ù natura
→ se raccolgo dei dati diventa una v.c.

IMBROGLIO: dire che $\theta \in I_\gamma$ perché una volta raccolto i dati $\hat{\theta}$ è un numero e non una v.c. $\Rightarrow$ o $\theta \in I_\gamma$ o $\theta \notin I_\gamma$ MA NON ha senso parlare di probabilità
È DIVERSO se parlo di PROB SOGGETTIVA

DIFESA del FISHERIANO: $\gamma$ è il n° di volte che ci azzecco se nella mia vita ripeto il "sondaggio" + volte $\Rightarrow \gamma$ = % di volte che ho ragione usando questa regola

*Esempio*: $X_i$ i.i.d, $Var[X_i] = \sigma^2$, $\sigma^2$ NOTA, N "grande". Trovare $I_\gamma$ per $M_1$. Parametro da cercare $\theta = m$

TEO. CENTRALE del LIMITE

$\hat{\theta} = M_1 \qquad \sigma_{\hat{\theta}} = \sqrt{\frac{\sigma^2}{N}} = \frac{\sigma}{\sqrt{N}} \qquad N \text{ "grande"} \Rightarrow M_1 \sim G\left(m, \frac{\sigma^2}{N}\right)$

sono soddisfatte tutte le HP del caso precedente

In virtù del caso precedente: $I_\gamma = \left[M_1 - z_0 \cdot \frac{\sigma}{\sqrt{N}},\ M_1 + z_0 \cdot \frac{\sigma}{\sqrt{N}}\right]$

| SE: $\frac{\sigma}{\sqrt{N}}$ è anche detto STANDARD ERROR |
|---|

$SD = \sigma \qquad SE = \frac{\sigma}{\sqrt{N}}$

↓ dev. std della media

Possibili PROBLEMI:

(a) noti $N = 100$ e $\gamma = 0{,}9$, trovare $I_\gamma$

(b) noti $N = 100$ e $I_\gamma = [M_1 - \delta, M_1 + \delta]$ Trovare $\gamma$ (con che prob posso garantire di non avere [illegible] > $\delta$ ?)

(c) noti $\gamma = 0{,}95$ e $I_\gamma = [M_1 - \delta, M_1 + \delta]$ Trovare N (es: Istituti Demoscopici)
↳ Quale N serve x essere sicuro con prob $\gamma$ di cadere dentro $I_\gamma$

(a) $\frac{1+\gamma}{2} = \frac{1{,}9}{2} = 0{,}95$ Sulle Tabelle trovo che $F_Z(z_0) = 0{,}95$ per

$z_0 = 1{,}645 \Rightarrow I_\gamma = \left[M_1 - \frac{1{,}645}{10}\sigma,\ M_1 + \frac{1{,}645}{10}\sigma\right]$

(b) Ipotizziamo che $\delta = \frac{\sigma}{10} \rightarrow \frac{z_0 \cancel{\sigma}}{\sqrt{N}} = \delta = \frac{\cancel{\sigma}}{10} \rightarrow z_0 = \frac{\sqrt{N}}{10} = 1$ N=100

$F_Z(z_0) = 0,8413$ (dalla tabella con $z_0 = 1$)

$= \frac{1+\gamma}{2} \Rightarrow \gamma = 0,6826 \rightarrow$ prob del 68% di commettere un errore $\leq \frac{\sigma}{10}$

(c) Suppongo $\delta = \frac{\sigma}{10}$ $\quad F_Z(z_0) = \frac{1+\gamma}{2} = \frac{1+0,95}{2} = 0,975$ per $z_0 = 1,96$

$\frac{z_0 \cancel{\sigma}}{\sqrt{N}} = \delta = \frac{\cancel{\sigma}}{10} \Rightarrow N = 100 \cdot (1,96)^2 = 384,16$

NOTA: $\boxed{\sqrt{N} = \frac{z_0 \sigma}{\delta}} \rightarrow \boxed{N = \left(\frac{z_0 \sigma}{\delta}\right)^2}$ per dimezzare la forchetta ($\delta$) dobbiamo quadruplicare i dati (N)!

$\Rightarrow$ la precisione COSTA! NON è tanto facile guadagnare

Se il n° di osservazioni è piccolo ci salviamo <u>SOLO</u> se sono distribuite gaussianamente $\Rightarrow$ media camp. gaussiana <u>VS</u> NO

ESEMPIO: $X_i$ i.i.d gaussiane, $\sigma^2$ non nota.

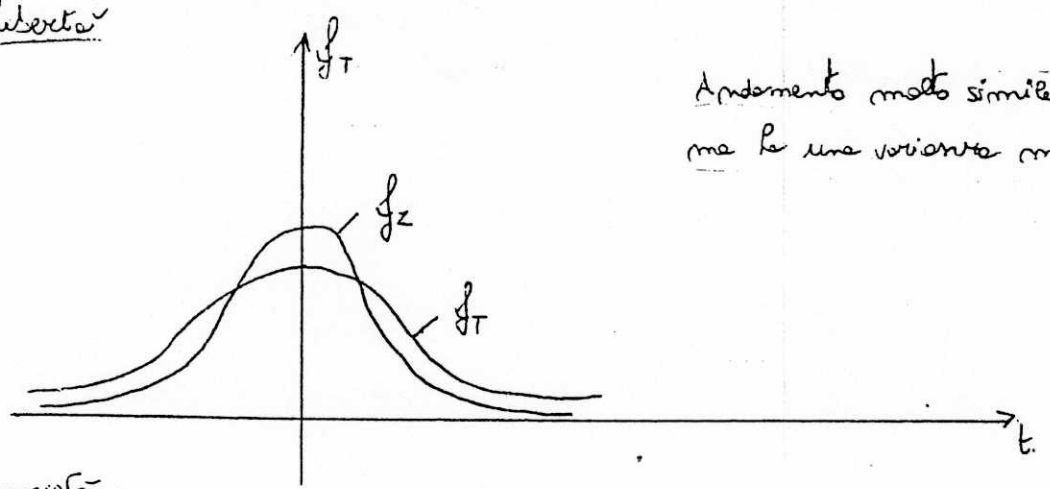
Trovare $I_\gamma$ per $M_1$

PARENTESI — La t di Student

La v.c

$$T_N := \frac{Z}{\sqrt{\underbrace{\sum_{i=1}^{N} Z_i^2}_{\chi_N^2}}}\sqrt{N} = \frac{Z}{\chi_N}\sqrt{N}$$

dove $Z$ e $Z_i$ sono v.c i.i.d $\sim G(0,1)$ è detta t di Student a N gradi di libertà



Andamento molto simile ad una gaussiana ma ha una varianza maggiore

Proprietà:

- Simmetrica
- $E[T_N] = 0$
- $Var[T_N] = \frac{N}{N-2}$
- Per $N \to \infty$, risulta che $T_N \xrightarrow[\text{in distribuzione}]{} G(0,1)$ (non per il teorema centrale del limite)

FINE PARENTESI

Si vede che:

la varianza è $\sigma^2/N$

effettuando una pseudonormalizzazione

$$\frac{M_1 - m}{\frac{S_c}{\sqrt{N}}} = \frac{M_1 - m}{S}\sqrt{N-1} = \frac{\frac{M_1 - m}{\sigma/\sqrt{N}}}{\sqrt{\frac{NS^2}{\sigma^2}}}\sqrt{N-1} = *$$

Moltiplico e divido per $\frac{\sigma}{\sqrt{N}}$ e ottengo Z

per il Teo di FISHER

$$S_c^2 = \frac{1}{N-1}\sum_{i=1}^{N}(X_i - M_1)^2$$

$$S^2 = \frac{1}{N}\sum_{i=1}^{N}(X_i - M_1)^2 = \frac{N-1}{N} S_c^2$$

$$S_c = \sqrt{\frac{N}{N-1}}\, S$$

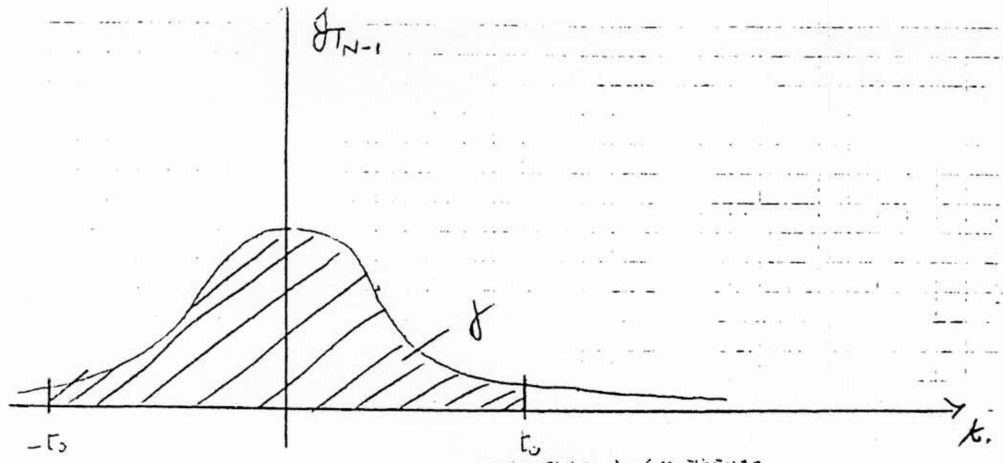
NB $S_c^2 = \frac{N}{N-1} S$ è la varianza campionaria corretta

Ho una t di Student

$$* = \frac{Z}{\sqrt{\chi_{N-1}^2}}\sqrt{N-1} = T_{N-1}$$

NOTA: ricordiamo che $M_1$ e $S^2$ sono indipendenti

1) Trovare $t_0$ t.c $P(|T_{N-1}| \leq t_0) = \gamma$




LIVELLO DI CONFIDENZA

$$P(-t_0 \leq T_{N-1} \leq t_0) = \gamma$$

$t_0$: ricavato dalle tabelle della $t$ di Student.

$$P\left(-t_0 \leq \frac{M_1 - m}{\frac{S_c}{\sqrt{N}}} \leq t_0\right) = \gamma$$

$$P\left(M_1 - t_0 \frac{S_c}{\sqrt{N}} \leq m \leq M_1 + t_0 \frac{S_c}{\sqrt{N}}\right) = \gamma$$

$$I_\gamma = \left[M_1 - t_0 \frac{S_c}{\sqrt{N}} \; ; \; M_1 + t_0 \frac{S_c}{\sqrt{N}}\right]$$

$\left( I_\gamma = \left[M_1 - z_0 \frac{\sigma}{\sqrt{N}} \; ; \; M_1 + z_0 \frac{\sigma}{\sqrt{N}}\right] \right.$, nel caso in cui era nota $\left. \sigma \right)$

se $\gamma = 0.95$

$z_0 = 1.96$

TABELLA prendo la colonna corrispondente a $1-\gamma = 0.05$

$t_0$ dipende da $N$

| | | $t_0$ |
|---|---|---|
| $N = 5$ | $\Rightarrow \nu = N-1 = 4$ | $\Rightarrow 2.776$ |
| $N = 10$ | $\Rightarrow \nu = N-1 = 9$ | $\Rightarrow 2.262$ |
| $N = 30$ | $\Rightarrow \nu = 29$ | $\Rightarrow 2.045$ |

Si nota che $t_0$ è sempre maggiore di $z_0$

$t_0 - z_0 \to 0$ per $N \to \infty$

(per $N$ "grande" posso usare $Z$ al posto di $T_{N-1}$).

ESEMPIO: SONDAGGIO ELETTORALE

- Quanto deve essere numeroso il campione?
- La numerosità in che modo dipende dal numero di elettori
- In un sistema uninominale le cose sono più facili o più difficili.

100'000 elettori di cui $X$ votano per CS

Interviste: prove di Bernoulli con $p = \frac{X}{100'000}$

Problema: stimare $p$ a partire dai risultati di $N$ prove di Bernoulli

Soluzione: Uso la frequenza relativa. $\hat{p} = \frac{K}{N}$ con $K$: n° di successi.

Intervallo di confidenza? (al 95%)

$$P(S_N = K) = \binom{N}{k} p^k (1-p)^{N-K} \qquad \text{(Binomiale)}$$

$S_N$: n° di successi in $N$ prove

$$E[S_N] = Np$$

$$Var[S_N] = Np(1-p)$$

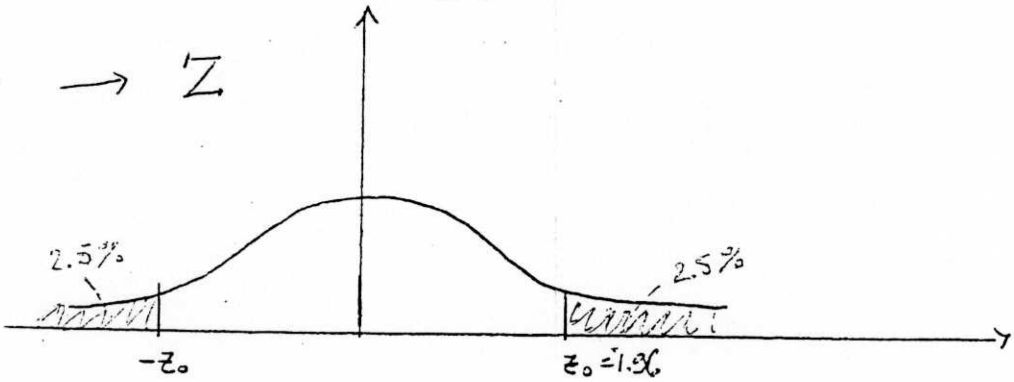
$$\hat{p} := \frac{S_N}{N} \Rightarrow E[\hat{p}] = p \qquad \text{(stimatore non polarizzato)}$$

$$Var[\hat{p}] = \frac{p(1-p)}{N}$$

$$Var[\hat{p}] = Var\left[\frac{S_N}{N}\right] = \frac{1}{N^2} Var[S_N] = \frac{p(1-p)}{N}$$

Per il teorema centrale del limite, $S_N \xrightarrow[N\to\infty]{\text{in distribuzione}}$ gaussiana

$$\Rightarrow \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{N}}} \longrightarrow Z$$

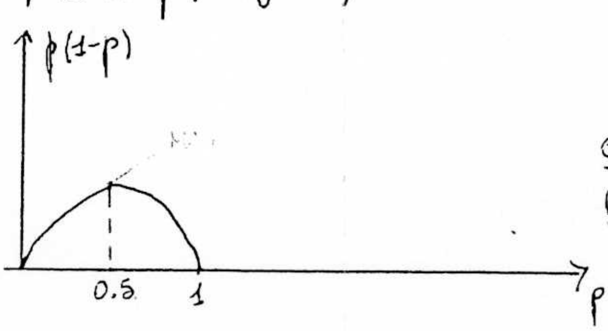



Per $N$ grande, nel 95% dei casi:

$$Z \cong \left| (\hat{p} - p) \frac{\sqrt{N}}{\sqrt{p(1-p)}} \right| < 1.96 \quad \text{ovvero}$$

$$|p - \hat{p}| \leq \underbrace{\frac{1.96\sqrt{p(1-p)}}{\sqrt{N}}}_{\text{dipende da } p \text{ (incognita)}}$$

OSSERVAZIONE




Caso peggiore (intervallo più lungo) per $p = 0.5$

$$|p-\hat{p}| \leq \frac{1.96\sqrt{p(1-p)}}{\sqrt{N}} < \frac{2\sqrt{0.5^2}}{\sqrt{N}}$$

nel 95% dei casi $\quad |p-\hat{p}| \leq \frac{1}{\sqrt{N}}$ ← quanto deve essere numeroso il campione

↳ la precisione costa

Per avere un errore $< 1\% \Rightarrow N \geq 10.000$

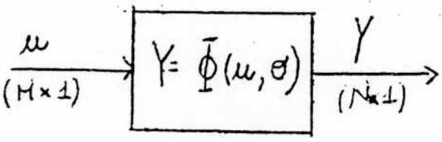
(e nel 5% dei casi può ancora andar male)

| $\gamma = 0.35$ | errore % |
|---|---|
| $N = 155$ | $\pm 8\%$ |
| $N = 200$ | $\pm 7\%$ |
| $N = 625$ | $\pm 4\%$ |
| $N = 2500$ | $\pm 2\%$ |
| $N = 10.000$ | $\pm 1\%$ |

- La numerosità del corpo elettorale non ha influenza sulla precisione (Pizzighettone = USA)
- Sistema maggioritario: ci vuole un sondaggio per ogni collegio. Se ci vuole precisione i costi aumentano rispetto al proporzionale

# IDENTIFICAZIONE DI MODELLI (STATICI).

non dinamici




$\theta = ?$
$(q \times 1)$

3 insiemi di variabili $u$, $Y$, $\theta$.

$Y$: variabili dipendenti, variabili misurate che voglio spiegare. (effetti)

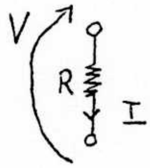
$u$: variabili indipendenti; variabili misurate oppure note a priori (cause) mediante le quali voglio spiegare gli effetti.

$\Phi(\cdot,\cdot)$ è il modello matematico che lega $u$ e $Y$.

$\theta$: parametri incogniti del modello

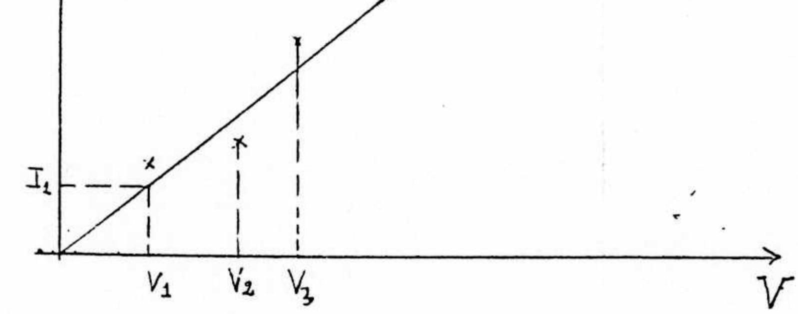
ESEMPIO:

$$I = \frac{V}{R}$$




$R = ?$

Impongo diversi valori di tensione $V$ ($V = V_1$, $V = V_2$, ..., $V = V_N$) e misuro $I$ ($I = I_1$, $I = I_2$, ..., $I = I_N$)




× punti reali.

La retta non si verifica nella realtà:
- errori di misura
- la relaz. fra $V$ e $I$ è lineare solo in un resistore ideale

Effetti
$$Y = \begin{bmatrix} I_1 \\ I_2 \\ \vdots \\ I_N \end{bmatrix}$$

Cause
$$u = \begin{bmatrix} V_1 \\ V_2 \\ \vdots \\ V_N \end{bmatrix}$$

Parametro incognito $\theta = \frac{1}{R}$ $\Rightarrow$ $\underbrace{\Phi(u, \theta) = u\theta}$

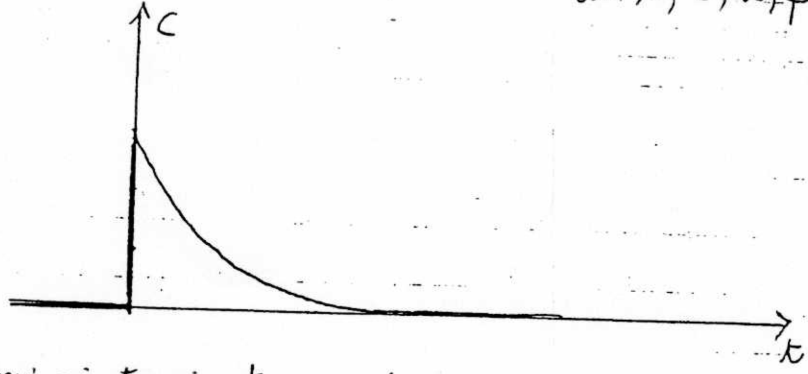
modello lineare nel parametro $\theta$.

Oppure $R$ $\Rightarrow$ $\Phi(u, \theta) = \frac{1}{\theta} u$ → modello non lineare nel parametro $\theta$

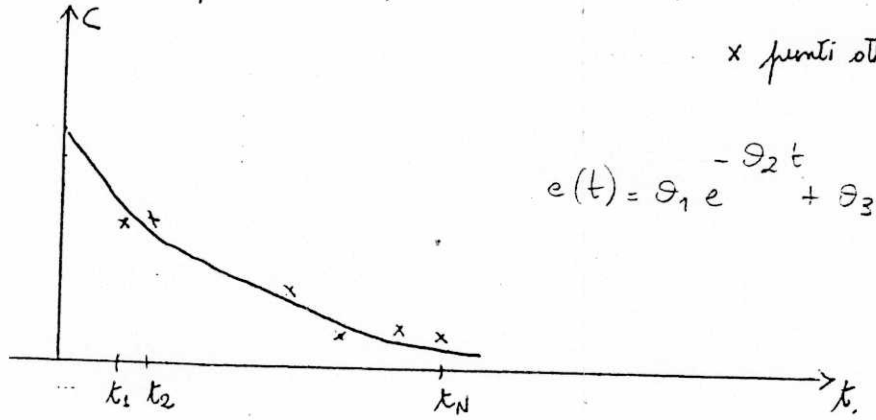
ESEMPIO: Supponiamo di sapere che la concentrazione plasmatica di un farmaco dopo una endovena evolve secondo la legge:

$$C(t) = a e^{-\alpha t} + b e^{-\beta t}$$

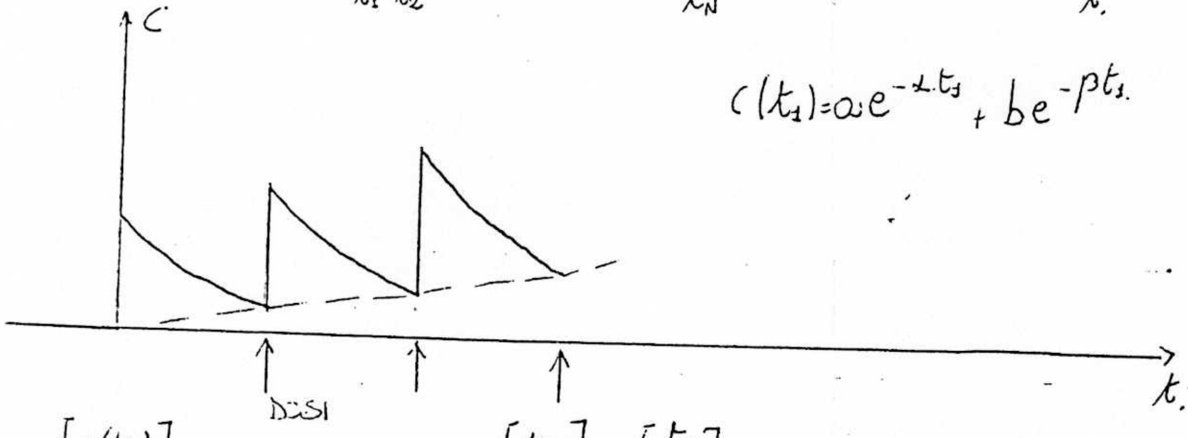
con $a$, $b$, $\alpha$, $\beta$ da stimare.



Effettuo dei prelievi ai tempi $t_1, \dots, t_N$




$$e(t) = \vartheta_1 e^{-\vartheta_2 t} + \vartheta_3 e^{-\vartheta_4 t}$$




$$C(t_1) = a e^{-\alpha t_1} + b e^{-\beta t_1}$$

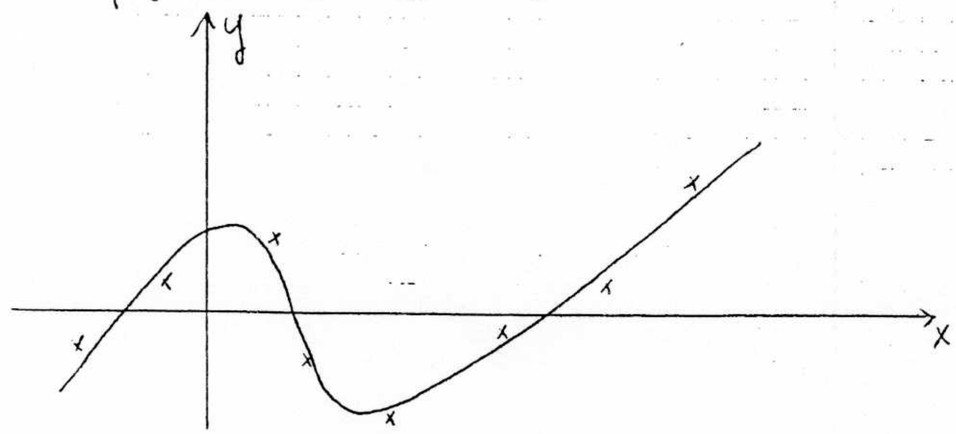
$$Y = \begin{bmatrix} C(t_1) \\ C(t_2) \\ \vdots \\ C(t_N) \end{bmatrix} \qquad u = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_N \end{bmatrix} = \begin{bmatrix} t_1 \\ t_2 \\ \vdots \\ t_N \end{bmatrix} \qquad \vartheta = \begin{bmatrix} a \\ \alpha \\ b \\ \beta \end{bmatrix} \begin{matrix} \to \vartheta_1 \\ \to \vartheta_2 \\ \to \vartheta_3 \\ \to \vartheta_4 \end{matrix}$$

$$\phi(u, \vartheta) = \begin{bmatrix} \vartheta_1 e^{-\vartheta_2 u_1} + \vartheta_3 e^{-\vartheta_4 u_1} \\ \vdots \\ \vartheta_1 e^{-\vartheta_2 u_N} + \vartheta_3 e^{-\vartheta_4 u_N} \end{bmatrix}$$

modello non lineare nei parametri
($\vartheta_1$ e $\vartheta_3$ compaiono linearmente mentre $\vartheta_2$ e $\vartheta_4$ sono all'esponente)

ESEMPIO : Voglio interpolare (approssimare) delle coppie di punti $(x_i, y_i)$ con una cubica

$$(y = a_0 + a_1 x + a_2 x^2 + a_3 x^3)$$




$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix} \qquad u = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix} = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_N \end{bmatrix} \qquad \vartheta = \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \end{bmatrix} = \begin{bmatrix} \vartheta_1 \\ \vartheta_2 \\ \vartheta_3 \\ \vartheta_4 \end{bmatrix}$$

$$\phi(u, \vartheta) =$$

$$y_1 = a_0 + a_1 x_1 + a_2 x_1^2 + a_3 x_1^3$$

$$\phi(u, \vartheta) = \begin{bmatrix} \vartheta_1 + \vartheta_2 u_1 + \vartheta_3 u_1^2 + \vartheta_4 u_1^3 \\ \vdots \\ \vartheta_1 + \vartheta_2 u_N + \vartheta_3 u_N^2 + \vartheta_4 u_N^3 \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & u_1 & u_1^2 & u_1^3 \\ & & & \\ 1 & u_N & u_N^2 & u_N^3 \end{bmatrix}}_{\phi(u)} \underbrace{\begin{bmatrix} \vartheta_1 \\ \vartheta_2 \\ \vartheta_3 \\ \vartheta_4 \end{bmatrix}}_{\vartheta}$$ modello lineare nei parametri

ESEMPIO : regressione lineare

Disponendo delle misure di $Q+1$ variabili $y(t), u_1(t), \dots u_q(t)$ per $t = 1, \dots N$, voglio spiegare $Y$ mediante una funzione lineare delle $u_j$

$$y(t) = \vartheta_1 u_1 + \vartheta_2 u_2 + \dots \vartheta_q u_q$$

regressione lineare della variabile $y$ sulle variabili $u_j$

$$Y = \begin{bmatrix} y(1) \\ y(2) \\ \vdots \\ y(N) \end{bmatrix} \qquad u = \begin{bmatrix} u_1(1) \\ u_1(2) \\ \vdots \\ u_1(N) \\ u_2(1) \\ \vdots \\ u_2(N) \\ \vdots \\ u_q(1) \\ \vdots \\ u_q(N) \end{bmatrix} \qquad \vartheta = \begin{bmatrix} \vartheta_1 \\ \vdots \\ \vartheta_q \end{bmatrix}$$

$$\Phi(u, \vartheta) = \Phi(u)\, \vartheta$$

$$\Phi(u) = \begin{bmatrix} u_1(1) & u_2(1) & \cdots & u_q(1) \\ \vdots & & & \vdots \\ u_1(N) & u_2(N) & \cdots & u_q(N) \end{bmatrix}$$

Def: <u>Un modello è lineare nei parametri se</u> $\boxed{\Phi(u, \vartheta) = \Phi(u)\, \vartheta}$

$\Phi = \Phi(u)$ <u>è detto matrice di sensitività</u>

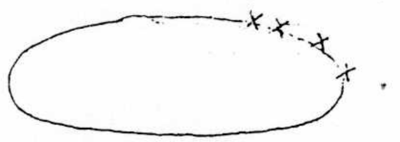
<u>minimi quadrati (LEAST SQUARES (LS))</u>

Cenni storici:

Gauss (1777 - 1855)

LS nel 1735.

Il 1° gennaio 1800 viene scoperto un asteroide: Cerere.



Da poche misure Gauss cerca di ricostruire l'orbita

↳ con i minimi quadrati e ci azzecca.

Nel 1807 diventa direttore dell'Osservatorio di Gottinga.

Nel 1809: " Teoria del moto dei corpi celesti "

(Boyer)

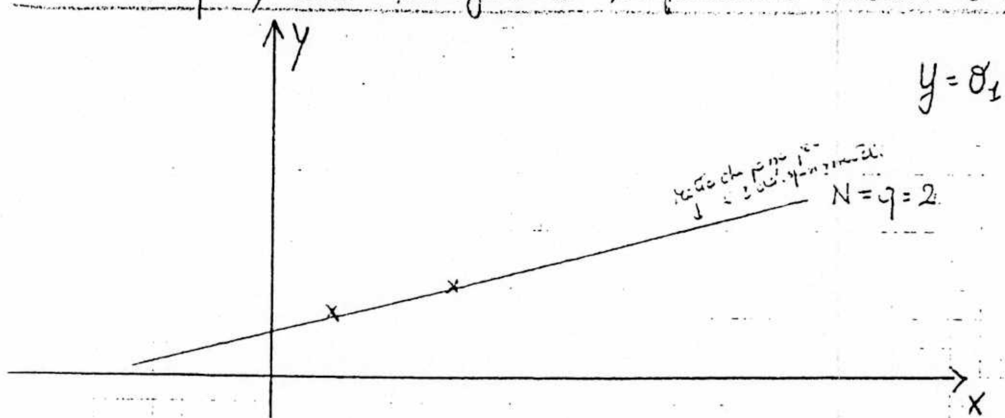
Fine dei cenni storici

D'ora in poi : $\boxed{q \leq N}$ ← <u>n° delle osservazioni disponibili</u>

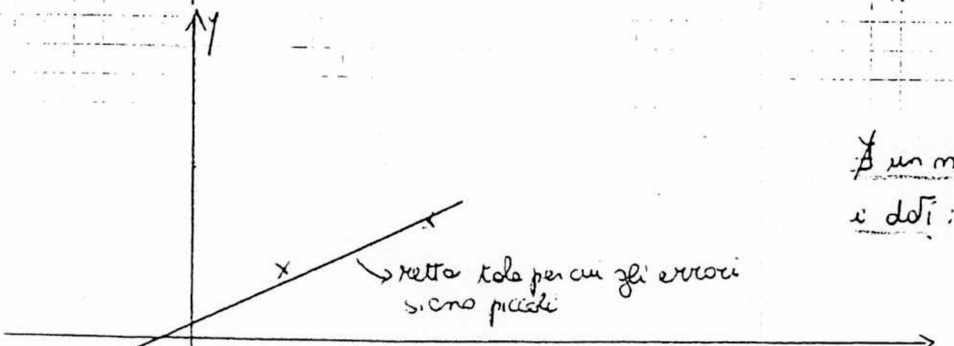
<u>n° di parametri del modello</u>

<u>Per $N > q$, sarà in generale impossibile trovare $\vartheta$ tale che $Y = \Phi\vartheta$.</u>

$$y = \vartheta_1 + \vartheta_2 x$$



$N = 3 > q = 2$



<u>∄ un modello che spiega perfettamente i dati</u>:
- ∃ errori di misura
- <u>nella realtà il legame non è mai perfettamente lineare</u>

↳ retta tale per cui gli errori siano piccoli

Allora, mi accorderò che l'errore

$$\boxed{\varepsilon := Y - \Phi\theta}$$ sia piccolo

$$\varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_N \end{bmatrix}$$

Per esempio posso chiedere che $\sum_{i=1}^{N} \varepsilon_i^2 = \|\varepsilon\|^2 = (Y-\Phi\theta)^T(Y-\phi\theta)$ sia piccol

Teorema: Si supponga che rank $[\Phi] = q$

($n^\circ$ di colonne di $\Phi$ linearmente indipendenti è uguale a $q$).

Allora, $J(\theta) := \|\varepsilon\|^2$ ha un minimo globale in corrispondenza di

$$\boxed{\theta^{LS} = (\Phi^T\Phi)^{-1}\Phi^T Y}$$

Dim:

PARENTESI: Funzioni matriciali

$$X = \begin{bmatrix} x_1 \\ \vdots \\ x_N \end{bmatrix} \qquad \underset{1\times 1}{f(x)}: \mathbb{R}^n \longmapsto \mathbb{R}^1$$

$$\frac{df}{dx} := \left[ \frac{\partial f}{\partial x_1} \ \frac{\partial f}{\partial x_2} \ \cdots \ \frac{\partial f}{\partial x_n} \right] \quad (1\times n)$$

$$\underset{(m\times 1)}{f(x)} = \begin{bmatrix} f_1(x) \\ f_2(x) \\ \vdots \\ f_m(x) \end{bmatrix} \qquad \frac{df}{dx} := \begin{bmatrix} \frac{df_1}{dx} \\ \frac{df_2}{dx} \\ \vdots \\ \frac{df_m}{dx} \end{bmatrix} = \underset{m\times n}{\begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & & \\ \frac{\partial f_m}{\partial x_1} & \cdots & \frac{\partial f_m}{\partial x_n} \end{bmatrix}}$$

- $\frac{d}{dx} x = I_n$

- $g(x), f(x) \in \mathbb{R}^{m\times 1}$

$$\frac{d}{dx}\underbrace{\left(g(x)^T f(x)\right)}_{\text{scalare}} = f(x)^T \frac{d}{dx} g(x) + g(x)^T \frac{d}{dx} f(x)$$

- $\frac{d}{dx}\left(f(x)^T f(x)\right) = 2 f(x)^T \frac{d}{dx} f(x)$
- $\frac{d}{dx} \underset{(m\times n)(n\times 1)}{(Ax)} = A$

$\bullet\ A = A^T$ $(n \times n)$ , $\frac{d}{dx}\underbrace{(x^T A x)}_{1\times 1} = x^T A + x^T A = 2x^T A$

$$\left[\frac{d^2 f(x)}{dx^2}\right]_{ij} := \left[\frac{\partial^2 f}{\partial x_i \partial x_j}\right]$$

$f(x): \mathbb{R}^n \to \mathbb{R}^1$

$$\frac{d^2 f}{dx^2} \in \mathbb{R}^{n \times n}$$

, $\frac{d^2(x^T A x)}{dx^2} = 2A$

Sviluppo di Taylor.

$$f(x) = f(x_0) + \left.\frac{d f(x)}{dx}\right|_{x=x_0}(x-x_0) + \frac{1}{2}(x-x_0)^T \left.\frac{d^2 f(x)}{dx^2}\right|_{x=x_0}(x-x_0) + \dots$$

Def. D é detto NADE se $D = A^T A$

Proprietà $\boxed{D = D^T \geq 0}$

Infatti, $x^T D x = x^T A^T A x = \|Ax\|^2 \geq 0 \quad \forall x$.