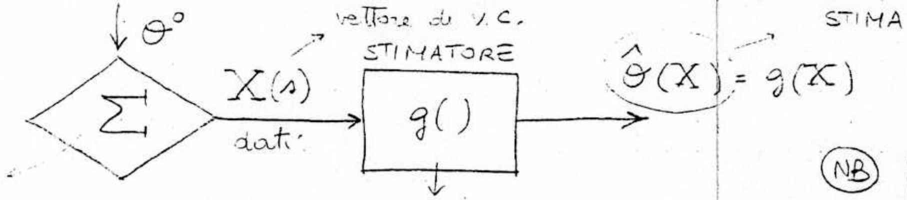
TEORIA DELLA STIMA

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 stimare $E[X]$ e $Var[X]$?

es Stima di λ 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

$g() = \frac{\sum x_i}{N}$ $\quad \hat{\theta} = \frac{\sum x_i}{N}$ = probabilità di avere Testa

L'esperimento Σ è caratterizzato da una densità congiunta $f_X^{\theta^\circ}(x)$

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 (σ) di una V.C. gaussiana

- $\theta^\circ = \begin{bmatrix} \theta_1^\circ \\ \theta_2^\circ \end{bmatrix} = \begin{bmatrix} m \\ \sigma \end{bmatrix}$
- Σ estrae N valori indipendenti della V.C. gaussiana
- $X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix}$ $\quad f_X^{(\theta^\circ)}(x) = \frac{1}{\sqrt{(2\pi)^N}(\theta_2^\circ)^N} \cdot e^{-\frac{1}{2}\sum_1^N \left(\frac{x_i - \theta_1^\circ}{\theta_2^\circ}\right)^2}$

la ddp è fz di θ°

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

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

Stimatore abbastanza buono xché obbedisce alla legge forte dei grandi numeri

lo vedremo in seguito

STIMA DEI MOMENTI DI UNA V.C. : MOMENTI CAMPIONARI

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

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

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

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

- 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:

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

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 $N \to \infty$

↓

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

MOMENTO DI ORDINE 2 (VALOR QUADRATICO MEDIO CAMPIONARIO)

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

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

PARENTESI: LA V.C. χ^2

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

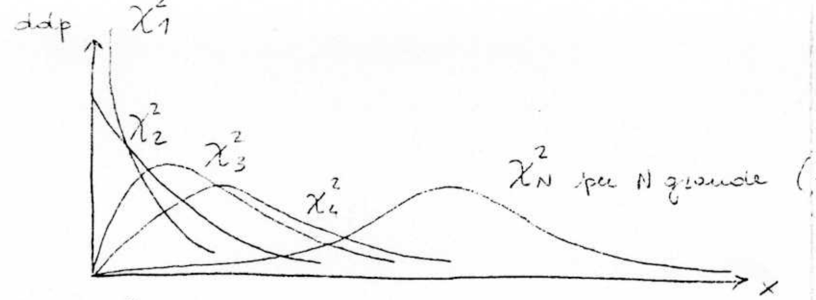
Allora costruisco una nuova v.c.:

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

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

↓

∃ un'intera famiglia di v.c. χ^2



Dovrebbe esserci una tabella del χ^2 ∀ valore di N

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

FINE PARENTESI

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

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

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

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

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

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

cvd

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

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

Più precisamente: $\quad \dfrac{NS^2}{\sigma^2} = \displaystyle\sum_{i=1}^N \underbrace{\left(\frac{x_i - M_1}{\sigma}\right)^2}_{\text{pseudonormalizzazione}} = \chi^2_{(N-1)}$

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).

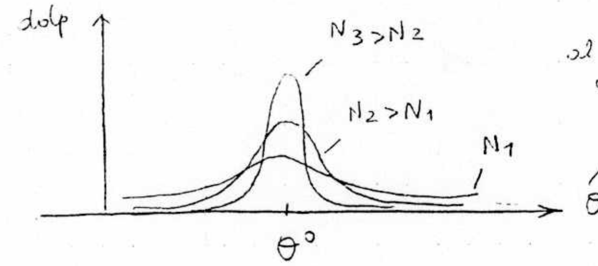
Cosa posso dire per M_2?

$Y := X^2 \Rightarrow M_2$ 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 θ^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 perché 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)

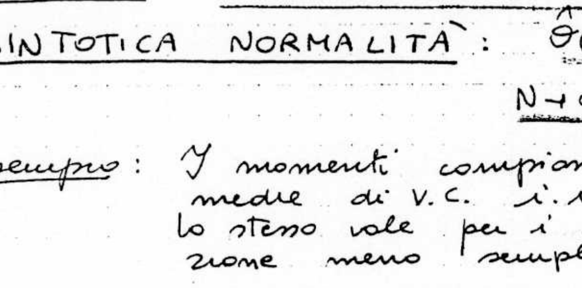
Considero uno stimatore non polarizzato ($E[\hat{\theta}] = \theta_0$) scalare. Allora, sotto condizioni di regolarità, vale

∃ 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. xché 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 ∃ perché 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 ⇒ varianza grande

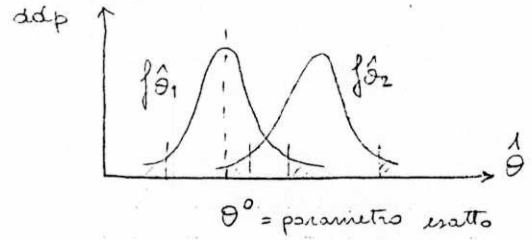
Esempi

- 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

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



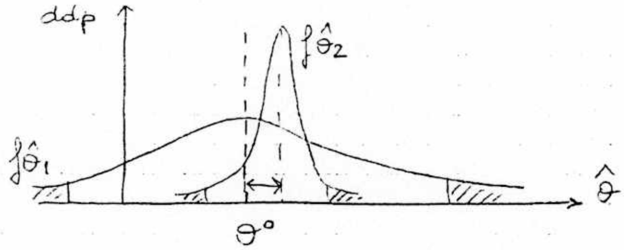
$f\hat{\theta}_1$ è NON polarizzato o CENTRATO perché il valor medio cade sul parametro vero

$f\hat{\theta}_2$ è polarizzato

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

Sul singolo caso può capitare che $\hat{\theta}_2$ sia + vicino a θ^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 θ^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 xché 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°

- <u>Estensione al caso vettoriale</u>

$S = \{ S_{ij} \}$ 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)

- <u>Nel caso di $\hat{\theta}$ polarizzato</u> <u>NON</u> 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

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

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

MEAN SQUARE ERROR

o

ERRORE QUADRATICO MEDIO

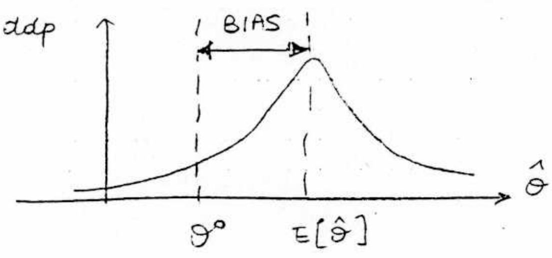
Si noti che $E[(\hat{\theta} - \theta^\circ)^2] = E[((\hat{\theta} - E[\hat{\theta}]) + (E[\hat{\theta}] - \theta^\circ))^2] =$

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

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

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

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



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

$Var[\hat{\theta}^m] \leq Var[\hat{\theta}]$, $\forall \hat{\theta}$ 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

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

$X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix}$ con x_i <u>iid</u>

ddp che hanno in comune (sono tutte =)

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

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

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

$\theta^\circ = m \qquad f^{(\theta)}(X_i) = \dfrac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{1}{2}\left(\frac{x_i-\theta}{\sigma}\right)^2}$ 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 $\sum_1^N \ldots$

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

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

24-4-2001

$(*)$ $$f \xrightarrow{X} \boxed{g} \rightarrow \hat{\theta}$$

θ = 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 perchè θ non è una V.C.

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

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

Per V.C. discrete $L(\theta, \bar{x}) := P^{(\theta)}(X = \bar{x})$

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

prove di Bernoulli (prob. di avere 4 successi)

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

Se la moneta è truccata e $\theta = 1 \Rightarrow L = 1$

" " è onesta e $\theta = \frac{1}{2} \Rightarrow L = \frac{1}{16}$

$\rightarrow$ 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 $\theta = 1 \Rightarrow L = 0$ $\rightarrow$ 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 perchè ciò che varia è θ, NON x

(parametro) (fisso)

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

$\rightarrow$ non ho usato $\pm\infty$ perchè potrebbe essere un integrale doppio e perchè potrebbe non $\exists$ per valori < 0

CRITERIO DELLA MAX VEROSIMIGLIANZA → Scuola dei FREQUENTISTI

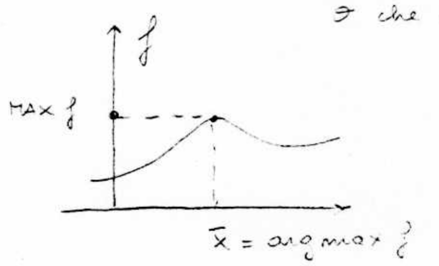
$$\theta^{ML}(X) = \arg\max_{\theta \in D_\theta} L(\theta, X)$$

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

MAXIMUM likelihood

cerco il valore di θ che massimizza $L(\cdot, X)$

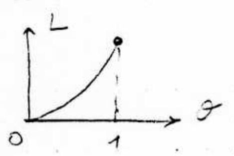
se, x es, non ammettessi valori < 0 non potrei accettare un $\bar{x} < 0$



$\bar{x} = \arg\max_i f$

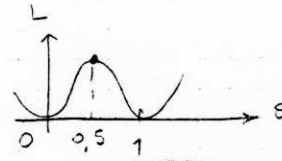
Esempio : Moneta, k = n° di teste, N lanci. $L(\theta, k) = \binom{N}{k} \theta^k (1-\theta)^{N-k}$ $\quad 0 \le \theta \le 1$ (D_θ)

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



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

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



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

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

NB
- 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,

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

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

Esempio : Dati N campioni indipendenti di una v.c. gaussiana, determina re lo stimatore congiunto ML di m e σ^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

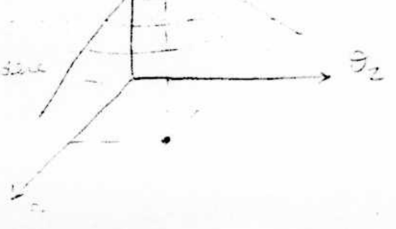
ddp congiunta dei dati sperimentali

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

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

media n camp.

dovrei verificare l'hessiana x vedere se è un max



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

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 λ e di $m := E[W_i]$ ▨ $\lambda=\frac{N}{\sum W_i}$ $m=\frac{\sum W_i}{N}$

<u>PROPRIETÀ DEGLI STIMATORI ML</u>

- <u>Sotto ipotesi di regolarità</u>, se $\eta=h(\theta)$ è un parametro funzione di θ (es: $\eta=\theta^3$) allora: $\underline{\eta^{ML}=h(\theta^{ML})}$ (es: $\eta^{ML}=(\theta^{ML})^3$)

  funzione

  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).

- <u>Sotto hp di regolarità θ^{ML} per campioni indipendenti</u> è <u>uno stimatore asintoticamente</u>:
  - <u>consistente</u>
  - <u>non polarizzato</u> (*)
  - <u>raggiunge il limite di C.R.</u> → tende a diventare imbattibile
  - <u>gaussiano</u>

  iid

  n° osservazioni → ∞

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

  <u>limite del metodo</u>: 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: θ è un parametro <u>DETERMINISTICO</u> ignoto

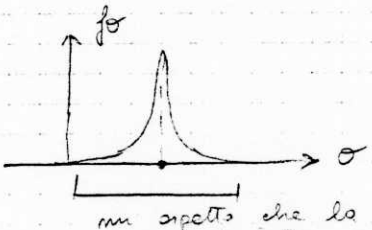
NON è una V.C.

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

26-4-2001

<u>Idea</u>: <u>Trattare θ come una V.C.</u>

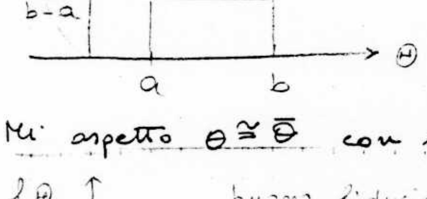
es:



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 $a\leq\Theta\leq b$ ma non ho motivi per preferire determinati valori.

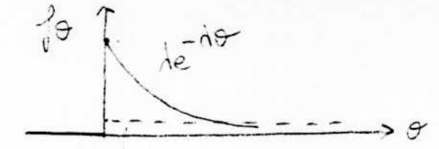


Mi aspetto $\theta\cong\bar{\theta}$ con un po' di incertezza

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

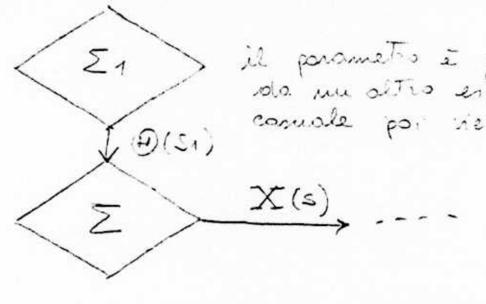


se conosco la media $\Rightarrow \lambda = \frac{1}{m}$

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

Σ_1: sorgente casuale caratterizzata da $f_\Theta(\theta)$

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

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

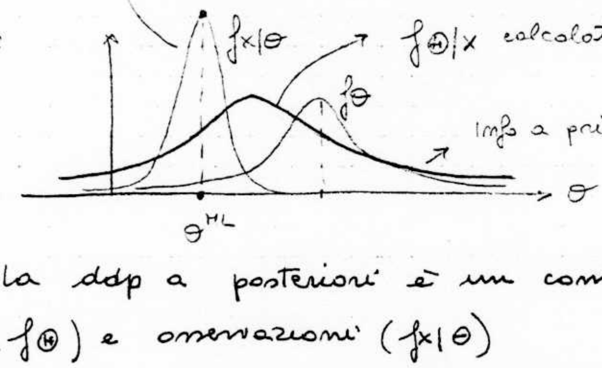
$$\underbrace{f_{\Theta|X}(\theta|X)}_{\text{Densità a posteriori}} \overset{\text{BAYES}}{=} \frac{\overbrace{f_{x|\Theta}(X|\theta)}^{\text{VEROSIMIGLIANZA}}\overbrace{f_\Theta(\theta)}^{\text{ddp a priori}}}{\underbrace{\int_{D\theta} f_{x|\Theta}(X|\theta) f_\Theta(\theta)\, d\theta}} \longrightarrow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

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 θ

evidenza sperimentale

es:



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

NOTO f_θ faccio l'esperimento e calcolo la verosimiglianza

Info a priori

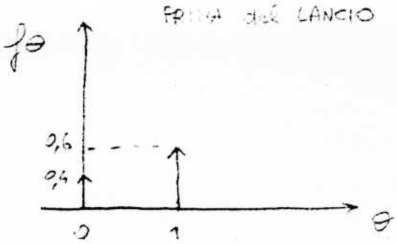
la ddp a posteriori è un compromesso tra informazioni a priori (f_Θ) e osservazioni ($f_{x|\theta}$)

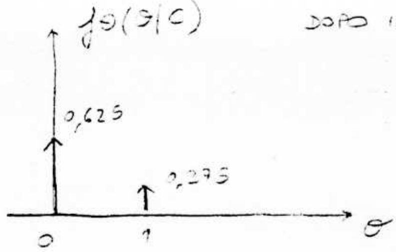
Esempio: In una scatola ci sono monete oneste (40%) e truccate (60%). Per quelle truccate $P(\text{Testa}) = 0{,}8$. Estraggo una moneta a caso. Qual è la prob. che sia truccata? Se dopo averla lanciata esce croce, quale è la prob. a posteriori che sia truccata

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

$P(c \mid \theta = 0) = 0{,}5 \qquad P(c \mid \theta = 1) = 0{,}2$

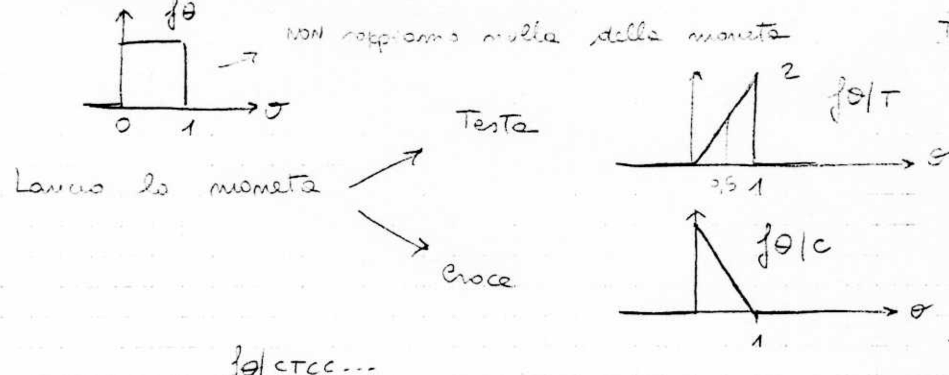
$$P(\Theta = 1 \mid c) = \frac{P(c \mid \Theta = 1)\, P(\Theta = 1)}{\underbrace{P(c \mid \Theta = 1) P(\Theta = 1) + P(c \mid \theta = 0) P(\theta = 0)}_{P(c)}} = 0{,}375$$

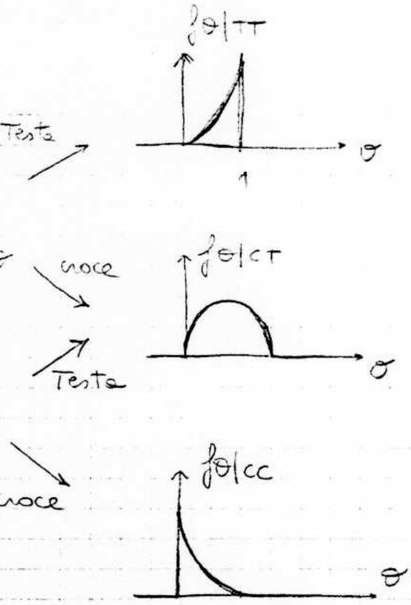




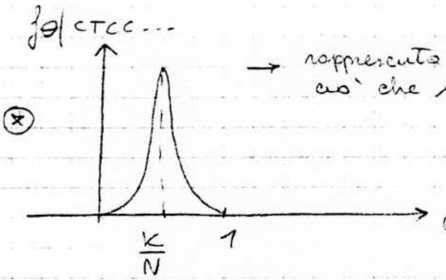
DOPO IL LANCIO (cambio opinione)

<u>Esempio</u>: moneta. Voglio stimare $p = P(T)$



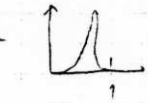


<u>Iterando</u>



$K =$ n° Teste su N lanci

6T su 10 lanci è ≠ da 600T su 1000 lanci ⇒ se N è molto grande si stringe

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

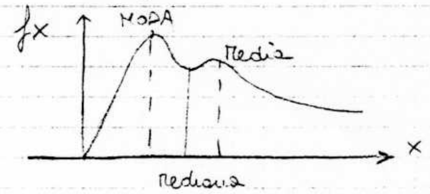
Solo se parto con 0 o 1 NON posso + correggere i miei pregiudizi

Come ottenere un pronostico dalla densità a posteriori? → pronostico

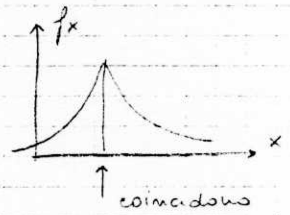
PROBLEMA: Una volta che conosco $f_{\theta|X}(\theta|X)$ come scelgo $\hat{\theta}$?

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?



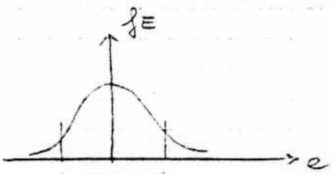
vengono usate tutte e 3
↓
posso scegliere una qualunque ma la prob. di avere esattamente un valore è $\underline{\varnothing}$
⇒ meglio usare l'errore (che è una v.c)



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 |·| piccolo ⇒ prob. elevata che e cada in un intervallo piccolo

(<u>minimizzo l'errore quadratico medio</u>)

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

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

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

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

c'è una certa preferenza per il valor medio (xché 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

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

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

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

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 medio)
(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) ⇒ $E[X \mid Y=y] = E[X]$ (perché se le 2 var. sono indip ⇒ la ddp di $x \mid y$ è uguale alla ddp di x ⇒ anche valor medio) ⇒ miglior stimatore che posso prendere è proprio il valor medio

$$E[(X-\hat{X}(y))^2 \mid Y=y] = E[(X-E[X])^2] = Var[X]$$

- 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 migliore resta il valor atteso condizionato).

$$\rightarrow \quad E[X \mid Y=y] = E[X] + V_{XY} V_Y^{-1} (y - E[Y])$$

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

(an1 an2 accanto a Y; fis1 geom accanto a X)

Espressione dello stimatore, errore quadr. medio:

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

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

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

Nel caso della gauss. è semplice:

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

↓
pseudo normalizzaz. della mia previsione

L'err. quadr medio che si commette è:

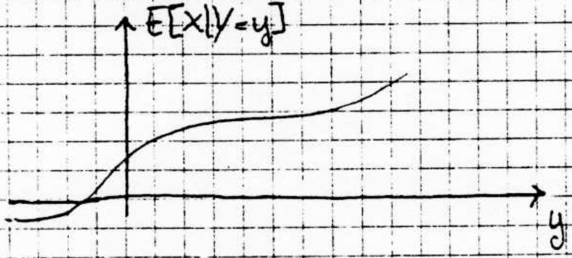
$$\rightarrow E\left[\left(X - E[X|Y=y]\right)^2 \middle| Y=y\right] = \sigma_x^2\left(1-r^2\right)$$

se trovo 2 var. fortissimam correlate fra loro ($r \cong \pm 1$) $\Rightarrow$ ~~ideale~~ ~~atteso~~ l'err. quadr. medio sarebbe molto piccolo (anz condizionato da a_{n1}) $\Rightarrow$ quando uso questo stimatore me ne accorgo, lo smaschero
(sì, aderenza su origine)

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 conosco $E[X|Y=y]$)
(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?

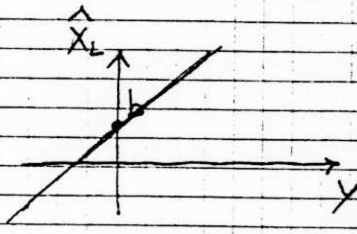
STIMATORE MS LINEARE

Caso non gaussiano: cerco una soluzione subottima per non dovere 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 $\hat{X}_L(Y) = aY + b$ (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: $\hat{X}_L(Y) = E[X] + V_{XY} V_Y^{-1} (Y - E[Y])$

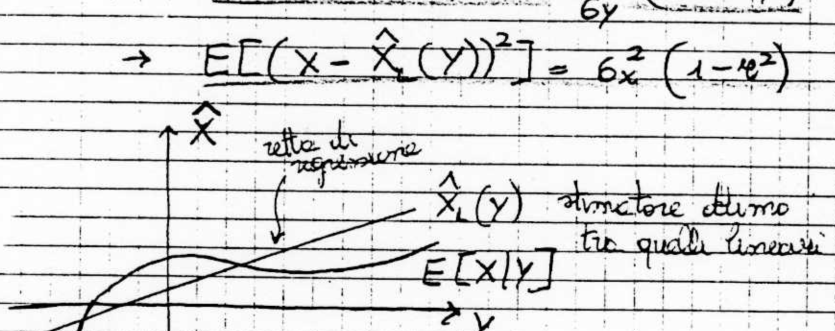
stimatore → $E[(X - \hat{X}(Y))(X - \hat{X}(Y))^T] = V_X - V_{XY} V_Y^{-1} V_{XY}^T$

(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, però devo dimostrarla in qualche modo, non solo da istogrammi; oppure sono + prudente e cerco uno stim. lineare e così trovo però uno