IDENTIFICAZIONE DEI MODELLI E ANALISI DEI DATI

Prof. G. De Nicolao

APPUNTI

I Parte

IMAD

Giuseppe De Nicolao - Dip. Info. e Sistemistica - piano C
Orario Ricevimento : Venerdì 13:45 - 15:45

Venerdì 11-12 (24-25)

[illegible] in SCALA

HODELLO : deve essere simile e diverso rispetto a ciò che voglio spiegare

Ci sono ≠ tipi di modelli :

- Modelli FISICI (dighe, automobili, aeroplani, edifici, navi ridotti in piccolo)
- Modelli "ANIMALI" (cavie)
- Modelli MATEMATICI : insieme di relazioni matematiche (equazioni) che descrivono quantitativamente le relazioni tra alcune grandezze di un fenomeno

  es: $F = m \cdot a$

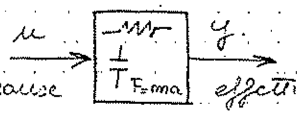
$\boxed{\text{PREGI}}$:
- Rigore deduttivo della matematica
- Versatile
- Poco Costosi (eccezione: modelli fluidodinamici che richiedono supercomputer)
- Funzionano!

Possibili USI dei Modelli :
- previsione
- Interpretazione, diagnosi, classificazione
- Simulazione
- Controllo
- Misure indirette

Come si costruiscono i modelli matematici ?

- Usando leggi fisiche

  si può pensare di costruire il modello come se fosse un puzzle trovando le leggi fisiche ∀ componente



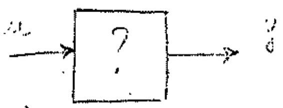
A partire dalle cause il modello deve prevedere gli effetti.

Detto " MODELLIZZAZIONE A SCATOLA TRASPARENTE" (white box)

DIFFICOLTÀ : - Sistemi troppo complessi (es: motore dell'automobile)

- Economia, Biologia, Ecologia ... (discipline in cui non c'è lo stesso formalismo di quelle + "matematiche").

- A partire dai dati



Detto " MODELLIZZAZIONE A SCATOLA NERA" (black box)

Motto: Dai dati al modello

- Approccio intermedio: conosco la legge fisica ma mi mancano i valori di alcuni parametri es: i pianeti di massa ignota ma che obbediscono alle leggi di gravitazione

ES: Motore elettrico di cui non conosco il coeff. di attrito
Uso i dati sperimentali x stimare i parametri (MODELLIZZAZIONE A SCATOLA GRIGIA o grey box)

La struttura del modello è dettata dalle leggi fisiche ma non conosco [illegible] dati

Diff tra black box e grey box: nel 2° conosco almeno la STRUTTURA del modello → a volte è difficile (det la complessità del modello con black box).

<u>IDENTIFICAZIONE</u> (dei sistemi o dei modelli) System Identification (Mode

↓

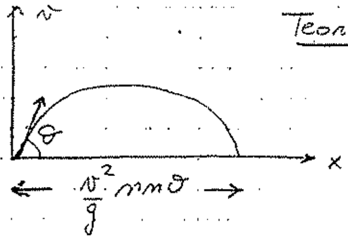
<u>Disciplina che si occupa della costruzione di modelli a partire da dati sperimentali</u>

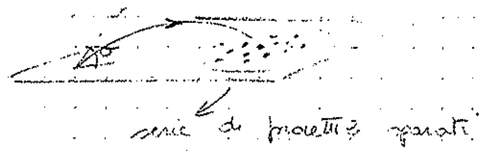
<u>PREGI</u>: <u>Non ho bisogno di conoscere leggi di natura</u>

<u>LIMITI</u>: i <u>modelli</u> black box fanno riferimento a <u>determinate</u> condizioni di funzionamento e <u>NON</u> possono avere validità generale

Esempio: "Caso e Necessità"

Proiettile <u>Teoria</u> <u>Realtà</u>





Realtà → Deviazioni dal modello che dipendono da:

- folate di vento
- resistenza dell'aria
- ....

Soluzioni: • Potrei complicare il modello MA NE VALE LA PENA?

• I fenomeni che non so (o non voglio) descrivere vengono attribuiti al "<u>CASO</u>"

<u>OBIETTIVI</u>

1) Culturale ("conoscenze"): Principi base di prob. e statistica

2) Progettuale ("abilità"): Saper effettuare una semplice analisi dati o costruire semplici modelli

<u>MODALITÀ D'ESAME</u> Teoria + Esercizi (3 esercizi + 3 domande)

2 prove in itinere + progetto = 1ª prova sulla 1ª parte (fondamenti di prob.)
Progetto sulla 2ª parte del corso
2ª prova sulla 3ª parte " "

12 pti x ciascuna prova in itinere + 6 per il progetto.

<u>LIBRI</u>:

- G. De Nicolao, R. Scattolini "Identificazione Parametrica" CUSL
- A. Papoulis "Probability, Random Variables and Stochastic Processes" McGraw Hill
- S. Bittanti Teoria della Predizione e del Filtraggio / Identificazione dei Modelli e Controllo Adattativo } Pitagora

7-3-2000 Compitino prima di Pasqua

1. PROBABILITÀ ← 1ª prova
2. IDENTIFICAZIONE DI MODELLI "STATICI" ← Progetto
3. PROCESSI CASUALI E IDENTIF. DI MODELLI DINAMICI ← 2ª Prova

SIGNIFICATO DELLA NOZIONE DI PROBABILITÀ

Almeno 3 (+1) possibili definizione

- CLASSICA $P(A) = \frac{N_A}{N}$

A: un evento (es: lancio un dado, $A = \{2\}$)

N: n° di esiti possibili (dado: $N=6$)

N_A: n° di esiti favorevoli ad A (dado: $N_A = 1$)

es Dado $N=6$ evento A: pari

Esiti favorevoli: $\{2, 4, 6\} \Rightarrow N_A = 3$ $P(A) = \frac{3}{6} = \frac{1}{2}$

Difficoltà: E se il dado è truccato? Potrei aggiungere: purché gli esiti siano equiprobabili

PB: Definizione circolare: cosa vuol dire equiprobabile se non so cosa è la probabilità?

- FREQUENZA RELATIVA: Si ripete l'esperimento N volte e sia NA il n° di volte che si verifica l'evento A

$$P(A) := \lim_{N \to \infty} N_A/N$$

→ pone rimedio al pb del dado truccato!

Difficoltà: il lim. non può essere calcolato in pratica (non posso fare ∞ prove) inoltre non è detto che il lim esista

[Vale x dadi truccati e NON

- ASSIOMATICA (del mate. russo Kolmogorov):

È un numero ≥0 associato agli eventi che rispetta certe regole:

POSTULATI
- $P(A) \geq 0$
- $P(\text{evento certo}) = 1$
- Se A e B si escludono a vicenda $P(A+B) = P(A) + P(B)$

Se $C = A + B \Rightarrow C$ è l'evento che è vero quando è vero A oppure B

es: $A = \{\text{pari}\}$ $B = \{1\}$ $\underset{\text{unione}}{A+B} = \{1, 2, 4, 6\}$

(NB) gli eventi sono tutto ciò su cui è possibile scommettere!

(NNB) la def. Assiomatica non richiede la ripetibilità di un evento (VANTAGGIO)

TEORIA DEDUTTIVA: se le probabilità hanno certi valori ⇒ ... varie regole
Dati dei postulati posso ottenere delle conclusioni

Problema: Collegamento con la realtà. Si fa risolvendo dei pb di stima. (STATISTICA)
↳ dice come analizzare la realtà a partire dalla teoria della probabilità.

- SOGGETTIVA (De Finetti, scuola bayesiana):

Prob = Misura del grado di fiducia (degree of belief) individuale riguardo il verificarsi di un evento.
(vale anche x eventi NON ripetibili)

Problema: come si misura?

Es: Paolo e Pietro sono 2 borseggiatori. Trovano una ricevuta del Toto nero che vale 1.000.000 se Roma vince lo scudetto. Pb: divisione del bottino.
Paolo da una parte mette la ricevuta e dall'altra parte la somma che secondo lui rappresenta il valore della ricevuta.
Pietro sceglie.

GLI ASSIOMI DELLA PROBABILITÀ

È definita su INSIEMI

- Insieme di tutti gli esiti possibili di un esperimento casuale: $\mathcal{S}$ (spazio degli esiti)

es: dado $\mathcal{S} = \{1, 2, 3, 4, 5, 6\}$

EVENTO: ciò che può essere oggetto di scommessa

es: dado → evento $A = \{2, 4, 6\}$

EVENTO: un sottoinsieme di $\mathcal{S}$

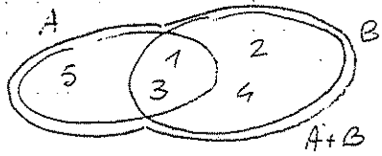
INSIEME DEGLI EVENTI: una classe $\mathcal{F}$ di sottoinsiemi di $\mathcal{S}$ che godono di 2 proprietà:

1) se $A \in \mathcal{F} \Rightarrow \bar{A} \in \mathcal{F}$

2) se $(A \in \mathcal{F})$ & $(B \in \mathcal{F}) \Rightarrow (A+B) \in \mathcal{F}$

PARENTESI: Unione ("somma") di eventi: $C = A+B$ è l'evento che si verifica se si verifica A oppure B

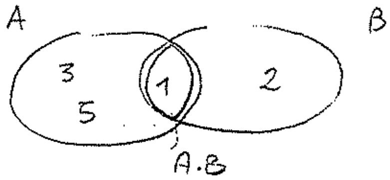
Es: $A = \{\text{dispari}\}$ $B = \{\leq 4\}$



$A+B = \{\leq 5\}$

Intersezione ("prodotto") di eventi: $C = A \cdot B$ è l'evento che si verifica se si verifica sia A che B

Es: $A = \{\text{dispari}\}$ $B = \{\leq 2\}$

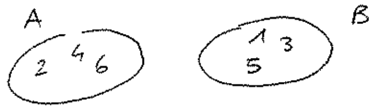


EVENTO NEGATO: $B = \bar{A}$ è l'evento che si verifica se e solo se non si verifica A

Es: $A = \{\text{dispari}\}$, $\bar{A} = \{\text{pari}\}$

EVENTI DISGIUNTI (mutuamente esclusivi): A e B sono disgiunti se non possono verificarsi simultaneamente $\Leftrightarrow AB = \{\emptyset\}$ insieme vuoto

Es: $A = \{\text{pari}\}$ $B = \{\text{dispari}\}$



EVENTO IMPOSSIBILE: è l'insieme vuoto $\{\emptyset\}$

1) Es $\{1\}$ evento $\Rightarrow \{2,3,4,5,6\}$ evento   cioè devo poter scommettere che non si verifichi $\{1\}$

2) Es $\{1\}$ evento & $\{2\}$ evento $\Rightarrow \{1,2\}$ evento

Conseguenza: $(A \in \mathcal{F})$ & $(B \in \mathcal{F}) \Rightarrow AB \in \mathcal{F}$

In particolare l'insieme $\mathcal{F}$ deve contenere sempre:

$\mathcal{S}$: evento certo $\{\emptyset\}$: evento impossibile

$P=1$

Es: Partita di calcio. Trovare S ed $\mathcal{F}$

$S = \{1, x, 2\}$ $\quad \mathcal{F} = \{\{\phi\}, \{1\}, \{x\}, \{2\}, \{1, x\}, \{1, 2\}, \{x, 2\}, S\}$

$P=0$ $\quad \{1, x, 2\}$

DEF. ASSIOMATICA DI PROBABILITÀ

P è una funzione def. sull'insieme degli eventi in modo tale che valgano le seguenti regole: (se $A \in \mathcal{F}$, $B \in \mathcal{F}$)

① $P(A) \geq 0$

② $P(S) = 1$

③ se $AB = \{\phi\} \Rightarrow P(A+B) = P(A) + P(B) \rightarrow$ propietà di ADDITIVITÀ SEMPLICE

A e B disgiunti

COROLLARI

a) $P(A) = 1 - P(\bar{A}) \leq 1$

b) $P(\{\phi\}) = 0$

c) $B \subset A \Rightarrow P(B) \leq P(A)$

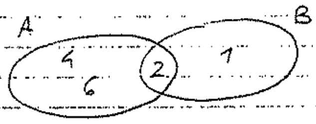
d) se $A_1, A_2 .. A_n$ sono disgiunti $\Rightarrow P(A_1 + A_2 + .. A_n) = P(A_1) + P(A_2) + .. P(A_n)$

e) $P(A+B) = P(A) + P(B) - P(AB) \leq P(A) + P(B) \rightarrow$ vale se i 2 eventi non sono disgiunti

Es: dado. Calcolare la prob di un n° pari o minore di 3

$A = \{2, 4, 6\}$ $\quad B = \{1, 2\}$ $\quad C = \{\text{pari o} < 3\} = A + B$

$P(C) = P(A+B) = P(A) + P(B) - P(AB)$



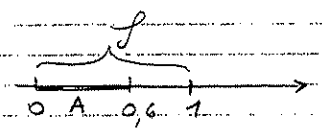
$AB = \{2\}$ $\quad P(A) = \frac{1}{2}$ $\quad P(B) = \frac{1}{3}$ $\quad P(AB) = \frac{1}{6}$

$P(C) = \frac{1}{2} + \frac{1}{3} - \frac{1}{6} = \frac{2}{3}$

SPAZI CON INFINITI RISULTATI

Es: S: n° reali tra 0 e 1

Possibile evento: $A = \{0 \leq x \leq 0,6\}$



Chiedo che $\mathcal{F}$ soddisfi un'ipotesi aggiuntiva

②' $(A_i \in \mathcal{F}, i = 1, 2 .. \infty) \Rightarrow \sum_{i=1}^{\infty} A_i \in \mathcal{F}$

NB L'additività semplice NON implica quella completa.

Inoltre chiedo che P soddisfi

③' Se A_i con $i = 1 .. \infty$ sono disgiunti e $A := \sum_{i=1}^{\infty} A_i \Rightarrow P(A) = \sum_{i=1}^{\infty} P(A_i)$

ADDITIVITÀ COMPLETA

Questo postulato NON è obbligatorio (la scuola soggettiva lo rifiuta).

8-3-2001

Esperimento casuale:

- S spazio degli esiti
- $\mathcal{F}$ insieme degli eventi
- P probabilità def su $\mathcal{F}$

SPAZI DEGLI ESITI EQUIPROBABILI

CARDINALITÀ FINITA → n° di esiti finito

Def: S è equiprobabile se, essendo $S = \{s_1, s_2 .. s_n\}$ s_i = evento elementare, la probabilità di un evento A composto da r_A eventi elementari è:

$$P(A) = \frac{r_A}{n}$$ def classica

CARDINALITÀ INFINITA

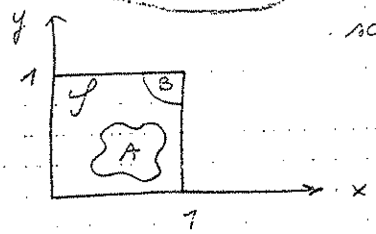
Def: S è equiprobabile se: → legata al concetto di uniformità

(i) su di esso è def una misura geometrica (lunghezza, area ...)

(ii) $m(S) \neq 0$ misura geometrica su S

(iii) $P(A) = \frac{m(A)}{m(S)}$

Es:



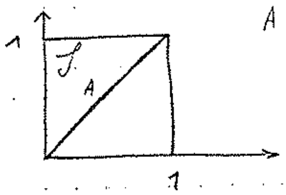
scegliere un pto a caso nel quadrato, scegliere un pto a caso di A

$m(S) = 1$ è l'area del quadrato

$P(A) = \frac{m(A)}{1}$ → rapporto tra l'area di A e l'area del quadrato

$P(B) = \frac{m(B)}{1}$

FATTO INTERESSANTE: Hp: S equiprobabile $P(A) = ?$



A è la diagonale

Dato che S è equiprobabile $P(A) = \frac{m(A)}{m(S)} = 0$

perchè l'area di un segmento è zero

$\Rightarrow P(\bar{A}) = 1 - P(A) = 1$

$P(A) = 0 \not\Rightarrow A$ impossibile cioè A impossibile $\Rightarrow P(A) = 0$, $\not\Leftarrow$

$P(B) = 1 \not\Rightarrow B$ certo B certo $\Rightarrow P(B) = 1$, $\not\Leftarrow$

$B = \bar{A}$

ESERCIZIO: 2 Dadi. Qual'è la prob. di avere almeno un "1" $= \frac{11}{36}$

perchè $= 1 - P(\text{neanche un } 1) = 1 - \frac{5}{6} \cdot \frac{5}{6}$

PROBABILITÀ CONDIZIONATA

Def: Hp: $P(M) \neq 0$ La prob. di A condizionata da M è data da

$$P(A|M) := \frac{P(AM)}{P(M)}$$

$P(AM)$ → evento congiunto

$P(M)$ → evento condizionante

Es: 52 carte. P(regina di picche) = ? So che è stata pescata una carta nera:

$P(Q \text{ picche} | \text{nera}) = ?$

$A = Q$ di picche M = carta nera AM = (carta nera) & (Q di picche) = Q di picche

$$P(A|M) = \frac{P(AM)}{P(M)} = \frac{P(A)}{P(M)} = \frac{1/52}{1/2} = \frac{1}{26}$$

TEOREMA DI BAYES

detto TEOREMA DELLA PROBABILITÀ DELLE CAUSE

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Dimostrazione

$$P(A|B) = \frac{P(AB)}{P(B)} \qquad P(B|A) = \frac{P(AB)}{P(A)} \implies P(AB) = P(B|A)P(A)$$

$$\downarrow$$

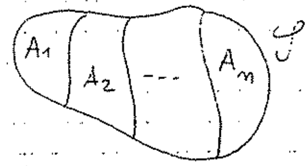
$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \qquad \text{c.v.d.}$$

VANTAGGIO del TEO: Permette di rovesciare il condizionamento!!

TEO DELLA PROB. TOTALE

Supponiamo di avere una partizione di S (A_i disgiunti t.c. $\sum_{i=1}^{N} A_i = S$) $\bigcup_{i=1}^{n} A_i$

Allora $P(A) = \sum_{i=1}^{n} P(A|A_i)P(A_i)$



es: $n=3$ $\quad P(A) = P(A|A_1)P(A_1) + P(A|A_2)P(A_2) + P(A|A_3)P(A_3)$

Posso usarlo x valutare il denominatore del Teo. di Bayes

$$P(A_i|A) = \frac{P(A|A_i)P(A_i)}{P(A)} = \frac{P(A|A_i)P(A_i)}{\sum_{j=1}^{n} P(A|A_j)P(A_j)}$$

Esercizio Scatola con 3 tipi di oggetti

A: 2500 di cui 10% difettoso
B: 500 " " 40% "
C: 1000 " " 30% "

1) Qual'è la prob che l'oggetto pescato sia difettoso? $P(D) = ?$
2) Dato l'oggetto qual'è la prob che sia A, B o C? $P(A|D) = ?$ $P(B|D)$ $P(C|D) = ?$

$$P(D) = P(D|A)P(A) + P(D|B)P(B) + P(D|C)P(C)$$

$$= 0{,}1 \cdot \frac{25}{40} + 0{,}4 \cdot \frac{5}{40} + 0{,}3 \cdot \frac{1}{4} = \frac{750}{4000} = \frac{3}{16}$$

$$P(A|D) = \frac{P(D|A)P(A)}{P(D)} = \frac{0{,}1 \cdot 25/40}{75/400} = \frac{1}{3}$$

$$P(B|D) = \frac{P(D|B)P(B)}{P(D)} = \dots = \frac{4}{15} \qquad P(C|D) = 1 - P(A|D) - P(B|D) =$$

$$= \frac{P(D|C)P(C)}{P(D)} = \frac{2}{5}$$

INDIPENDENZA

Def: A e B si dicono indipendenti se $P(AB) = P(A)P(B)$

Interpretazione: Supponiamo che A, B indipendenti $P(A|B) = \frac{P(AB)}{P(B)} =$

$$= \frac{P(A)P(B)}{P(B)} = P(A)$$

Esempio: Scatola con 5 palline rosse numerate e 3 palline nere numerate

$A = \{\text{pallina rossa}\}$ $\quad B = \{\text{pallina n° 3}\}$ A e B sono indipendenti?

$P(A) = 5/8$ $\quad P(B) = 2/8 = 1/4$ $\quad P(AB) = 1/8$ $\rightarrow P(AB) \neq P(A)P(B) \Rightarrow$ A e B non sono indipendenti (se fossero 5 nere $\Rightarrow$ A e B indipendenti)

INDIPENDENZA TRA PIU' EVENTI

Def: $A_1, A_2 \dots A_n$ sono (statisticamente) indipendenti se

$P(A_iA_j) = P(A_i)P(A_j) \quad \forall i,j \dots i \neq j$

$P(A_iA_jA_k) = P(A_i)P(A_j)P(A_k) \quad \forall i,j,k$ diversi

$P(A_1 A_2 \dots) = P(A_1)P(A_2) \dots P(A_n)$

Controesempio : eventi indip. a coppie ma non tutti insieme

2 Dadi $A = \{1°\text{ dado dispari}\}$ $B = \{2°\text{ dado dispari}\}$

$C = \{\text{somma dei dadi dispari}\}$

$P(ABC) = P(\{\emptyset\}) = 0 \neq P(A)P(B)P(C)$

PROVE RIPETUTE = esperimenti composti in cui si ripetono N volte eventi con 2 possibili esiti

Un esperimento che consiste nel ripetere N volte una prova con 2 possibili risultati (successo/insuccesso).

Def: A_i : il risultato dell'i-ma prova è un successo

Def: PROVE DI BERNOULLI

(i) $P(A_i) = P(A_j) \quad \forall i, j$ la prob. di successo è sempre la stessa

(ii) A_i indipendenti tra di loro. Ciò implica che $P(A_i A_j) = P(A_i)P(A_j)$

D'ora in poi $p = P(A_i)$ cioè prob. di successo

$q = P(\bar{A}_i) = 1 - p$ → Conoscere p vuol dire conoscere tutto (x Bernoulli)

TEOREMA La probabilità $p_n(k)$ di avere k successi su n prove di Bernoulli è :

$$p_n(k) = \binom{n}{k} p^k q^{n-k}$$

NOTA : $\binom{n}{k} = \frac{n!}{k!(n-k)!} = \frac{n(n-1)\cdots(n-k+1)}{k(k-1)(k-2)\cdots 2\cdot 1}$

9-3-2001

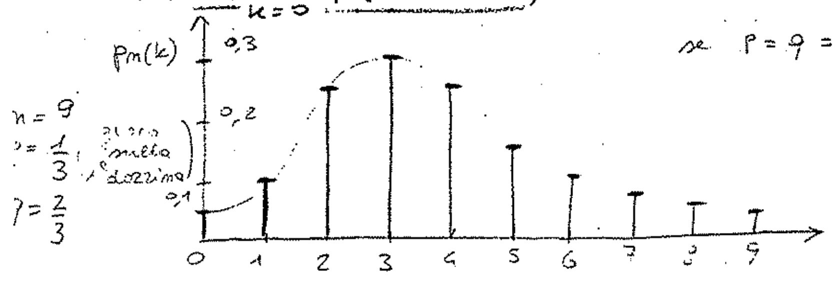
Es 4 lanci di dado. Trovare la prob. che il "6" esca 2 volte

$n = 4 \quad k = 2 \qquad p_4(2) = \binom{4}{2}\left(\frac{1}{6}\right)^2\left(\frac{5}{6}\right)^2 = \frac{4\cdot 3}{2\cdot 1}\left(\frac{1}{6}\right)^2\left(\frac{5}{6}\right)^2$

NOTA : $p_n(k)$ viene detta BINOMIALE di ordine n poiché coincide con il k-mo termine dell'espansione della potenza del binomio $(p+q)^n$

$$\underset{\substack{\parallel \\ 1}}{(p+q)^n} = \sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k} = \sum_{k=0}^{n} p_n(k)$$

$$1 = \sum_{k=0}^{n} P(k \text{ successi})$$



se $p = q = 0{,}5 \Rightarrow$ campana simmetrica

il MAX della prob si raggiunge per $k = \text{int}[(n+1)p]$

Es : Roulette onesta senza zero → vuol dire che $\int$ è equiprobabile

Calcolare la prob. che il rosso esca per la 1ª volta al (k+1)-esimo tentativo. Valutare per $k = 5$

$S_1 :=$ numero d'ordine del 1° successo (se il rosso esce per la 1ª volta al terzo lancio $\Rightarrow S_1 = 3$).

$P(S_1 = k+1) = ? = P(\underbrace{\text{insuc, insuc, insuc} \ldots \text{insuc, successo}}_{\text{PROVE DI BERNOULLI}}) =$ → eventi congiunti

$= P(\text{insucc})\, P(\text{insucc}) \cdot P(\text{insucc}) \cdots P(\text{insuc}) \cdot P(\text{successo}) =$

$= \underbrace{0,5 \cdot 0,5 \cdot 0,5 \cdots \cdot 0,5}_{k} \cdot 0,5 = (0,5)^{k+1}$ in generale sarebbe

$= q \cdot q \cdot q \cdot \cdots \, q \cdot p = q^k p$

Se $k=5$, $p=0,5 \Rightarrow 0,5^6 = 0,015625$

Es: Calcolare la prob. che il rosso esca al $(k+1)$-esimo lancio sapendo che nei primi k lanci è uscito sempre il nero (ritardo)

Uso la prob. condizionata (def)

$$P(S_1 = k+1 \mid S_1 > k) = \frac{P(S_1 = k+1, S_1 > k)}{P(S_1 > k)} = \frac{P(S_1 = k+1)}{P(S_1 > k)} = \frac{q^k p}{q^k} = p = 0,5$$

$S_1 > k := \underbrace{\text{insucc, insucc, insucc, insucc} \ldots}_{k}$

Esercizio

Emofilia che colpisce gli uomini mentre le donne sono portatrici sane. I figli di un uomo sano e una donna portatrice hanno il 50% di prob. di ereditare il germe.
La signora X pur avendo il padre ~~malato~~ sano ha un fratello emofiliaco ($\Rightarrow$ madre portatrice)
la signora X ha 2 figli maschi sani avendo sposato un uomo sano.

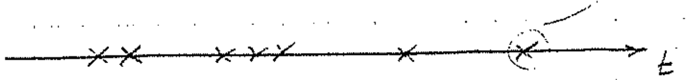
usare Bayes

1) Prob(X sia portatrice) = ?

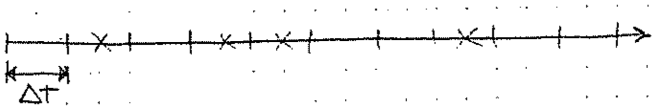
2) Prob (3° figlio maschio sia emofiliaco) = ? $\left(\frac{1}{2}\right)^2 \frac{1}{2} = 1/8$?

EVENTI DI POISSON No

eventi nel tempo (casuali)



Voglio simulare degli eventi casuali con frequenza media λ [eventi / unità di tempo]



faccio un esperimento casuale di Bernoulli + vedere se nell'intervallo Δt si verifica un evento.

$p = \lambda \cdot \Delta t$ prob. parametrica in Δt

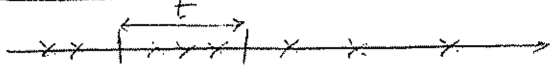
λ è detta DENSITÀ

Oppure diamo una prob ad ogni intervallo (es: $p = 0,9$) e poi usiamo la fz random. x avrà un valore tra 0 e 1 (con distribuzione uniforme) Se $x < p \Rightarrow a$, $x > p \Rightarrow \bar{a}$ (simulazione stocastica)

Facendo il limite per $\Delta t \to 0$ ottengo gli EVENTI DI POISSON (gli eventi possono accadere in qualsiasi istante di tempo).

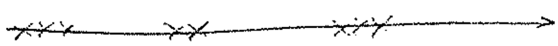
UNIFORMITÀ: non c'è nessuna parte dell'asse reale privilegiata

es

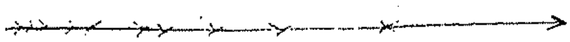


POISSON

$N(t) = 3$ (*)



no Poissoniano xchè gli eventi si verificano sempre a gruppi $\Rightarrow$ non è bernoulliano



no Poissoniano xchè gli eventi si diradano

TEOREMA La prob. che nell'intervallo di lunghezza t cadono k eventi:

$\longrightarrow P(N(t)=k) = \frac{(\lambda t)^k}{k!} e^{-\lambda t}$

Es: Nascite in clinica come eventi di P. con $\lambda = 10$ neonati / giorno

a) Prob. di avere 0, 1, 2 neonati in 1 ora

b) Prob. di avere almeno una nascita nel 1° minuto dell'anno

a) $t = \frac{1}{24}$ $k=0 \Rightarrow P(0) = 0,659$ $k=2 \Rightarrow P(2) = 0,057$

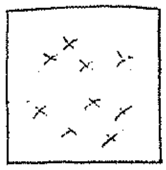
$k=1 \Rightarrow P(1) = 0,274$

b) $p(1), p(2), p(3) \ldots$ serie VS $p(0)$ $\frac{10}{24 \cdot 60}$

$P(\text{0 neonati nel 1° minuto}) = \frac{(\lambda t)^0}{0!} e^{-\lambda t} = e^{-\lambda t} = 0,99306$

$p(\text{almeno 1 neonato nel 1° MIN}) = 1 - e^{-\lambda t} = 0,00694$

EVENTI DI POISSON NEL PIANO



Es Bombardamento di Londra

$\lambda = 3,729$ bombe / km²

Prob di 0, 1, ≥2 bombe in un km² ?

$P(N(S)=k) = \frac{(\lambda S)^k}{k!} e^{-\lambda S}$ $S=1$ xché è una superficie di 1 km²

$P(N(S)=0) = 0,0240$ $P(N(S)=1) = 0,0895$ $P(N(S) \geq 2) = 0,8864$

13-3-2001

VARIABILI CASUALI

Un esperimento casuale i cui risultati (esiti) sono numeri reali.

es: moneta $S_1 = \{T\}$ $X(S_1) = 1$ } funzione def. sullo spazio degli esiti.
$S_2 = \{C\}$ $X(S_2) = 0$

D'ora in poi indichiamo una generica variabile casuale come $X(\cdot)$

Se ho un particolare esito $\bar{s}$, allora $X(\bar{s})$ indica il valore numerico corrispondente

Invece $X(\cdot)$ indica la FUNZIONE che, def. su $\mathcal{S}$, mi restituisce un n° reale

Esperimento casuale $= \{\mathcal{S}, \mathcal{F}, P(\cdot)\}$

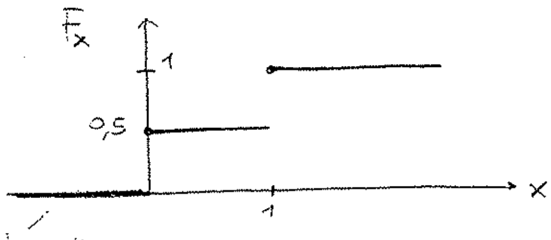
Idea: adesso voglio ragionare più comodamente usando i numeri reali.

DESCRIZIONE DI UNA VAR. CASUALE

Def: Funzione di distribuzione (di probabilità) F_X della var. casuale $X(\cdot)$

$F_X(x) = P(\underbrace{X}_{\text{var. cas.}} \leq \underbrace{x}_{\text{n° qualsiasi}})$

Es: Moneta $X(\{T\}) = 1$ $X(\{C\}) = 0$



prob nulla

$F_X(0) = P(X \leq 0) = P(\{C\})$

$F_X(0,5) = P(X \leq 0,5) = P(X=0) = 0,5$

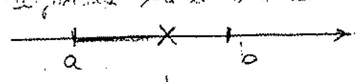
$F_X(1) = P(X \leq 1) = 1$

Nota la f.z di distribuzione conosco tutto

Es: Punto t scelto in modo equiprobabile nell'intervallo [a, b]. Poniamo

es. telefonata fra le 15 e le 16

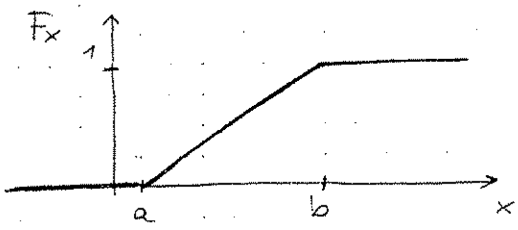
$X(t) = t$



esperimento casuale con un n° ∞ di risultati

$$P(X \le x) = \frac{x-a}{b-a}$$

dato che è equiprobabile devo fare il rapporto fra le misure



PROPRIETÀ

1. $F(+\infty) = 1, \quad F(-\infty) = 0$
2. $x_1 < x_2 \Rightarrow F(x_1) \le F(x_2)$ (MONOTONIA)
3. $P(X > x) = 1 - F(x)$?
4. $F(x^+) = F(x)$ (continua da destra) ¿F?
5. $P(x_1 < X \le x_2) = F(x_2) - F(x_1)$
6. $P(X = x) = F(x) - F(x^-)$ (caso della moneta) → la prob. che Tizio telefoni proprio alle 15:30 è = 0
7. $P(x_1 \le X \le x_2) = F(x_2) - F(x_1^-)$

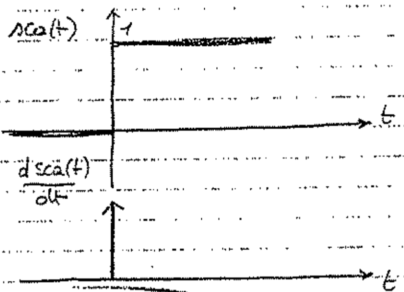
Quando la fz di distribuzione è a gradini ⇒ la var casuale è discreta VS è continua

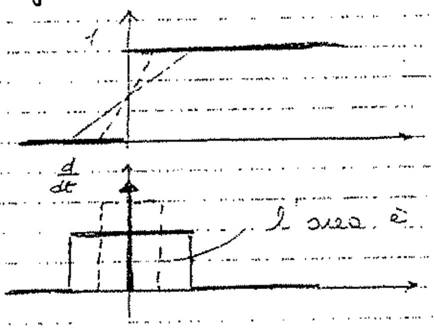
Def: $f_X(x) := \dfrac{dF_X(x)}{dx}$ è detta DENSITÀ DI PROBABILITÀ della var. casuale X

(PB) Nel caso continuo tutto bene. Ma ho pb con la derivazione di v. discrete

DELTA DI DIRAC

- Calcolo della derivata della funzione "scalino"



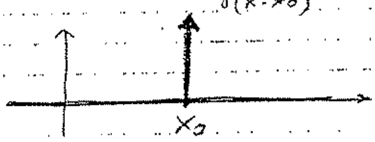


al limite x rompe sempre + ripida la fz converge allo scalino mentre la derivata converge al retta la con base nulla e h=∞ in modo che l'area sia 1

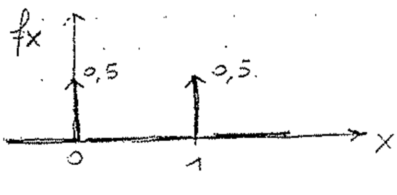
l'area è sempre = 1

$$\frac{d\,sca(t)}{dt} = \delta(t)$$

- Proprietà
  - $\int_{-\infty}^{+\infty} g(x)\,\delta(x-x_0)\,dx = g(x_0)$
  - $\int_{-\infty}^{+\infty} \delta(x)\,dx = 1 \quad (g(x) = 1,\ x_0 = 0)$

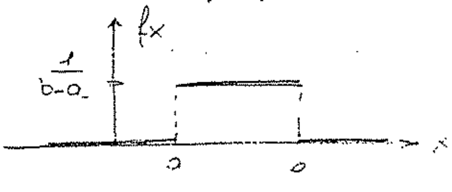


Es: Moneta



$$f_X(x) = 0{,}5\,\delta(x) + 0{,}5\,\delta(x-1)$$

Es: pto scelto in modo equiprobabile in [a b]



PROPRIETÀ

Nota la densità di probabilità è noto tutto

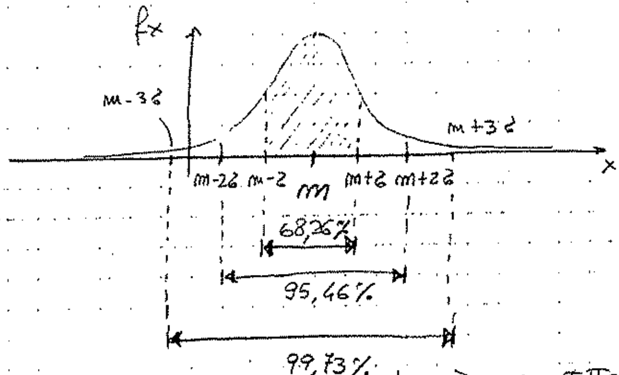
- $f_X(x) \geq 0 \quad \forall x$, non X ?
- $F_X(x) = \int_{-\infty}^{+\infty} f_X(\xi)\, d\xi$
- $\int_{-\infty}^{+\infty} f_X(x)\, dx = 1$
- $P(x_1 < X \leq x_2) = \int_{x_1^+}^{x_2^+} f_X(x)\, dx$
- $f_X(x) = \lim_{\Delta x \to 0} \frac{P(x \leq X \leq x + \Delta x)}{\Delta x}$

Alcune Variabili Casuali Notevoli

VARIABILE CASUALE GAUSSIANA

$$f_X(x) = \frac{1}{\sqrt{2\pi \sigma^2}}\, e^{-\frac{(x-m)^2}{2\sigma^2}}$$

è completamente caratterizzata da 2 parametri, m e σ, noti i quali è noto tutto



↳ è quasi tutto compreso tra m-3σ e m+3σ

VARIABILE CASUALE UNIFORME

$$f_X(x) = \begin{cases} \frac{1}{b-a}, & a \leq x \leq b \\ 0 & \text{altrove} \end{cases}$$

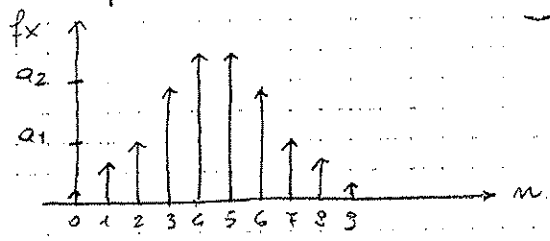
VAR. CASUALE BINOMIALE DI ORDINE n

X = n° di successi in n prove di Bernoulli

$$f_X(x) = \sum_{k=0}^{\infty} \binom{n}{k} p^k q^{n-k}\, \delta(x-k)$$

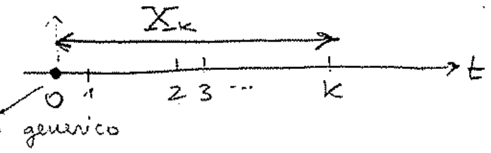
? n ?

es n = 9 $p = q = \frac{1}{2}$



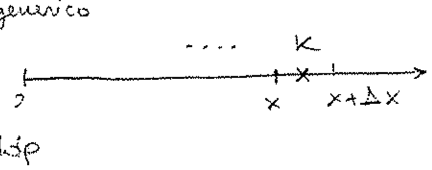
TEMPI DI INTERCORRENZA TRA EVENTI DI POISSON

Dati degli eventi di Poisson con densità λ, calcolare fdd e ddp della v.c.
X_k = tempo di attesa tra un generico evento ed il k-esimo evento successivo.



$P(x \leq X_k \leq x + \Delta x) = P($k-1 eventi in $[0,x]$, k-esimo evento in $[x, x+\Delta x]) \overset{\text{indip.}}{=} P($k-1 eventi in $[0\,x])\cdot$
$P($k-esimo evento in $[x, x+\Delta x]) =$

$$= \frac{(\lambda x)^{k-1}}{(k-1)!} e^{-\lambda x} \cdot \lambda \Delta x$$

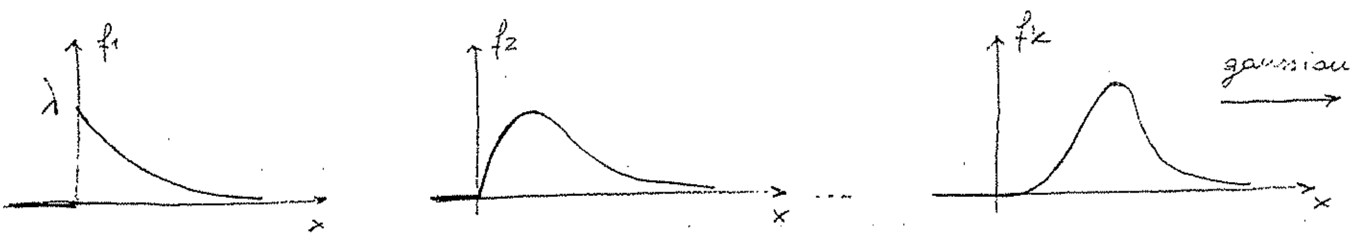


ddp

$$\lim_{\Delta x \to 0} \frac{P(x \leq X_k \leq x + \Delta x)}{\Delta x} = \frac{(\lambda x)^{k-1}}{(k-1)!} \lambda e^{-\lambda x}, \quad x \geq 0 \qquad \text{ERLANG - k}$$

La Erlang-1 è più nota come ESPONENZIALE

$F_X(x) = 1 - e^{-\lambda x}$ $f_X(x) = \lambda e^{-\lambda x}$



DISTRIBUZIONI CONDIZIONATE

Def: Sia $P(M)>0$. Distribuzione Condizionata:

$$F_X(x|M) := P(X \le x|M) = \frac{P(X \le x, M)}{P(M)}$$

Def: Densità Condizionata

$$f_X(x|M) = \frac{dF_X(x|M)}{dx} = \lim_{\Delta x \to 0} \frac{P(x \le X \le x+\Delta x | M)}{\Delta x}$$

OSSERVAZIONE: Disgiunzione e Indipendenza sono concetti diversi!

Es: Dado onesto $A=\{pari\}$ $B=\{dispari\}$ A (2 4 6) (1 3 5) B

- A e B sono disgiunti

$P(AB) = P(\{\emptyset\}) = 0 \ne P(A)P(B) \Rightarrow$ A e B non sono indip.!!

14-3-2001

Soluzione dell'esercizio

$X = \{$ Signora X è portatrice del gene $\}$

$2S = \{$ Signora X ha avuto 2 figli maschi sani $\}$

$3M = \{$ Signora X ha un 3° figlio maschio malato $\}$

1) $P(X|2S) = ?$ → si usa il Teo. di Bayes

$$P(X|2S) = \frac{P(2S|X)P(X)}{P(2S)}$$

$P(2S|X) = 0{,}5^2 = 0{,}25$ $\quad P(X) = 0{,}5$ $\quad$ Per calcolare $P(2S)$ uso il th. della probabilità Totale.

$P(2S) = P(2S|X)P(X) + P(2S|\bar{X})P(\bar{X})$

$= 0{,}25 \cdot 0{,}5 + 1 \cdot 0{,}5 = 5/8$

$$P(X|2S) = \frac{0{,}25 \cdot 0{,}5}{5/8} = \frac{1}{5}$$

2) $P(3M|2S) = ?$ $\quad = 1/10 = 1/2 \cdot 1/5$

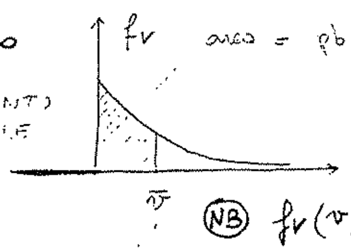
$$P(3M|2S, X)P(X|2S) + P(3M|2S,\bar{X})P(\bar{X}|2S) = 0{,}5 \cdot \frac{1}{5} + 0 \cdot \frac{4}{5} = \frac{1}{10}$$

condizioni condizionanti separate da ,

Esercizio: Un componente ha una durata di funzionamento distribuita in modo esponenziale. Sapendo che ha già funzionato per un Tempo τ, calcolare la ddp della durata Z del tempo di funzionamento restante

↳ PB di AFFIDABILITA'

V : var. casuale "VITA" del pezzo



$F_Z(z) = P(Z \le z) =$ — CONDIZIONAMENTO FONDAMENTALE

$= P(V-\tau \le z | V > \tau) =$

$= \frac{P(V-\tau \le z, V>\tau)}{}$ =

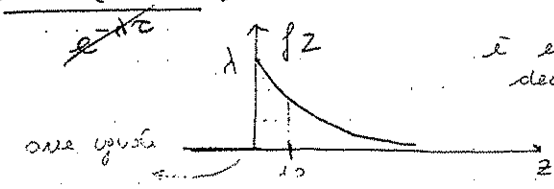
$Z := V - \tau$ sapendo che $V > \tau$

(NB) $f_V(v) = \lambda e^{-\lambda v}$ $\quad F_V(v) = 1 - e^{-\lambda v}$

$$= \frac{P(V \le z+\tau,\ V>\tau)}{1-F_V(\tau)} = \frac{P(\tau \le V \le z+\tau)}{1-(1-e^{-\lambda\tau})} = \frac{F_V(z+\tau)-F_V(\tau)}{e^{-\lambda\tau}} =$$

$$= \frac{1-e^{-\lambda(z+\tau)}-(1-e^{-\lambda\tau})}{e^{-\lambda\tau}} = \frac{\cancel{e^{-\lambda\tau}}(1-e^{-\lambda z})}{\cancel{e^{-\lambda\tau}}} = 1-e^{-\lambda z}$$

$$f_Z(z) = \lambda e^{-\lambda z}$$



è esattamente la stessa densità di probabilità di prima

Gli eventi di Poisson sono come le prove di Bernoulli ⇒ i ritardi NON hanno nessun valore ⇒ **VALE LA PROPRIETÀ DI NON MEMORIA**

anche se il fenomeno è già durato per un tempo τ, da quell'istante è come se ricominciasse da zero.

(Questo modello non va bene per es: x rappresentare la vita umana)

Il modello Poissoniano non va bene per tutte le volte che ci sono dei "pezzi" soggetti ad USURA. MA se stiamo guardando un IMPIANTO IN generale (con tanti pezzi in parallelo) l'evento è poissoniano (xché dato un guasto non posso prevedere il successivo). Si spera che il passaggio dell'autobus non sia poissoniano.

TEOREMA DELLA PROBABILITÀ TOTALE (per VAR. CASUALI):

Ipotesi: $A_1 + A_2 + \dots + A_n = S$, A_i disgiunti

Tesi: $F_X(x) = \sum_{i=1}^{n} F_X(x|A_i)\,P(A_i)$ $\qquad f_X(x) = \sum_{i=1}^{n} f_X(x|A_i)\,P(A_i)$

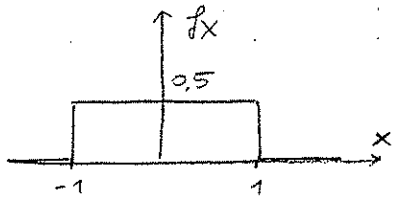
FUNZIONI DI UNA VAR. CASUALE

Data una var. casuale $X(\cdot)$ e una funzione $g(\cdot)$: allora $Y := g(X)$ è una nuova var. casuale

ATTENZIONE: Non tutte le $g(\cdot)$ vanno bene. Se $X(\cdot)$ è gaussiana $Y := \sqrt{X}$ non ha senso! xché la gaussiana è def. anche x valori negativi.

PROBLEMI: calcolo di fdd e ddp di Y

Esempio: X è una var. casuale uniforme in $[-1,1]$. Trovare la fdd di $Y := X^2$

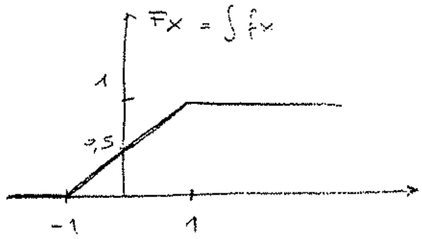


$F_Y(y) = P(Y \le y)$ Distinguiamo 2 casi:

- $y \ge 0$ $\quad P(Y \le y) = P(X^2 \le y)$ ovvero che $\boxed{X^2 \le y \iff -\sqrt{y} \le X \le \sqrt{y}}$
  $= P(-\sqrt{y} \le X \le \sqrt{y}) = F_X(\sqrt{y}) - F_X(-\sqrt{y})$

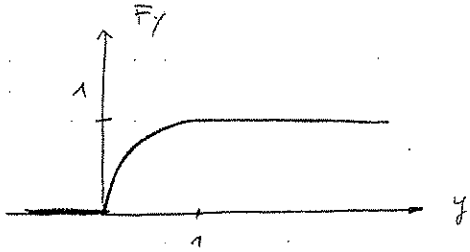
- $y < 0$ $\quad P(Y \le y) = 0 \Rightarrow F_Y(y) = 0,\ y<0$
  perché $Y = X^2$ non può essere negativo (al + è zero)

Troviamo $F_X(x)$



$$F_X(x) = \begin{cases} 0, & x < -1 \\ \frac{1}{2} + \frac{x}{2} & |x| \le 1 \\ 1, & x > 1 \end{cases}$$

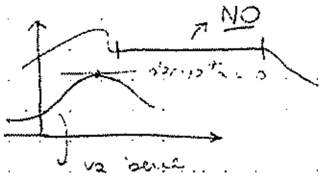
$$F_Y(y) = \begin{cases} 0 & y < 0 \\ \frac{1}{2} + \frac{\sqrt{y}}{2} - \left(\frac{1}{2} - \frac{\sqrt{y}}{2}\right) = \sqrt{y} & 0 \le y \le 1 \end{cases}$$



(NB) Quando trasformiamo una variabile casuale la sua densità cambia !!
↓
in questo caso una v.c. uniforme diventa NON + uniforme !

ddp di $Y = g(X)$: Posso calcolare la fdd e poi derivare

In alternativa : TEOREMA : Sia X una var. casuale continua e $g(\cdot)$ una funzione continua, derivabile, con derivata prima $\neq 0$ quasi ovunque (cioè non ci sono tratti orizzontali di lunghezza finita).
Dette $x_1, x_2 \dots x_n$ le radici dell'eq. $y = g(x)$ la ddp di $Y = g(X)$ è continua ed è data da:



$$f_Y(y) := \begin{cases} \sum_{i=1}^{n} \dfrac{f_X(x_i)}{|g'(x_i)|} \\ 0 \quad \text{se } y = g(x) \text{ non ha radici} \end{cases}$$

NOTA : $x_i = x_i(y)$

Es: $Y := e^X$ con X gaussiana $f_Y(y) = ?$

- $y \leq 0$ $\quad y = e^x$ non ha radici $\Rightarrow f_Y(y) = 0,\ y \leq 0$
- $y > 0$ $\quad y = e^x \Rightarrow x = \ln y$ è l'unica radice

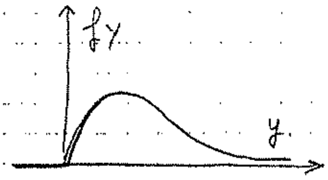
$$\frac{dg(x)}{dx} = \frac{de^x}{x} = e^x = y$$

Perciò $f_Y(y) = \dfrac{f_X(\ln y)}{|y|} = \boxed{\dfrac{1}{\sqrt{2\pi\sigma^2}\, y}\, e^{-\frac{(\ln y - m)^2}{2\sigma^2}}, \quad y > 0}$

(NB) NON fare MAI il log di una gaussiana !!

DENSITÀ LOGNORMALE
↓
$\neq$ gaussiana

detta così xchè facendone il logaritmo ottengo una normale (cioè una gaussiana)



Es: $Y = aX + b$ trovare f_Y $\quad g(x) = ax + b \quad |g'(x)| = |a| \quad$ hp: $a \neq 0$

$y = ax + b \longrightarrow x = \dfrac{y-b}{a}$ è l'unica radice

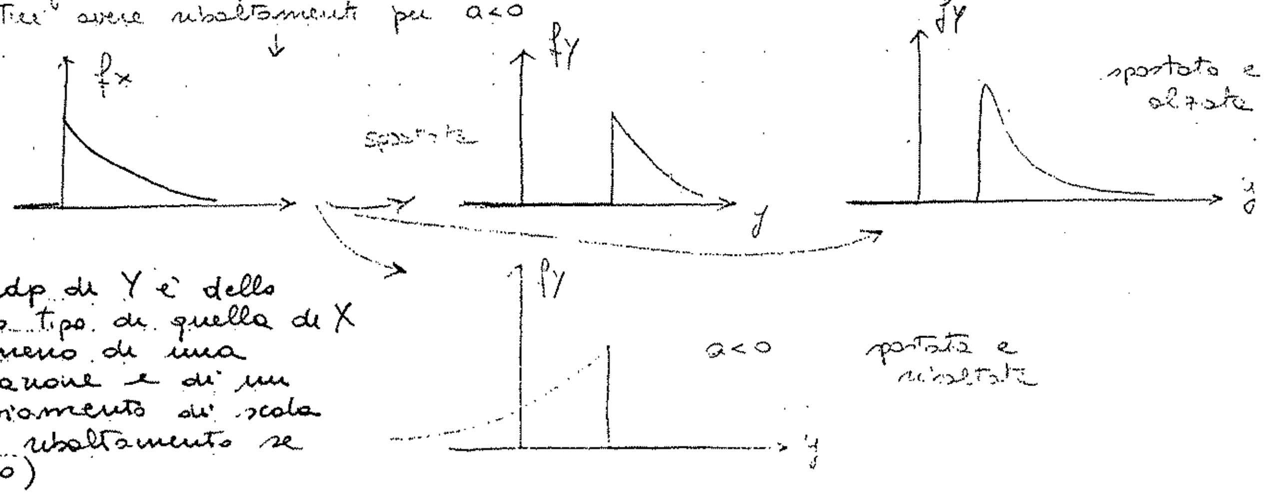
$$f_Y(y) = \frac{1}{|a|} \cdot f_X\left(\frac{y-b}{a}\right)$$

] Regola x ottenere la ddp di una trasf. lineare

$\parallel$
$f_X(x)$ a meno di
- eventuali Traslazioni $(y-b)$
- potrebbe essere alzata o schiacciata $\left(\frac{1}{|a|}\right)$

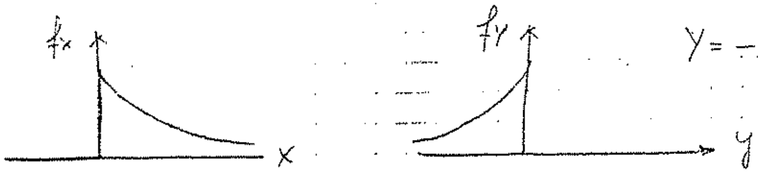
↓
se ci fosse un exp potrei avere ribaltamenti per $a<0$

Es:



La ddp di Y è dello stesso tipo di quella di X a meno di una traslazione e di un cambiamento di scala (con ribaltamento se $a<0$)

es:

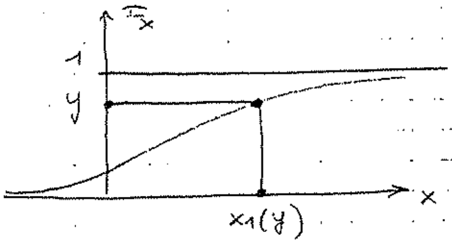


$Y = -X$

Es: Data X con F_X strettamente crescente, Trovare la ddp di $Y = F_X(X)$

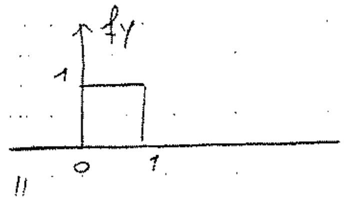
NB: $g := F_X$

Se F_X strett. crescente $\Rightarrow$ $y = F_X(x)$ ammette al più una sola radice



Allora $f_Y(y) = \dfrac{f_X(x_1)}{|F_X'(x_1)|} = \dfrac{f_X(x_1)}{f_X(x_1)} = 1$, $0 \le y \le 1$

Tanto è sempre > 0



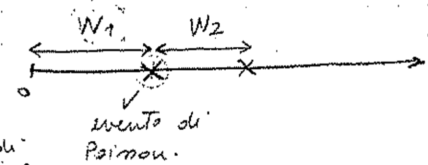
Ho trovato una procedura Tc. qualunque X io abbia posso sempre ottenere una ddp uniforme usando $g = F_X$!! Serve al contrario!

15-3-2001

Problema: Data una var. casuale U uniforme in $[0,1]$, trovare $g(\cdot)$ t.c. $X = g(U)$ abbia una fdd assegnata F_X

un generatore di n casuali dà qualcosa distribuita in modo uniforme

Serve x simulare nel calcolatore variabili casuali con fdd NON uniforme



tempo di attesa del 1° evento

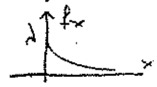
Esempio di applicazione: simulazione di eventi di Poisson $\Rightarrow$ W_n exp

Ricordando l'esempio precedente, $g(U) = F_X^{-1}(U)$ inversa.

Voglio ottenere una v. casuale X distribuito in modo exp.

$F_X(x) = 1 - e^{-\lambda x}$, $x \ge 0$ $\quad u = 1 - e^{-\lambda x} \rightarrow e^{-\lambda x} = 1 - u \rightarrow x = -\dfrac{\ln(1-u)}{\lambda} = F_X^{-1}(u)$

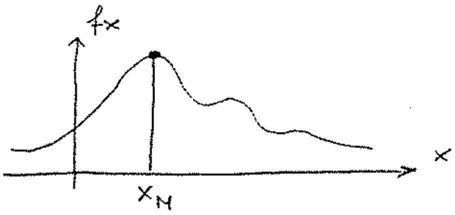
$X = g(U) = -\dfrac{\ln(1-U)}{\lambda}$ $\rightarrow$ nuova variabile casuale che ha una distribuzione exp.



(PB) Per generare var. gaussiane non posso usare questa tecnica xchè la gaussiana non ha F_X ma ha f_X

PARAMETRI CARATTERISTICI DELLE DDP

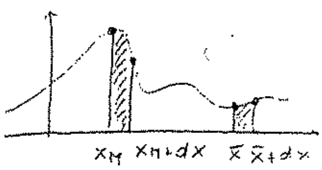
MODA



è la coordinata x_M cui corrisponde il max valore di f_X

In qualche senso è il valore + probabile

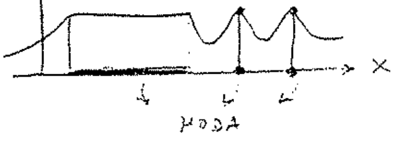
OSS • $P(x_M \le X \le x_M + dx) \cong f_X(x_M)\,dx$



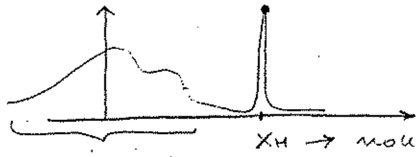
$\Rightarrow$ x_M è in un certo senso il valore più probabile (in realtà se la var. casuale è continua $P(X = \bar{x}) = 0$ $\forall \bar{x} \in \mathbb{R}$).

• Può non essere unica

Quando è unica la ddp è detta UNIMODALE



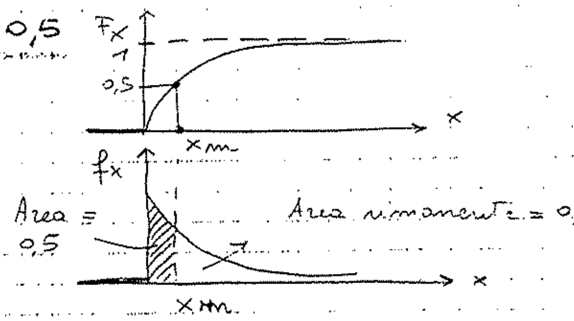
- Può essere poco significativa



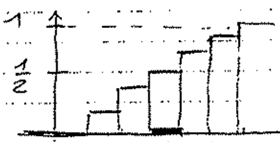
(NB) la moda è facile da trovare xchè basta porre a zero la derivata della ddp.

MEDIANA

Il valore x_m tale che $F_X(x_m) = 0{,}5$



OSS
- $P(X \leq x_m) = P(X > x_m) = 0{,}5$
- Può non essere unica



la mediana può essere usata quando la distribuzione non è ben definita (come forma)

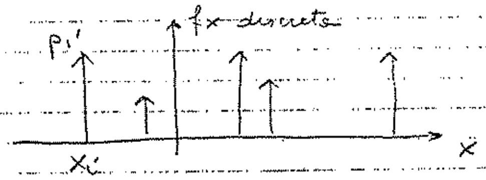
MOMENTI

- Momenti di ordine k $\quad m_k := \int_{-\infty}^{+\infty} x^k f_X(x)\, dx$

Per la var. casuale discreta si ottiene $m_k = \sum_{i=-\infty}^{\infty} x_i^k p_i$

aree dei corrispondenti impulsi

valori possibili

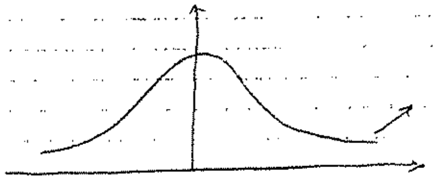


- Momenti centrali di ordine k $\quad \mu_k := \int_{-\infty}^{+\infty} (x - m_1)^k f_X(x)\, dx$

Per var. casuali discrete si ottiene $\mu_k = \sum_{i=-\infty}^{\infty} (x_i - m_1)^k p_i$

OSS Non è detto che esistono xchè non è detto che $\int_{-\infty}^{\infty}$ converga

Per es: la var. casuale di Cauchy $\left(f_X(x) = \left(\frac{1}{\pi} \frac{1}{1+x^2} \right) \right)$ non ammette momenti



va a zero come $\frac{1}{x^2}$ (vs la gaussiana va a zero come e^{-x^2})

$$m_1 = \int_{-\infty}^{\infty} x\, f_X(x)\, dx$$

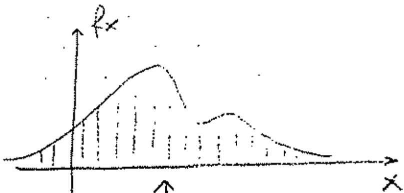
$\int \frac{1}{x}$ non converge

MEDIA O VALORE ATTESO

m_1, $E[X]$, m_X

expectation — valore atteso

Interpretazione: baricentro della ddp.



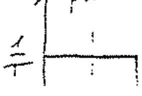
$$m_1 := \int_{-\infty}^{+\infty} x\, f_X(x)\, dx$$

formula del baricentro quando la distribuzione di massa è f_X

In generale NON coincide con la mediana

Es: $X(\cdot)$ v.c. uniforme in $[0, T]$

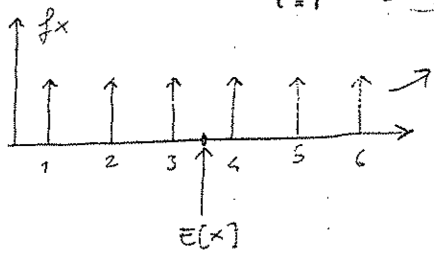
$$E[X] = \int^{T} x \cdot \frac{1}{T}\, dx =$$



$= \frac{1}{T}\left[\frac{x^2}{2}\right]_0^T = T/2$ Ogni volta che c'è symm nella ddp baricentro = mediana

<u>OSS</u>: Non è detto che coincida con uno dei valori possibili per la v.c.

es: Dado $E[X] = \sum_{i=1}^{6} x_i p_i = \frac{1}{6}\sum_{i=1}^{6} i = 3,5$ → NON è un valore possibile

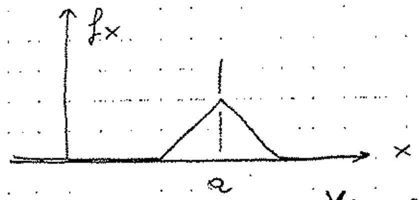


VALORE ATTESO CONDIZIONATO

$$E[X|M] = \int_{-\infty}^{+\infty} x\, f_X(x|M)\, dx$$

PROPRIETA': Se $f_X(x)$ è simmetrica attorno a $x=a$ e m_1 esiste ⇒

$m_1 = x_m = a$



TEOREMA Data la v.c. X e consideriamo $Y := g(X)$

$$E[Y] := E[g(X)] = \int_{-\infty}^{+\infty} g(x) f_X(x)\, dx$$

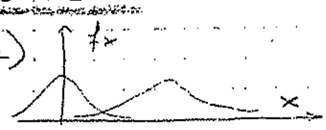
VANTAGGIO: Non è necessario calcolare f_Y !!

CONSEGUENZE

- Per la def di m_2, $m_2 = E[X^2]$ (ricordiamo che $m_2 := \int_{-\infty}^{\infty} \underbrace{x^2}_{g(x)} f_X(x)\, dx$)

Tutti i momenti vengono ricondotti a delle medie!!

dice quanto una densità è vicina all'origine (compattezza) ← VALOR QUADRATICO MEDIO



- Se $\boxed{Y = aX + b}$ $E[Y] = \int_{-\infty}^{+\infty} (ax+b) f_X(x)\, dx =$

$= a\int x f_X(x)\, dx + b\int f_X(x)\, dx = \boxed{a\, E[X] + b}$ → L'operatore di aspettazione $E[\cdot]$ è lineare!!

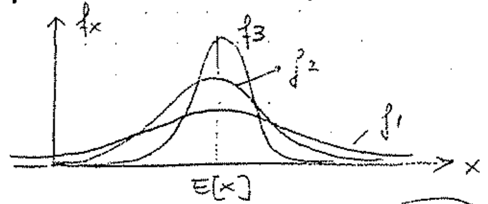
VARIANZA

misura la dispersione intorno al valor medio

$\mu_2 := \int_{-\infty}^{+\infty} (x-m_1)^2 f_X(x)\, dx$ è detta VARIANZA della v.c. e si indica con

$Va[X]$, σ_X^2 → è sempre > 0

La quantità $\sigma_X = \sqrt{Va[X]}$ è detta deviazione standard.



stessa media
MA varianze ≠

$\sigma_1^2 > \sigma_2^2 > \sigma_3^2$

es: f_3 → classe di studenti + o - con gli stessi voti
f_1 → voti molto vari

NOTA: $Var[X] = \int_{-\infty}^{+\infty} (x-m_1)^2 f_X(x)\, dx = E[(X-m_1)^2] = E[(X-E[X])^2]$

$g(x) = (x-m_1)^2$

PROPRIETA': $\mu_2 = m_2 - m_1^2$

$Var[X] = E[X^2] - E[X]^2$

$$Var[X] = E[(X - m_1)^2] = E[X^2 - 2Xm_1 + m_1^2] = E[X^2] + E[-2Xm_1] + E[m_1^2] =$$

x la linearità   sono NUMERI deterministici   non è casuale

$$= m_2 - 2m_1 E[X] + m_1^2 =$$

$$= m_2 - 2m_1^2 + m_1^2 = m_2 - m_1^2 \quad \text{cvd}$$

Esercizio : Conoscendo σ_X^2, trovare la varianza di $Y = aX + b$

$$\sigma_Y^2 = E[(Y - E[Y])^2] = E[(aX + b - (aE[X] + b))^2] = E[(aX + \cancel{b} - aE[X] - \cancel{b})^2] =$$

$$= E[a^2 (X - E[X])^2] = a^2 E[(X - E[X])^2] = a^2 \sigma_X^2$$

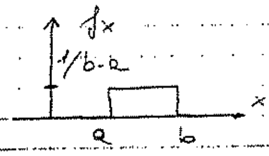
$$\Rightarrow \sigma_Y^2 = a^2 \sigma_X^2$$

$$\boxed{Var[aX + b] = a^2 Var[X]}$$ non è lineare come $E[X]$

cost non conta nulla

MEDIA E VARIANZA DI ALCUNE V.C. NOTEVOLI

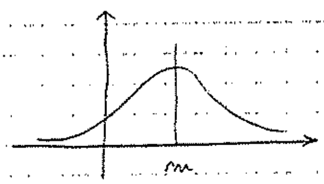
Uniforme in [a b]



$$E[X] = \frac{a+b}{2}$$

$$Var[X] = \frac{(b-a)^2}{12}$$

Gaussiana



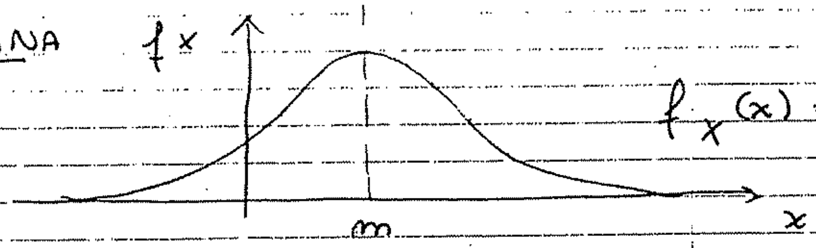
$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-m)^2}{2\sigma^2}}$$

$$E[X] = m$$

$$Var[X] = \sigma^2$$

$E[X] = \frac{a+b}{2}$, $Var[X] = \frac{(b-a)^2}{12}$

GAUSSIANA

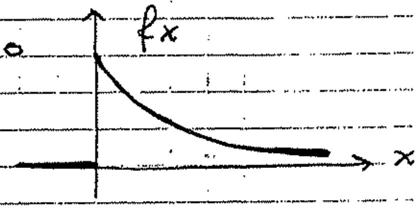


$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-m)^2}{2\sigma^2}}$$

$E[X] = m$, $Var[X] = \sigma^2$ — sono i 2 parametri che caratterizzano la gaussiana

ESPONENZIALE

$$f_X(x) = \lambda e^{-\lambda x} , \quad x \geq 0$$



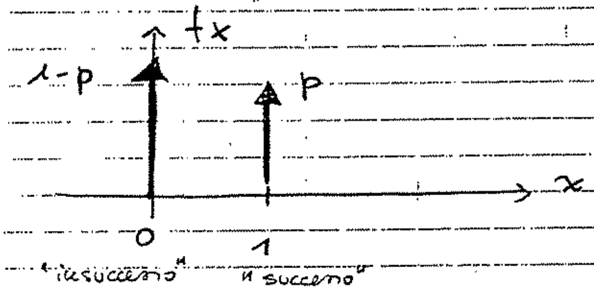
(bisogna fare un'integrazione per parti)

$E[X] = \frac{1}{\lambda}$ $\qquad Var[X] = \frac{1}{\lambda^2}$

è il tempo medio. $\qquad$ freq. eventi Poisson

$\sum_{i=0}^{1} (x_i - p)^2 p_i$

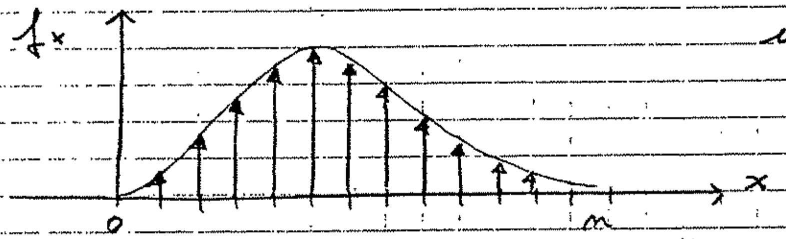
BERNOULLI



$E[X] = p$ $\qquad Var[X] = p(1-p) = pq$

$\sum_{i=0}^{1} x_i p_i = 0 \cdot (1-p) + 1 \cdot p$

BINOMIALE



essendo discreta $\Longrightarrow$ la ddp è un pettine di impulsi

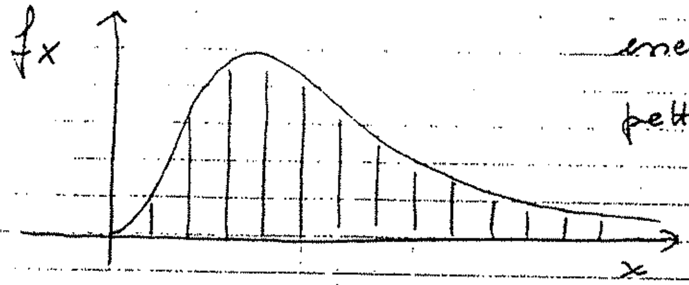
n prove di Bernoulli

$n \to \infty$?

$$f_X(x) = \sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k} \delta(x-k)$$

$E[X] = np$ $\qquad Var[X] = npq$

POISSONIANA (= n° eventi di Poisson in un intervallo di lunghezza t)



essendo discreta => ddp è un pettine di impulsi

se estende fino a +∞

Questo anche se l'intervallo di tempo è FINITO → perché non c'è limite al n° di eventi di Poisson.

$E[X] = \lambda t$

$Var[X] = \lambda t$

essendo un n° puro è possibile che vengano sequiti (è molto difficile che lo siano in genere)

$f_X(x) = \sum_{k=0}^{\infty} \frac{(\lambda t)^k}{k!} e^{-\lambda t} \delta(x-k)$

λ = freq.
t = tempo } => N° puro

FORMULE APPROSSIMATE PER IL CALCOLO DI MEDIA E VARIANZA DI UNA FUNZIONE DI V.C.

$Y = g(X)$ mi interessano $E[Y]$ e $Var[Y]$

$E[g(x)]$

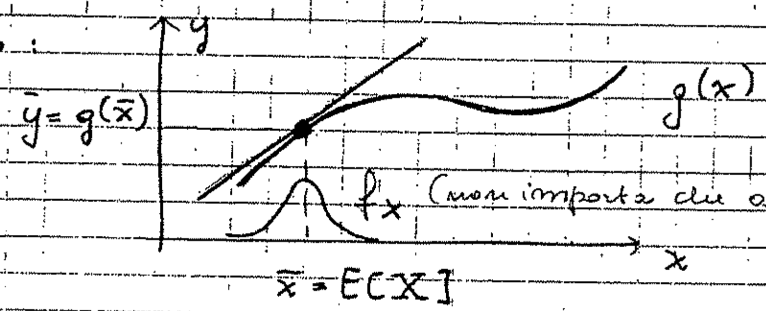
$Var[Y] = \cancel{Var} E[(g(X) - E[g(X)])^2]$

Situazione sfavorevole: conosco la $E[X]$ ma non la densità di probabilità.

Come fare se non conosco $f_X(x)$? => non posso calcolare $E[Y]$ e $Var[Y]$ in maniera precisa.

Sotto particolari condizioni xò posso calcolare $E[Y]$ e $Var[Y]$ in modo approssimato.

Grafico:



la curva si lascia approssimare bene dalla tg.

(non importa che abbia forma a campana o meno)

Ipotesi fondamentale: f_X è sufficientemente concentrata intorno a $\bar{x} = E[X]$ (cioè $g(x)$ deve essere ben approssimata dalla tg intorno a $\bar{x}$)

→ deve avere Varianza piccola

Se $X - E[X]$ è "piccola"

=> taylor fino al 1° ordine

$$Y = g(X) \cong g(\bar{x}) + \left.\frac{dg(x)}{dx}\right|_{x=\bar{x}} (X - \bar{x})$$

Perciò: applico il valore atteso

$E[X]-E[X]=$

$$\underline{E[Y]} \cong g(\bar{x}) + \left.\frac{dg}{dx}\right|_{x=\bar{x}} E[X - \bar{x}] = $$

$$\underset{E[X]}{\bar{x}}$$

$$= \underbrace{g(\bar{x}) = g(E[X])}$$

posso pensare che il valore $g(\bar{x})$ è ben approssimato da $g(E[X])$. Questo in generale non vale a meno che l... non sia verificato ($Var[X]$ deve essere piccola)

$$Var[Y] \cong \left(\left.\frac{dg}{dx}\right|_{x=\bar{x}}\right)^2 \cdot Var[X]$$

coeff. angol. della Tg $\Leftarrow$

ESERCIZIO: Dati $E[X]$ e $Var[X]$, trovare la trasformazione lineare del tipo $Y = aX+b$: tale che $E[Y]=$ ... e $Var[Y]=1$. (in gergo si dice che Y e' "standardizzata")

· Devo imporre che $E[Y]=0 \Rightarrow E[aX+b]=0 \Rightarrow$
$a \cdot E[X]+b=0$

· Devo imporre $Var[Y]=1 \Rightarrow Var[aX+b]=1$
$\Rightarrow a^2\, Var[X]=1$

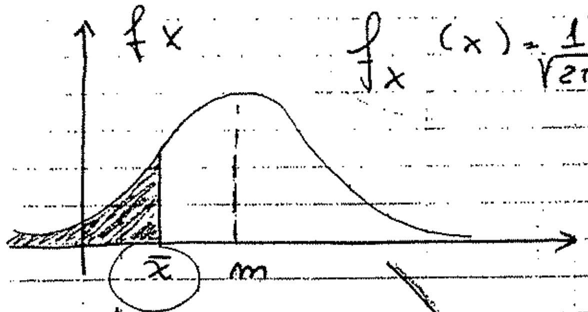
$$\Rightarrow \begin{cases} b = -\dfrac{E[X]}{\sqrt{Var[X]}} \\ a = \dfrac{1}{\sqrt{Var[X]}} \end{cases}$$

$$\underline{Y} = \frac{X}{\sigma_x} - \frac{E[X]}{\sigma_x} = \underline{\frac{X - E[X]}{\sigma_x}}$$

questa è un'operazione di standardizzazione (serve x uniformare i dati)

Un esempio di applicazione: Formule per la gaussiana (es tabella per la gaussiana standard che ha $E[X]=0$ e $Var[X]=1$)

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-\frac{(x-m)^2}{2\sigma^2}}$$



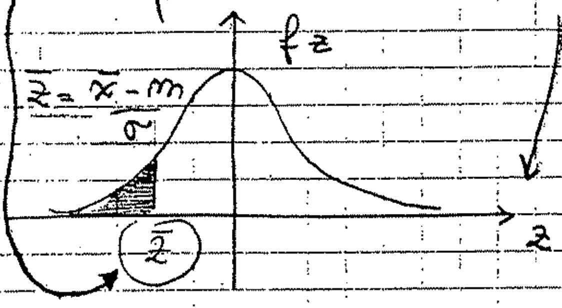
$P(X \le \bar{x}) = ?$

non posso calcolarne l'integrale (xchè non è un integrale che si risolve con una formula chiusa

=> Devo per forza usare la TABELLA

trasf. di standardiz. $Z = \frac{X - m}{\sigma}$

che prende in considerazione la gaussiana standard



$$f_Z(z) = \frac{1}{\sqrt{2\pi}}\, e^{-\frac{z^2}{2}}$$

$F_Z(\bar{z})$ (

$$P(X \le \bar{x}) = P(\sigma Z + m \le \bar{x}) = P\left(Z \le \frac{\bar{x} - m}{\sigma}\right) = P(Z \le \bar{z})$$

$\left(Z = \frac{X - m}{\sigma} \Rightarrow X = \sigma Z + m\right)$

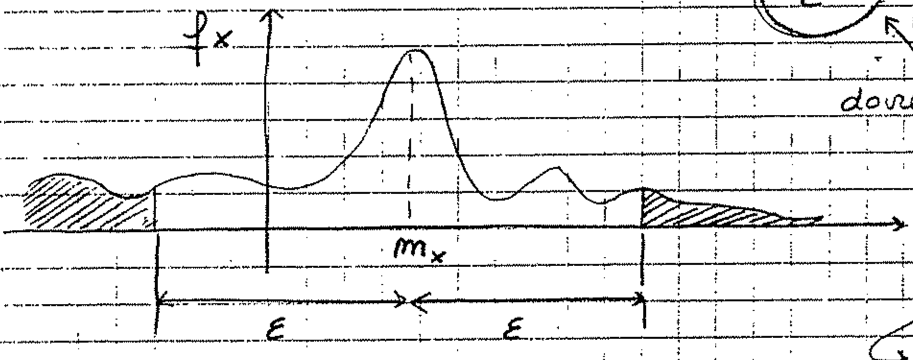
dove $\bar{z} := \frac{\bar{x} - m}{\sigma}$

DISUGUAGLIANZA DI TCHEBYCHEFF (CEBICEV)

$\forall \varepsilon > 0, \quad P(|X - m_x| \ge \varepsilon) \le \frac{\sigma_x^2}{\varepsilon^2}$ se conosco la f_X => mi dà il limite superiore

INVECE

dovrei calcolarmi i 2 integrali (



ε è importante

σ_x^2 " mi dice quanto grande è la distanza dalla media

Interpretazione: se il rapporto $\frac{\varepsilon^2}{\sigma_x^2}$ è sufficientemente grande la probabilità che X cada all'esterno di $[m_x - \varepsilon, m_x + \varepsilon]$ è trascurabile

$\varepsilon = 100\sigma$ la P(cadere fuori dalla media) $= \frac{1}{100}$ mi dà una misura di quanto sono lontano dalla media.

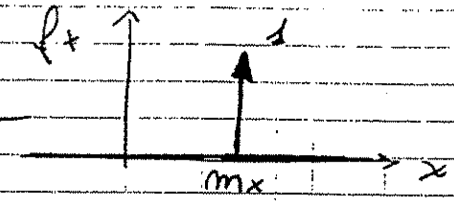
Media e varianza devono $\exists$ => vale per qualunque distribuzione (

(tranne per Cauchy xchè non ha momento)

OSSERVAZIONE:

Se $\sigma_x = 0 \Rightarrow P(|X - m_x| \geq \varepsilon) = 0, \quad \forall \varepsilon \Rightarrow$

$\Rightarrow P(X = m_x) = 1$



è una V.C. degenere

perché non è più una V.C.

ma una costante $= m_x$

VARIABILI CASUALI CONGIUNTE

Esperimento casuale che produce una n-upla di numeri:

$$X(\omega) = \begin{bmatrix} X_1(\cdot) \\ X_2(\cdot) \\ \vdots \\ X_n(\cdot) \end{bmatrix} \quad \text{V.C. vettoriale } n\text{-dimensionale}$$

perché è una fz dell'esito casuale

Caso $n = 2$: per comodità invece di scrivere $X_1(\cdot)$ e $X_2(\cdot)$

usiamo $\begin{bmatrix} X(\cdot) \\ Y(\cdot) \end{bmatrix}$

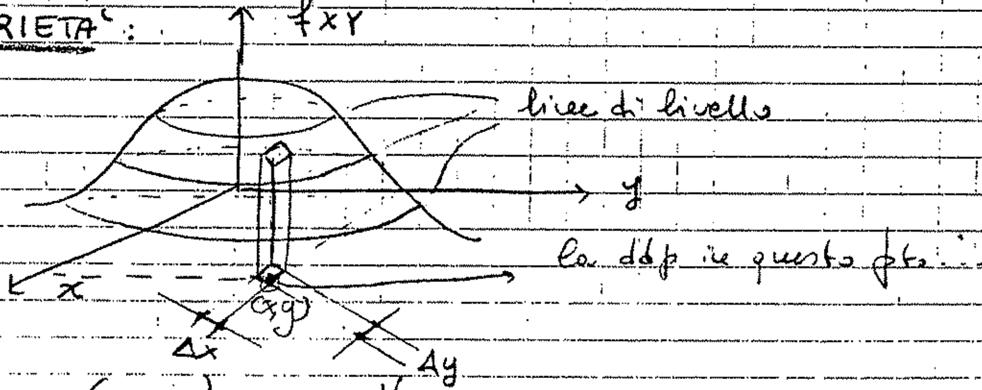
Def: Funzione di distribuzione congiunta

$$F_{XY}(x, y) := P(X \leq x, Y \leq y)$$

è fz di 2 eventi — eventi congiunti $\Rightarrow$ verifica simultanea dei 2.

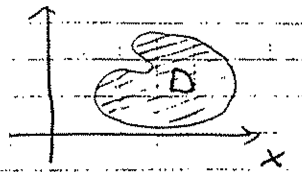
Def: ddp congiunta $f_{XY}(x, y) := \dfrac{\partial^2 F_{XY}(x, y)}{\partial x \, \partial y}$

PROPRIETA':



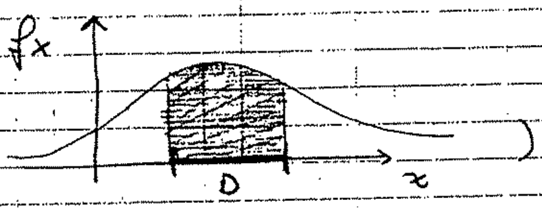
- $f_{XY}(x, y) \geq 0, \quad \forall x, y$
- $F_{XY}(x, y) = \displaystyle\int_{-\infty}^{x} \int_{-\infty}^{y} f_{XY}(\alpha, \beta) \, d\alpha \, d\beta$
- $f_{XY}(x, y) = \displaystyle\lim_{\substack{\Delta x \to 0 \\ \Delta y \to 0}} \frac{P(x \leq X \leq x + \Delta x, \; y \leq Y \leq y + \Delta y)}{\Delta x \, \Delta y}$ → vedere parallel

- $P((X,Y)\in D) = \int_D f_{XY}(x,y)\,dx\,dy$



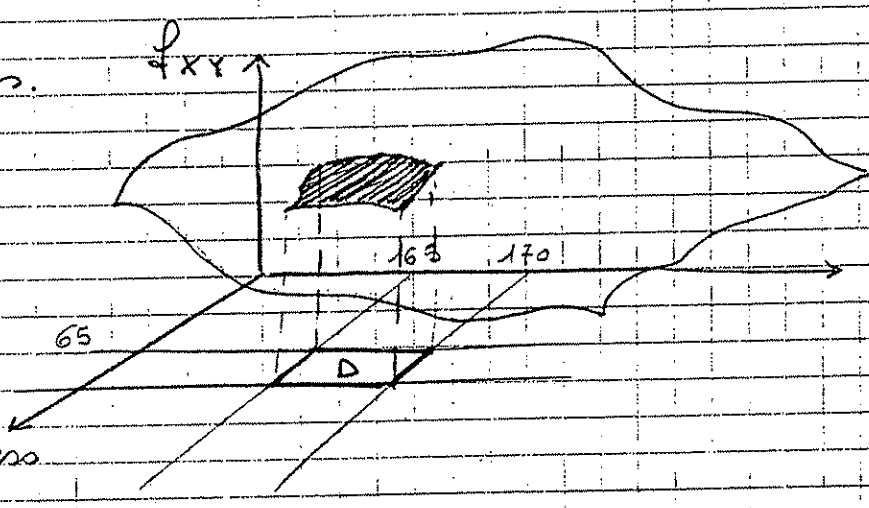
è un' integrale DOPPIO

(nel caso lineare era un' area:



)

es.



dovrò poi calcolare il volume del solido => INTEGRALE DOPPIO

DISTRIBUZIONE E DENSITÀ MARGINALI

Data una coppia di V.C. $X(\cdot)$, $Y(\cdot)$ le fdd F_X, F_Y e le ddp f_X, f_Y delle SINGOLE V.C. $X(\cdot)$ e $Y(\cdot)$ sono dette fdd e ddp marginali (in contrapposizione alle fdd e ddp congiunte F_{XY} e f_{XY})

es. coppie peso-altezza

es. posso considerarle sing. peso (senza altezza) o altezza (senza peso!).

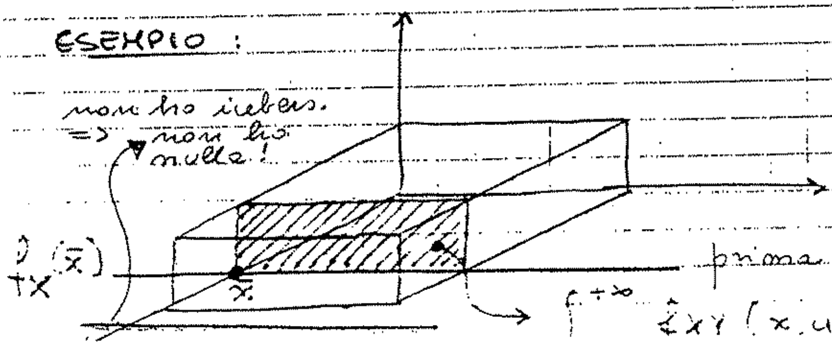
PROPRIETÀ:

- $F_X(x) = F_{XY}(x,\infty)$; $F_Y(y) = F_{XY}(\infty, y)$
- $f_X(x) = \int_{-\infty}^{+\infty} f_{XY}(x,y)\,dy$; $f_Y(y) = \int_{-\infty}^{+\infty} f_{XY}(x,y)\,dx$

ESEMPIO:

$f_X(\bar{x}) = ?$

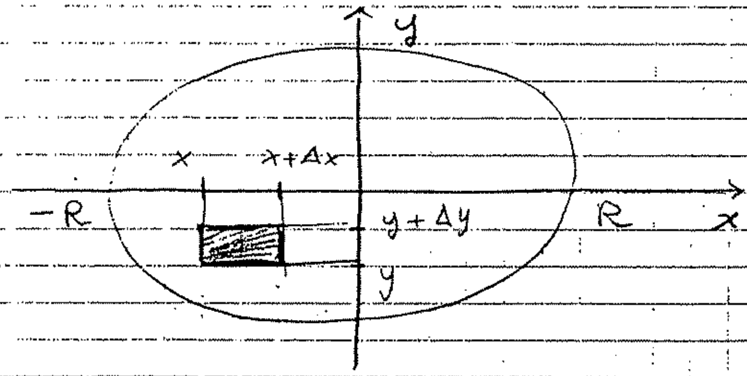
non ho iceberg => non ho nulla!



prima effetto (tengo la x fissa)

$\int^{+\infty} f_{XY}(x,y)\,dy$

ESEMPIO: trovare la ddp congiunta e marginale delle V.C. X, Y coordinate di un pto Q scelto in modo equiprob. in un cerchio di raggio R e centro nell'origine

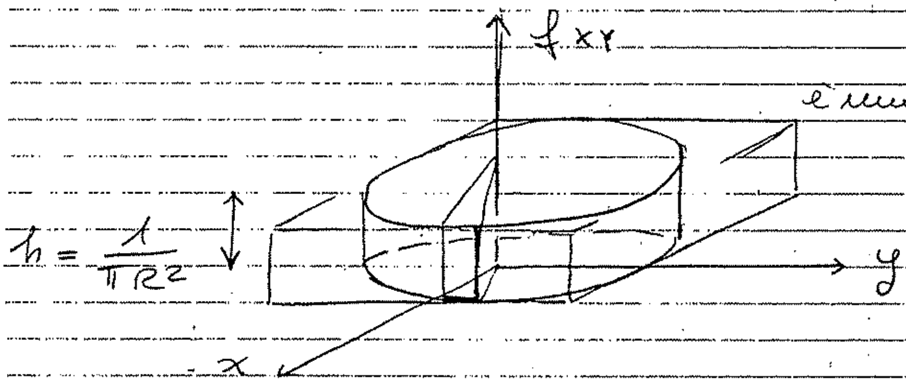


pb: prendere un pto casuale all'interno del cer

$$P(x \le X \le x+\Delta x,\ y \le Y \le y+\Delta y) = \frac{\Delta x \Delta y}{\pi R^2}$$

$$\Rightarrow f_{XY}(x,y) = \frac{1}{\pi R^2} \quad \text{(dentro il cerchio)}$$

fuori la probabilità di cader è nulla.

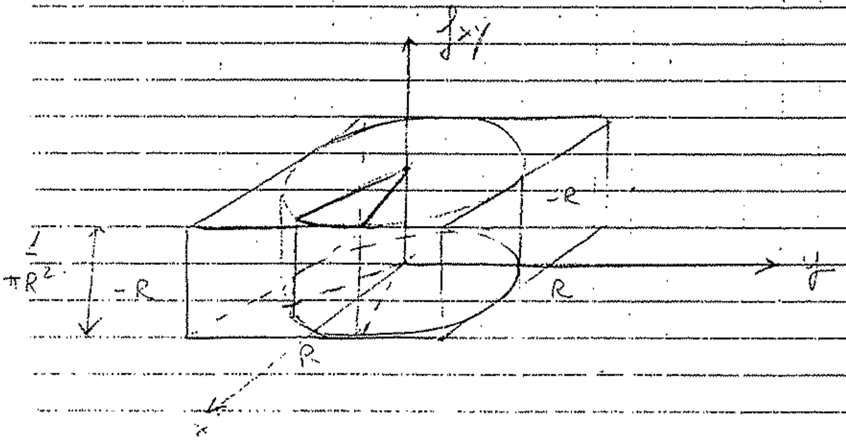


è una torta!

$-R \le x \le R$

$-R \le y \le R$

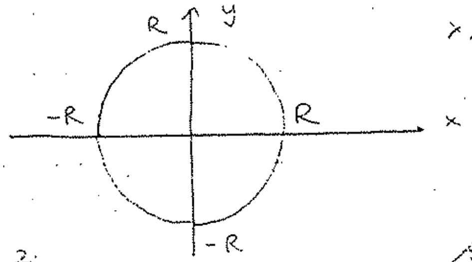
- Il volume sotteso è UNITARIO (come era unitario nel caso lineare



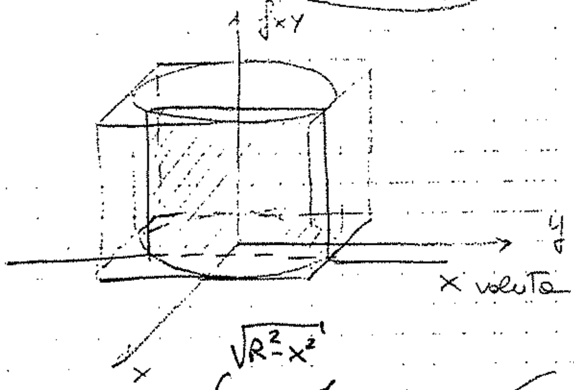
20 - 3 - 2001

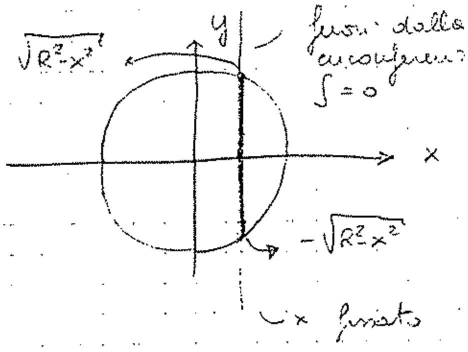
Esempio pto scelto in modo equiprobabile in una circonferenza

x, y = coordinate



1) $f_{XY}(x,y) = \frac{1}{\pi R^2}$ (dentro la circonferenza)



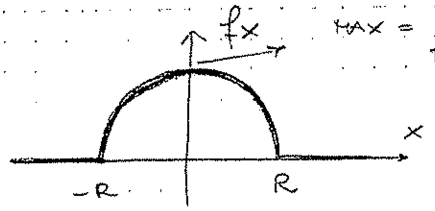


$x^2 + y^2 = R^2$

$y = \pm\sqrt{R^2-x^2}$

densità marginale

2) $f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x,y)\,dy = \int_{-\sqrt{R^2-x^2}}^{\sqrt{R^2-x^2}} \frac{1}{\pi R^2}\,dy = \frac{2}{\pi R^2}\sqrt{R^2-x^2}, \quad -R \le x \le R$



MAX = taglio la torta esattamente a metà

Fuori dalla circonferenza è zero, in R è tg ⟹ è zero

DENSITÀ CONDIZIONATE

Def := $f_{Y|X}(y|X=x) := \frac{f_{XY}(x,y)}{f_X(x)}$

TEOREMA DELLA PROBABILITÀ TOTALE

$f_Y(y) = \int_{-\infty}^{+\infty} f_{Y|X}(y/X=x) f_X(x)\,dx$ → $\int$ e non Σ perchè la densità è continua.

TEOREMA DI BAYES

$f_{Y|X}(y/X=x) = \frac{f_{X|Y}(x|Y=y) f_Y(y)}{f_X(x)}$

Def: X, Y sono var. casuali INDIPENDENTI se $F_{XY}(x,y) = F_X(x)F_Y(y)$

PROPRIETÀ: Le seguenti relazioni sono equivalenti

1. $F_{XY}(x,y) = F_X(x)F_Y(y)$
2. $f_{XY}(x,y) = f_X(x) f_Y(y)$
3. $f_{X|Y}(x|Y=y) = f_X(x)$
4. $f_{Y|X}(y/X=x) = f_Y(y)$

ognuna di queste è VERA ⟺ lo sono anche tutte le altre 3

1. ⟺ 2. ⟺ 3. ⟺ 4.

conoscere y non mi dà nessuna informazione su x (non è il caso in cui, nota l'altezza, deduco qualcosa sul peso!)

se conosciamo la densità congiunta possiamo fare la verifica → di solito note le densità marginali ottengo la congiunta come prodotto

PROPRIETÀ: Se le v.c. X e Y sono indipendenti ⟹ le v.c. $Z = g(X)$ e $W = h(Y)$ sono anch'esse indipendenti

funzioni

FUNZIONI DI VARIABILI CASUALI CONGIUNTE

Date 2 v.c. X, Y e una funzione $g(x,y)$, $Z = g(X,Y)$ definisce una nuova v.c. Z es la media dei voti su 2 esami $g = x+y$ [illegible] v.

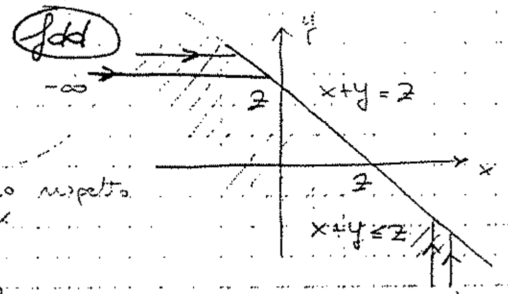
CALCOLO DELLA FUNZIONE DI DISTRIBUZIONE

Definisco $D_z = \{(x,y) : g(x;y) \le z\}$

$$F_Z(z) : P(Z \le z) = P(g(X;Y) \le z) = P((X,Y) \in D_z) = \iint_{D_z} f_{XY}(x,y)\,dx\,dy$$

Esempio: Data $f_{XY}(x,y)$, trovare la fdd e la ddp di $Z = X+Y$

fdd



$D_z = \{(x,y) \mid x+y \le z\}$

$$F_Z(z) = \int_{-\infty}^{+\infty} \int_{-\infty}^{z-y} f_{XY}(x,y)\,dx\,dy = \int_{-\infty}^{+\infty} \int_{-\infty}^{z-x} f_{XY}(x,y)\,d$$

tutti i possibili valori di y

Per calcolare la densità riscrivo la formula con un cambio di variabile:

$\tilde{y} := y + x$ $\quad F_Z(z) = \int_{-\infty}^{\infty} \int_{-\infty}^{z} f_{XY}(x, \tilde{y} - x)\,d\tilde{y}\,dx$ $\quad$ ho sostituito z in $\tilde{y}$

ddp Faccio la derivata $\dfrac{dF_Z(z)}{dz} \Rightarrow f_Z(z) = \int_{-\infty}^{+\infty} f_{XY}(x, z-x)\,dx = \int_{-\infty} f_{XY}(z-y, y)\,dy$

se non so altro sulla densità congiunta devo fermarmi qui

TEOREMA Se X e Y sono indipendenti ($\Leftrightarrow f_{XY}(x,y) = f_X(x) f_Y(y)$) allora la ddp di $Z = X+Y$ è

$$f_Z(z) = \int_{-\infty}^{+\infty} f_X(x) f_Y(z-x)\,dx = \int_{-\infty}^{+\infty} f_X(z-y) f_Y(y)\,dy = f_X * f_Y$$

CONVOLUZIONE

NB NON è vero che la densità della somma è la somma delle densità !! → si usa la convoluzione !

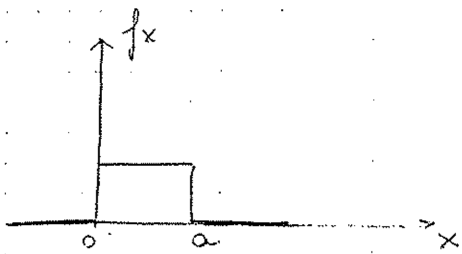
Esempio: Trovare la ddp della v.c. $Z = X+Y$ quando X e Y sono entrambe

(a) uniformi $\left(f_X(x) = \frac{1}{a},\ 0 \le x \le a\right)$

(b) esponenziale $\left(f_X(x) = \lambda e^{-\lambda x},\ x \ge 0\right)$

a – $f_Z(z) = \int_{-\infty}^{\infty} f_X(x) f_X(z-x)\,dx$

perché hanno la stessa distribuzione



Integrando $\neq 0$ quando $(0 \le x \le a)$ & $(0 \le z-x \le a)$

$\Rightarrow (0 \le x \le a)$ & $(z-a \le x \le z)$

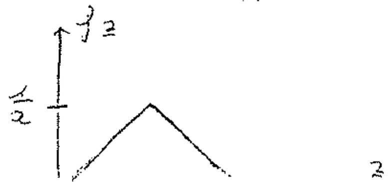
Quindi $z < 0 \Rightarrow f_Z(z) = 0$

$z > 2a \Rightarrow f_Z(z) = 0$

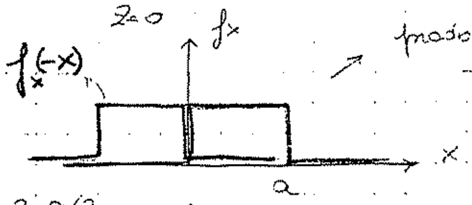
$$\int_{\max(0,z-a)}^{\min(a,z)} \dots$$

$0 \le z \le a \Rightarrow f_Z(z) = \dfrac{1}{a^2}\int_0^z dx = z/a^2$

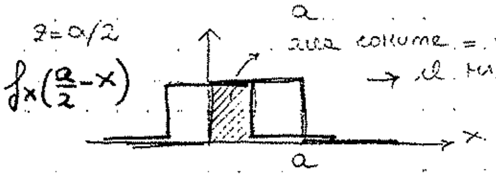
$a \le z \le 2a \Rightarrow f_Z(z) = \dfrac{1}{a^2}\int_{z-a}^{a} dx = \dfrac{2a-z}{a^2}$



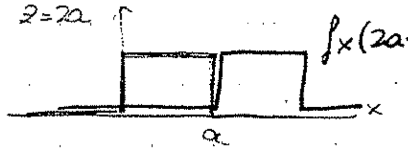
NB la densità della v.c. Z ottenuta come somma di 2 v.c. [illegible] densità [illegible] ≠



prodotto nullo $\Rightarrow \int = 0$



area colonne = prodotto
→ il max dell'integrale si ha quando le due densità si coprono a vicenda

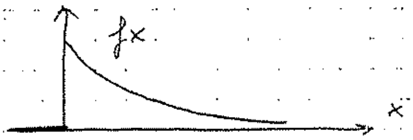


→ prodotto ancora nullo $\Rightarrow \int = 0$

b.- $f_Z(z) = \int_{-\infty}^{\infty} f_X(x) f_X(z-x)\, dx$

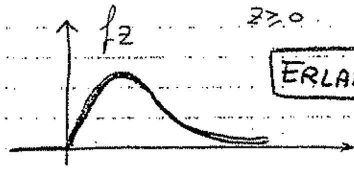
$f_X \neq 0$ per $x \geq 0$

$f_X(z-x) \neq 0$ per $z-x \geq 0$



Integrando $\neq 0$ quando $(x \geq 0)$ & $(x \leq z)$. Quindi:

$$f_Z(z) = \int_0^z \lambda e^{-\cancel{\lambda x}} \lambda e^{-\lambda(z-\cancel{x})} dx = \underline{\lambda^2 z e^{-\lambda z}}$$



$\boxed{\text{ERLANG - 2}}$

__Esercizio__: Date X e Y indipendenti con la medesima ddp esponenziale trovare la fdd e la ddp di $Z = \frac{Y}{X}$

Se Y è il tempo che aspetto l'autobus e X il tempo d'attesa del mio amico chi è il + sfortunato dei 2?

__Esercizio__: Trovare la ddp della v.c. somma di 2 dadi (X e Y). In particolare P(somma = 7) = ? METODO GRAFICO

MOMENTI DI V.C. CONGIUNTE

MEDIA DI $Z = g(X,Y)$

Per def $\underline{E[Z] = \int_{-\infty}^{\infty} z f_Z(z)\, dz}$ In realtà posso evitare il calcolo di f_Z

__TEOREMA__ $\underline{E[g(X,Y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y) f_{XY}(x,y)\, dx\, dy}$

__LINEARITÀ DI $E[\cdot]$__: $\underline{E[X+Y] = E[X] + E[Y]}$

Dimostrazione: $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x+y) f_{XY}(x,y)\, dx\, dy = \int_{-\infty}^{\infty} x \int_{-\infty}^{\infty} f_{XY}(x,y)\, dy\, dx +$

$+ \int_{-\infty}^{\infty} y \int_{-\infty}^{\infty} f_{XY}(x,y)\, dx\, dy = \int_{-\infty}^{\infty} x f_X(x)\, dx + \int_{-\infty}^{\infty} y f_Y(y)\, dy = E[X] + E[Y]$

cvd

Più in generale $\underline{E[\sum a_i X_i] = \sum a_i E[X_i]}$ somma e $E[\cdot]$ possono sempre commutare tra di loro

MOMENTI CONGIUNTI

- Momenti di ordine n

$$\underline{m_{kr} := E[X^k Y^r] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x^k y^r f_{XY}(x,y)\, dx\, dy} \qquad \boxed{k + r = n}$$

__Momenti di ordine 1__: $E[X], E[Y]$

__Momenti di ordine 2__: $E[X^2], E[XY], E[Y^2]$ $k = 0, 1, 2$

- Momenti centrali di ordine n

$$\mu_{kr} := E\left[(X-E[X])^k (Y-E[Y])^r\right], \quad \boxed{k+r=n}$$

e $r=0$
$k=2 \Rightarrow$ Varianza di X
$r=2 \Rightarrow$ " di Y

NOTA In genere $E[XY] \neq E[X]E[Y]$

COVARIANZA

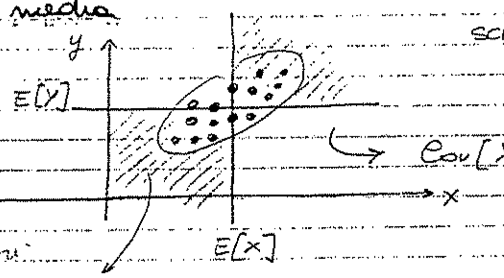
μ_{11} è detto COVARIANZA e si indica con $Cov[X,Y]$ oppure σ_{XY}

Interpretazione $Cov[X,Y] := E\left[\underbrace{(X-E[X])}_{\text{diff. tra } X \text{ e il suo valore atteso}}\underbrace{(Y-E[Y])}_{\text{idem}}\right]$

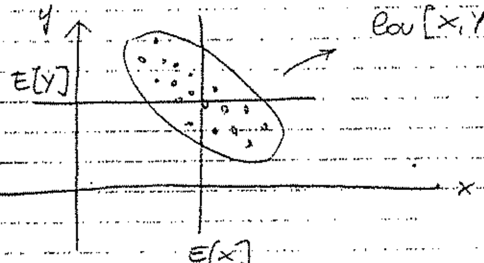
se è >0 è uno scostamento in eccesso

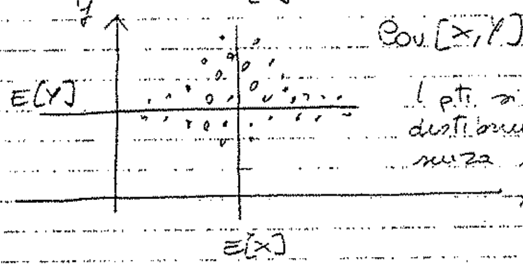
Se $Cov[X,Y]>0$
valore atteso del prodotto >0
$\Rightarrow$ o entrambi $+$ o entrambi $-$

in media valori di X superiori alla media sono associati a valori di Y superiori alla media



i valori devono essere concentrati in questi 2 quadranti





VANTAGGIO: per studiare la densità di probabilità dobbiamo ricorrere a grafici in 2 dimensioni → difficile ⇒ la cov. è un n° che ci dice qualcosa di utile

21 - 3 - 2001

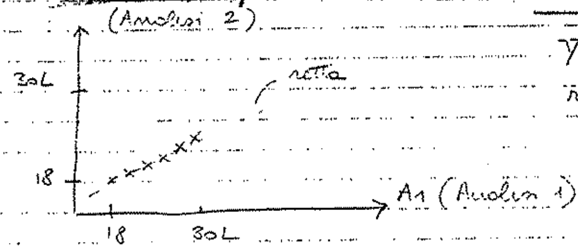
PROPRIETÀ (a) $\sigma_{XY} = Cov[X,Y] = E[XY] - E[X]E[Y]$

(b) $\sigma_{XY}^2 \le \sigma_X^2 \sigma_Y^2 \quad (|\sigma_{XY}| \le \sigma_X \sigma_Y)$

CASO LIMITE ← (c) $\sigma_{XY}^2 = \sigma_X^2\sigma_Y^2$ se e solo se $P\left(Y = \frac{\sigma_{XY}}{\sigma_X^2}(X-m_X)+m_Y\right)=1$

NON c'è libertà tra X e Y ⇒ NON ci sono 2 fonti di casualità

Y si può ottenere da X con una relazione lineare con probabilità 1



Serve x "quantificare" la covarianza

Def: Il coefficiente di correlazione r è dato dalla seguente formula:

$$r_{XY} := \frac{\mu_{11}}{\sqrt{\mu_{20}\mu_{02}}} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$$

→ dice quanto forte è il legame tra le 2 variabili → $r \approx 0,1 \Rightarrow$ legame debole
$r \approx 1 \Rightarrow$ " forte

NOTA: Proprietà (b) $\Rightarrow |r_{XY}| \le 1$ sempre

Proprietà (c) $\Rightarrow |r_{XY}| = 1$ se e solo se c'è un legame lineare tra X e Y (con probabilità 1)

OSSERVAZIONE: Le v.c. X,Y hanno lo stesso coeff. di correlazione delle v.c.

$(X-E[X]), (Y-E[Y])$ → traslare la densità di prob. non cambia la covarianza → il coefficiente di [illegible] ma la [illegible] [illegible]

[illegible] (cambia la media che [illegible]

Inoltre, $Var[X-E[X]] = Var[X]$ → sapere che uno sv. ha scostamento positivo NON mi dice nulla se l'altra ha scost. +

Def: X e Y tali che $\sigma_{XY}=0$ ($\Leftrightarrow$ $E[XY]=E[X]E[Y]$, vedi proprietà (a)) si dicono INCORRELATE

↳ (NB) se σ è zero NON posso dividere per zero

TEOREMA: Se X e Y sono indipendenti ne segue che X e Y sono anche incorrelate

NOTA: il viceversa NON è vero

X e Y indipendenti $f_{XY}=f_X f_Y$	X e Y incorrelate $E[XY]=E[X]E[Y]$ oppure $\sigma_{XY}=0$

→ L'indipendenza ha a che fare con la densità ⇒ è una condizione + forte mentre la correlazione è una relazione tra numeri

Indipendenza $\Rightarrow$ incorrelazione

$\not\Leftarrow$

PROPRIETÀ: Se $Z=X+Y$ Allora $\sigma_Z^2=\sigma_X^2+\sigma_Y^2+2\sigma_{XY}$

Dimostrazione:
$$E[(Z-E(Z))^2]=E[(X\pm Y-E[X]\mp E[Y])^2]=$$
$$=E[((X-E[X])\pm(Y-E[Y]))^2]=E[(X-E[X])^2+(Y-E[Y])^2\pm$$
$$+2(X-E[X])(Y-E[Y])]=E[(X-E[X])^2]+E[(Y-E[Y])^2]+$$
$$2E[(X-E[X])(Y-E[Y])]=\sigma_X^2+\sigma_Y^2\pm 2\sigma_{XY}$$ cvd

Se $Z=X-Y$ Allora $\sigma_Z^2=\sigma_X^2+\sigma_Y^2-2\sigma_{XY}$ → (demo uguale)

PROPRIETÀ: Se $Z=X\pm Y$ con X e Y incorrelate ($\sigma_{XY}=0$). Allora
$$\sigma_Z^2=\sigma_X^2+\sigma_Y^2$$

ERRORE GRAVE $Var[X-Y]=\sigma_X^2-\sigma_Y^2$ (X e Y incorrelate) NO!!

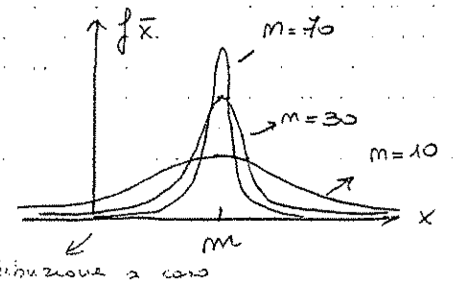
Esempio: Data una sequenza di v.c. $X_1, X_2 \dots X_n$ i.i.d. (= indipendenti identicamente distribuite ⇒ stessa densità di prob. e indip.), con valor medio m e varianza σ^2, calcolare valor medio e varianza della media aritmetica ↳ $\bar{X}_n=\frac{1}{n}\sum_{i=1}^{n}X_i$ ↳ è anch'essa una v.c.

es. voti di analisi 1

$$E[\bar{X}_n]=E\left[\frac{1}{n}\sum_1^n X_i\right]=\frac{1}{n}\sum_1^n E[X_i]=$$
$$=\frac{\cancel{n}m}{\cancel{n}}=m$$ → perché sono identicamente distribuite

$$Var[\bar{X}_n]=Var\left[\frac{1}{n}\sum_1^n X_i\right]=\frac{1}{n^2}\sum_{i=1}^{n}Var[X_i]=\frac{\cancel{n}\sigma^2}{n^{\cancel{2}}}=\frac{\sigma^2}{n}$$ → xchè hanno tutte la stessa varianza

NON ci sono prodotti incrociati xchè sono indipendenti!



→ Su questo fatto si basano tutte le tecniche di sondaggio

VARIABILI CASUALI VETTORIALI ($n>2$)

$$X(\cdot)=\begin{bmatrix}X_1(\cdot)\\X_2(\cdot)\\\vdots\\X_n(\cdot)\end{bmatrix}$$

ddp $f_X(x)=f_X(x_1,x_2\dots x_n):\mathbb{R}^n\to\mathbb{R}^1$

Valore atteso $E[X]=\begin{bmatrix}E[X_1]\\\vdots\\E[X_n]\end{bmatrix}$

PROPRIETÀ di $E[X]$ • X e Y v.c. vettoriali, $Z = AX + BY$ (A e B sono matrici ⇒ Z potrebbe essere un vettore) ⇒ $E[Z] = A E[X] + B E[Y]$

matrice, matrice

Def: La matrice varianza di $X(\cdot)$ è

$$\underbrace{Var[X]}_{\text{matrice quadrata}} := E\left[(X - E[X])(X-E[X])^T\right] = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \dots & \sigma_{1m} \\ \sigma_{21} & \sigma_{22} & \dots & \vdots \\ \vdots & & & \vdots \\ \sigma_{m1} & \sigma_{m2} & \dots & \sigma_{mm} \end{bmatrix}$$

coi vettori non ha senso fare il quadrato

interessa solo la parte sup. xchè sotto è ...

dove:

$$\sigma_{ij} = E\left[(X_i - E[X_i])(X_j - E[X_j])\right]$$

$i \neq j$ covarianza
$i = j$ varianza di i

↳ non c'è il trasposto xchè sono numeri

- $\sigma_{ij} = \sigma_{ji}$ ⇒ $Var[X]$ è simmetrica
- $m = 2$ ⇒ $Var\left[\begin{bmatrix} X \\ Y \end{bmatrix}\right] = \begin{bmatrix} \sigma_X^2 & \sigma_{XY} \\ \sigma_{XY} & \sigma_Y^2 \end{bmatrix}$ covarianze

diagonale principale = varianze delle singole v.c.

TEOREMA: $\underset{(m\times1)}{Y} = \underset{(m\times m)}{A}\underset{(m\times1)}{X}$ ⇒ $Var[Y] = A\, Var[X] A^T$

↳ nel caso scalare avrei $a^2 Var[x]$

Dimostrazione: $E[(AX - E[AX])(AX - E[AX])^T] = E[A(X - E[X])(X - E[X])^T A^T]$
$= A\, E[(X - E[X])(X-E[X])^T] A^T = A\, Var[X] A^T$ cvd

TEOREMA: $Var[X] \geq 0$ semidef. positiva

(si ricorda che $S = S^T \in \mathbb{R}^{m\times m}$ è ≥ 0 se $\underbrace{x^T S x}_{\text{forma quadratica}} \geq 0 \quad \forall x \in \mathbb{R}^m$)

Dimostrazione: Sia $a \in \mathbb{R}^m$ $\left(a = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_m \end{bmatrix}\right)$

Si consideri la v.c. $Y = a^T X = a_1 X_1 + a_2 X_2 + \dots a_m X_m$

Allora $Var[Y] = a^T Var[X] a$ Dato che $Var[Y] \geq 0$, $\forall a \in \mathbb{R}^m$ ⇒

$a^T Var[X] a \geq 0 \quad \forall a \in \mathbb{R}^m$ cvd

Def X e Y v.c. vettoriali si dicono indipendenti se, essendo $Z := \begin{bmatrix} X \\ Y \end{bmatrix}$ vett.

risulta che $f_Z\left(\begin{vmatrix} x \\ y \end{vmatrix}\right) = f_X(x) f_Y(y)$

Def: X e Y v.c. vettoriali si dicono incorrelate se

$$Cov[XY] := \underbrace{E\left[(X - E[X])(Y - E[Y])^T\right]}_{m \times m} = 0 \quad (\Leftrightarrow E[XY^T] = E[X]E[Y]^T)$$

Matrice NON necessariamente quadrata

NOTA Se X, Y sono indipendenti ⇒ incorrelate ($\not\Leftarrow$)

doppio prodotto

PROPRIETÀ: Se $Z = X \pm Y$, allora $Var[Z] = Var[X] + Var[Y] \pm Cov[X,Y] \pm Cov[Y,X]$

PROPRIETÀ: $Z = X \pm Y$, X e Y incorrelate $Var[Z] = Var[X] + Var[Y]$

LEGGI DEI GRANDI NUMERI

→ Importante xchè permette di legare tutta la teoria della prob. alla realtà fisica

CONVERGENZA DI SUCCESSIONI DI V.C.

Def Una successione di v.c. è una sequenza infinita di v.c. $X_1(\cdot) X_2(\cdot) \dots X_n(\cdot) \dots$

Esempio: Replico n volte in modo indipendente l'esperimento che mi fornisce il valore di una certa v.c. (× es, lancio n volte un dado). In tal modo ottengo $X_1, X_2 \dots X_n$ indip. tra loro (e con la stessa ddp se l'esperimento è sempre lo stesso).

Definisco $\bar{X}_n := \frac{1}{n}\sum_{i=1}^{n} X_i$ (media aritmetica dei primi n risultati) ↳ $\bar{X}_1, \bar{X}_2 \dots \bar{X}_n$

Per $n \to \infty$ sia $\{X_n(\cdot)\}$ che $\{\bar{X}_n(\cdot)\}$ sono delle successioni di v.c.

$\{X_n(\cdot)\}$: NON ci aspettiamo che converga

$\{\bar{X}_n(\cdot)\}$: ↳ $n \to \infty$ dovrebbe convergere alla media che è 3,5

CONVERGENZA IN DISTRIBUZIONE: $\{X_n(\cdot)\}$ converge a $X(\cdot)$ in distribuzione

$$\lim_{n\to\infty} F_{X_n}(x) = F_X(x)$$ (conv. puntuale) in ogni pto di continuità di F

c'è anche la conv. uniforme

Convergenza + debole possibile xché non mi dice nulla su X_n e X. (es: mi dice che tendono a diventare gaussian...)

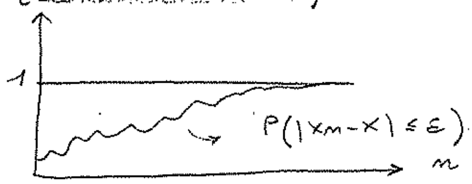
22-3-2001

CONVERGENZA IN PROBABILITÀ: $\{X_n(\cdot)\}$ converge a $X(\cdot)$ in prob. se, $\underline{\forall \varepsilon > 0}$

$$\lim_{n\to\infty} P(\underbrace{|X_n - X| \le \varepsilon}_{}) = 1$$

il modulo dell'errore è piccolo quanto vogliamo con probabilità 1

- ciò che converge non sono le v.c. ma la sequenza delle probabilità $\{P(|X_n - X| \le \varepsilon)\}$



- Convergenza in probabilità $\Rightarrow$ convergenza in distribuzione ($\not\Leftarrow$)

CONVERGENZA IN MEDIA QUADRATICA: $\{X_n(\cdot)\}$ converge a $X(\cdot)$ in media quadratica se

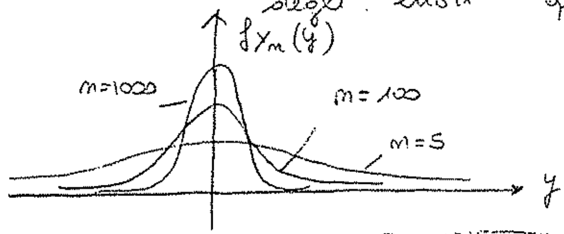
$$\lim_{n\to\infty} E\underbrace{[(X_n - X)^2]}_{\text{errori quadratici}} = 0 \qquad \left(\underset{n\to\infty}{\text{l.i.m.}}\, X_n = X\right)$$

l.i.m. → limite in media quadratica

Ciò che converge non è la v.c. ma la successione degli errori quadratici

Graficamente:

$Y_n := X_n - X$



Il valor quadratico medio dice quanto la densità si schiaccia intorno alla origine. Per $n \to \infty$ la densità di prob di Y_n tende ad un impulso $\Rightarrow X_n - X \to 0 \Rightarrow X_n \to X$

PROPRIETÀ: Se $X = a$ = costante

$$\left.\begin{aligned} &\lim_{n\to\infty} E[X_n] = a \\ &\lim_{n\to\infty} Var[X_n] = 0 \end{aligned}\right\} \Leftrightarrow \underset{n\to\infty}{\text{l.i.m.}}\, X_n = a$$

PROPRIETÀ: La convergenza in media quadratica $\Rightarrow$ convergenza in prob. ($\not\Leftarrow$)

CONVERGENZA CERTA: $\{X_n(\cdot)\}$ converge a $X(\cdot)$ certamente (o ovunque) se

$\underline{\forall \varepsilon > 0}$, $\underline{\exists n_0(\varepsilon)}$ t.c. $\underline{\forall n \ge n_0}$ e $\underline{\forall s \in S}$ $\quad |X_n - X| < \varepsilon$ → Convergenza + FORTE MA NON è MAI ottenibile

CONVERGENZA QUASI CERTA (CON PROB 1): $\{X_n(\cdot)\}$ converge a $X(\cdot)$ quasi certamente (o quasi ovunque, o con prob 1) se $\forall \varepsilon > 0 \; \exists \, n_0(\varepsilon)$ t.c. $\forall n \geq n_0$ e $\forall s \in A$ con $P(A) = 1$ $|X_n - X| \leq \varepsilon$

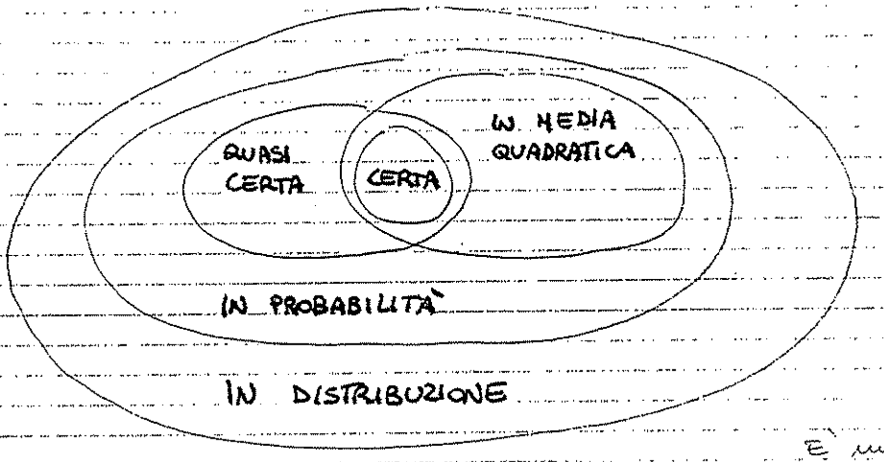
condizione verificata x tutti gli esiti che hanno prob = 1 → il che non vuol dire che gli esiti siano certi

NB certo ⇒ P = 1
$\not\Leftarrow$

Allora si scrive $P(\lim_{n\to\infty} X_n = X) = 1$

NB NON confondere il limite della probabilità (prima) con la probabilità del limite (ora)

Relazioni tra le nozioni di convergenza



E' una var. casuale

Media campionaria

LEGGE DEI GRANDI NUMERI

Sia $\{X_n(\cdot)\}$ una successione di v.c. e sia $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ e $\bar{m}_n = E[\bar{X}_n]$ (media della media campionaria).

Def Si dice che la successione $\{X_n(\cdot)\}$ obbedisce alla legge debole (forte) dei grandi numeri se $\bar{X}_n$ converge a $\bar{m}_n$ in probabilità (converge quasi certamente)

(media campionaria $\xrightarrow{?}$ media teorica della media campionaria)

TEOREMA (KOLMOGOROV) → RISULTATO FONDAMENTALE

Condizione necessaria e suff. affinché una successione $\{X_n(\cdot)\}$ di v.c. i.i.d. segua la legge forte dei grandi numeri è che ∃ il valor medio.

($\bar{X}_n \to m_n = E[X_i]$: la media campionaria converge alla media teorica)

↳ es: la media campionaria dei lanci del dado converge alla media teorica

Hp molto forte

TEOREMA CENTRALE DEL LIMITE

caso + semplice

con questa Hp la varianza della somma = somma delle varianze

Data una successione $\{X_n(\cdot)\}$ di v.c. i.i.d. con $E[X_n] = m$ e $Var[X_n] = \sigma^2$, sia

$$\tilde{X}_n := \sum_{i=1}^{n} X_i, \quad S_n := \frac{\tilde{X}_n - n \cdot m}{\sqrt{n\sigma^2}} = \frac{\tilde{X}_n - E[\tilde{X}_n]}{\sqrt{Var[\tilde{X}_n]}}$$

Somma delle v.c. normalizzata (in modo da avere media nulla e varianza unitaria).

NB le v.c. possono avere distribuzione come pare !!!

somma dei valori attesi = valore atteso della somma

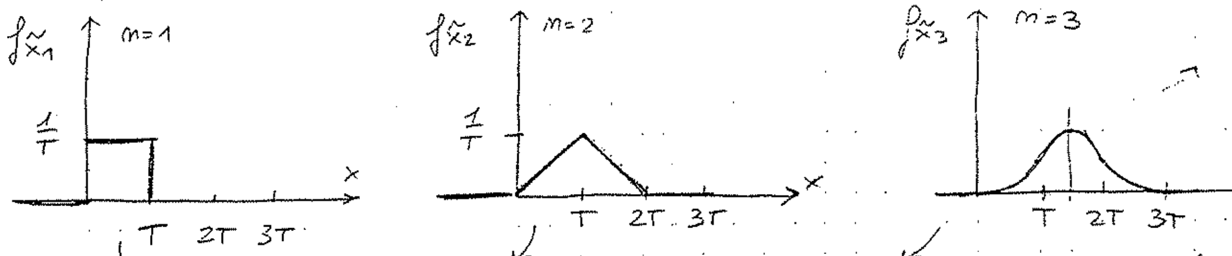
CONVERGENZA IN DISTRIBUZIONE

$\lim_{n\to\infty} F_{S_n}(x) = F_Z(x)$ dove $F_Z(x)$ è la fdd della gaussiana standard

OSSERVAZIONI

- La normalizzazione serve per enunciare correttamente il Teorema
  (PE. $\tilde{X}_n$ ha una media e una varianza che, al crescere di n, continuano a variare $\Rightarrow$ sarebbe difficile farli convergere a qualcosa).
  In realtà sia $\tilde{X}_n$ che $\bar{X}_n := \tilde{X}_n / n$ tendono a diventare gaussiane per $n \to \infty$
  (Tra $\tilde{X}_n$ ed S c'è solo una trasformazione lineare $\Rightarrow$ la forma NON cambia)

- L'approssimazione (spesso) è già buona per piccoli valori di n.



già per $n=5$ si vede bene la gaussiana (MA NON è STANDARD)

la media continua a spostarsi

OVVERO: la forma tende ad una gaussiana, solo che continua a spostarsi.

- Il teorema centrale del limite mi spiega perché ci sono in circolazione così tante gaussiane
  In tante situazioni una v.c. è il risultato di tanti fenomeni diversi che contribuiscono ad influenzarla (es: la statura, oltre alla genetica, può dipendere dall'ambiente, da patologie etc). e che sono + o - indipendenti tra di loro $\Rightarrow$ vale il Teo. centrale del limite.

(NB) Il Teorema centrale del limite vale anche in condizioni + complesse !!

VARIABILI CASUALI CONGIUNTAMENTE GAUSSIANE

Consideriamo la ddp congiunta $f_X(x_1, x_2)$ di 2 var. cas. X_1, X_2 indipendenti gaussiane con $E[X_i] = m_i$, $Var[X_i] = \sigma_i^2 \neq 0$

Dato che sono indip. $f_X(x_1, x_2) = f_{X_1}(x_1) \cdot f_{X_2}(x_2)$

$$f_X(x) = \frac{1}{\sqrt{(2\pi)^2 \sigma_1^2 \sigma_2^2}} e^{-\frac{1}{2}\left[\left(\frac{x_1 - m_1}{\sigma_1}\right)^2 + \left(\frac{x_2 - m_2}{\sigma_2}\right)^2\right]}$$

covarianza = 0 xché sono indipendenti $\Rightarrow$ incorrelate

$$m := E[X] = \begin{bmatrix} E[X_1] \\ E[X_2] \end{bmatrix} = \begin{bmatrix} m_1 \\ m_2 \end{bmatrix} \qquad V := Var[X] = \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix} \quad (*)$$

Allora $f_X(x)$ può essere riscritta in modo + compatto come

$$f_X(x) = \frac{1}{\sqrt{(2\pi)^2 \det V}} \cdot e^{-\frac{1}{2}(x-m)^T V^{-1} (x-m)} \qquad V^{-1} = \begin{bmatrix} 1/\sigma_1^2 & 0 \\ 0 & 1/\sigma_2^2 \end{bmatrix}$$

Idea: generalizziamo questa ddp congiunta al caso in cui V non è più diagonale

Def: Le v.c. $X_1, X_2 \dots X_N$ sono dette congiuntamente gaussiane se la loro densità congiunta è:

$$f_X(x) = \frac{1}{\sqrt{(2\pi)^N \det V}} e^{-\frac{1}{2}(x-m)^T V^{-1} (x-m)} \qquad m = \begin{bmatrix} m_1 \\ \vdots \\ m_N \end{bmatrix} \qquad \begin{matrix} V = V^T \geq 0 \\ V \in \mathbb{R}^{N \times N} \end{matrix}$$

PROPRIETÀ:

es: peso e altezza sono gaussiane (ma non indipendenti) e sono congiunt. gaussiane

- Si vede che $E[X] = m$, $Var[X] = V$
- Se X è un vettore di v.c. congiunt. gaussiane $\Rightarrow$ le ddp marginali $f_{X_i}(x_i)$

TEOREMA: Se X è un vettore di v.c. gaussiane, $\underset{(M\times1)}{Y} = \underset{(M\times N)}{A}\underset{(N\times1)}{X}$ è ancora un vettore di v.c. cong. gaussiane

Per un vettore di v.c. gaussiane bisogna conoscere il vettore m delle medie e la matrice V delle varianze.

La trasformazione lineare conserva la gaussianità

$$\left(E[Y] = A E[X] \ , \ Var[Y] = A\, Var[X]\, A^T \right)$$

NB: Se prendo 2 v.c. gaussiane e le sommo (o ne faccio una combinazione lineare) ottengo ancora una gaussiana e posso tenere traccia di media e varianza dopo ogni trasformazione. Se la trasformazione è lineare NON conserva la gaussianità

TEOREMA: Se delle v.c. congiunt. gaussiane $x_1, x_2 \dots x_N$ sono tra loro incorrelate ($Cov[x_i, x_J] = 0$, $i \neq J$ cioè $Var[X]$ è diagonale) $\Rightarrow$ sono anche indipendenti

Dimostrazione: Se V è diagonale $\Rightarrow f_X(x) = \prod_{i=1}^{n} f_{x_i}(x_i)$

Per le v.c. congiunt. gaussiane $\boxed{\text{INCORRELAZIONE} \Leftrightarrow \text{INDIPENDENZA}}$

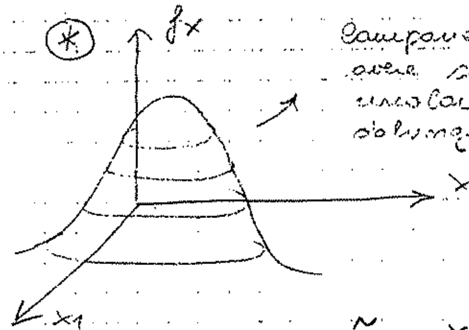
NON è vero in generale

Densità Congiunta della Gaus. per $n=2$

Se $r=0$ (cioè c'è indipendenza)

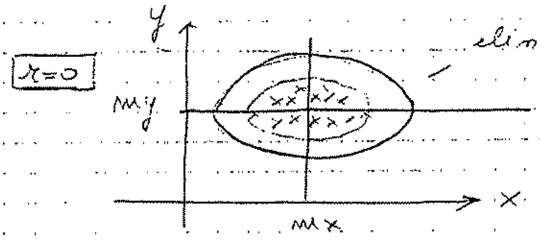
$f_{XY}(x,y) = \text{cost} \Rightarrow$ linee di livello cost $\rightarrow$ exp cost $\rightarrow$ esponente cost

$$\frac{(x-m_X)^2}{\sigma_x^2} + \frac{(y-m_Y)^2}{\sigma_y^2} = \text{cost}$$

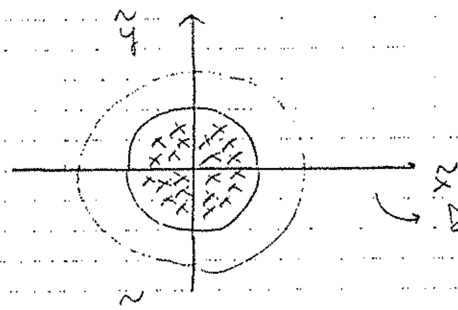


Campana che può avere sezione circolare o oblunga

$\boxed{r=0}$



ellissoidali



$\tilde{x} := \dfrac{x - m_x}{\sigma_x}$

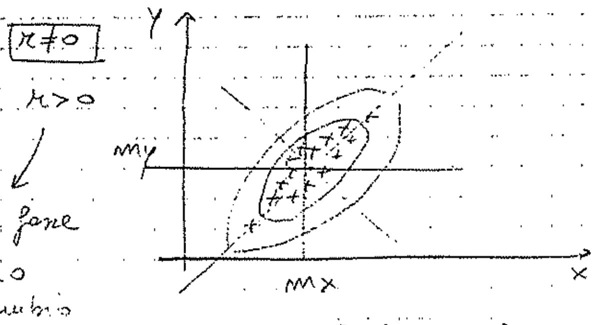
$\tilde{y} := \dfrac{y - m_y}{\sigma_y}$

Diventa un cerchio.

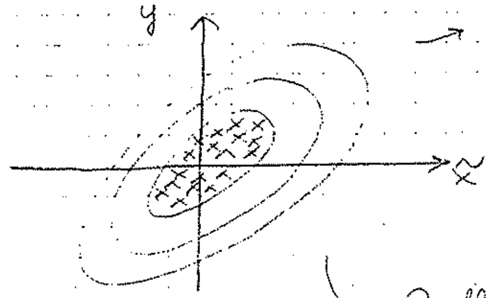
$\boxed{r\neq0}$

$r>0$

se fosse $r<0$ cambia l'inclinazione



Se la correlazione aumenta le ellissi si schiacciano di +



Resta un ellisse ma è centrata

Quelle normalizzate sono + facili da "leggere".

23 - 3 - 2001

DENSITA' e MOMENTI CONDIZIONATI

TEOREMA: Se X e Y sono v.c. vettoriali congiuntamente gaussiane (cioè $\begin{bmatrix} X \\ Y \end{bmatrix}$ è gaussiana), allora anche $f_{Y|X}(y|X=x)$ è gaussiana. Inoltre, essendo

$$Var\left[\begin{bmatrix} X \\ Y \end{bmatrix}\right] = \begin{bmatrix} V_{XX} & V_{XY} \\ V_{YX} & V_{YY} \end{bmatrix},$$

Si ha che :

matrice ...

- $E[Y|X=x] = m_Y + V_{YX} V_{XX}^{-1} (x - m_X)$

- $Var[Y|X=x] = V_{YY} - V_{YX} V_{XX}^{-1} V_{XY}$

(NB) $V_{XY} = V_{YX}^T$

COROLLARIO Per $N=2$ $\quad E[Y|X=x] = m_Y + r \dfrac{\sigma_Y}{\sigma_X}(x - m_X)$

$Var[Y|X=x] = \sigma_Y^2 (1 - r^2)$

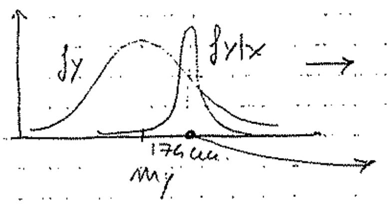
$$\frac{\sigma_{XY}}{\sigma_X \sigma_Y} \cdot \frac{\sigma_Y}{\sigma_X} = \frac{\sigma_{XY}}{\sigma_X^2} = \frac{V_{XY}}{V_{XX}}$$

(NB) Se le 2 v.c. sono incorrelate,
↳ $\sigma_{XY} = 0 \Rightarrow E[Y|X=x] = m_Y$

REGOLA MNEMONICA : $\dfrac{E[Y|X=x] - m_Y}{\sigma_Y} = r\left(\dfrac{x - m_X}{\sigma_X}\right)$ → se $x = m_X$ ⇓ $E[Y|X=x] = m_Y$

sorta di normalizzazione

Se non so niente per prevedere qual è l'h della prossima persona che entra dalla porta devo usare il valor medio (174 cm).
Se so che pesa 110 kg mi faccio un'idea ≠ della E

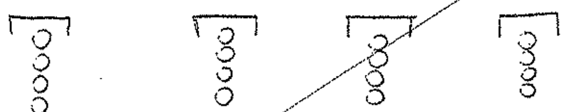


→ Conoscendo X la media si sposta e la varianza diminuisce
↳ se $r \approx 1$ ho una grande precisione
se $r = 0,5$ la varianza si ridurrebbe di 3/4 (0,75).

FINE della 1ª PARTE

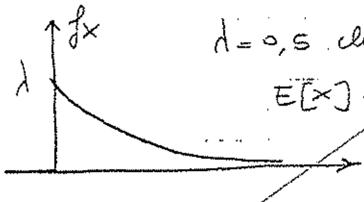
Es 1 del 7-4-2000

1.a



$\lambda = 0,5$ clienti/minuti

$E[X] = \dfrac{1}{\lambda} = 2$ minuti

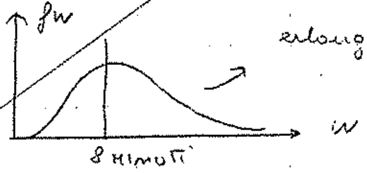


W = tempo di attesa $= X_1 + X_2 + X_3 + X_4$ (ERLANG-4)

$E[W] = ?$ $\quad Var[W] = ?$

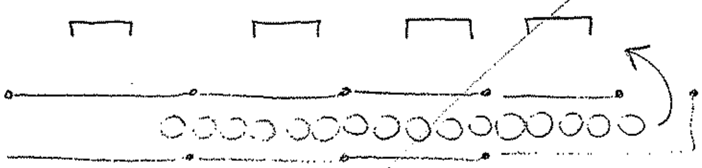
$E[W] = \sum_{i=1}^{4} E[X_i] = 4\,E[X] = 8$ minuti ↓ ci dice poco

i tempi di servizio sono indipendenti
⇒ le covarianze = 0



$Var[W] = \sum_{i=1}^{4} Var[X_i] = \dfrac{4}{\lambda^2} = 16 \text{ min}^2$
↳ deviazione standard = 4 minuti

1.b

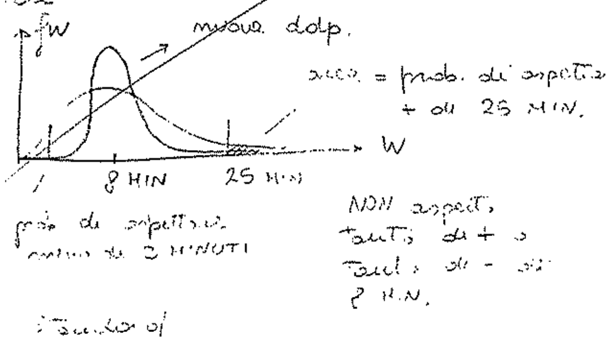


il ritardo è distribuito in modo + equo

Istanti in cui comincia ad essere servito un cliente sono eventi di P. con $\lambda = 2$

$E[W] = \sum_{i=1}^{16} E[X_i] = \dfrac{16}{\lambda} = 8$ minuti $\quad$ in effetti cambiare l'ordine dei clienti non può cambiare la coda

$Var[W] = \dfrac{16}{\lambda^2} = 4 \text{ min}^2$ $\quad$ dev. std. = 2 min



1.c $\quad P(W > 10) = ?$

Le Erlang tendono a diventare gaussiane perché, per il teo del limite centrale, sono somme di exp indipendenti

Tutti ... sono ... ... gaussiana standard

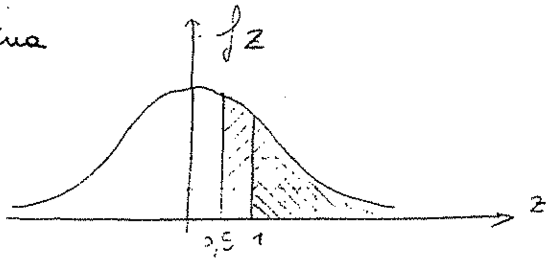
$$P(W>10) = P\left(\frac{W-E[W]}{\sigma_W} > \frac{10-E[W]}{\sigma_W}\right) \overset{\sim}{=} P\left(Z > \frac{2}{\sigma_W}\right) = 1 - F_Z\left(\frac{2}{\sigma_W}\right)$$

gaussiana standard

W circa gaussiana

1° caso : $\sigma_W = 4 \Rightarrow 1 - F_Z(0{,}5) = 1 - 0{,}69 = 0{,}31$

2° " : $\sigma_W = 2 \Rightarrow 1 - F_Z(1) = 1 - 0{,}84 = 0{,}16$

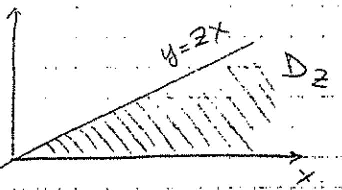


<u>Esercizio</u> per casa (Soluzione)

X, Y indip esponenziali $Z = \frac{Y}{X}$ $F_Z = ?$ $f_Z = ?$

$F_Z(z) = P(Z \le z) = P\left(\frac{Y}{X} \le z\right) =$

$x \ge 0$
$y \ge 0$
xché exp.



$f_{XY}(x,y) = \lambda e^{-\lambda x} \lambda e^{-\lambda y}$ $x, y \ge 0.$

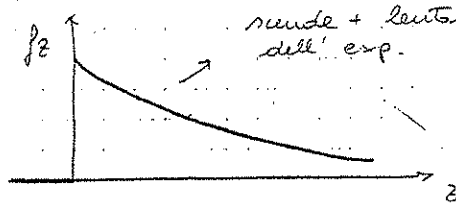
$$= \int_0^{+\infty} \int_0^{zx} \lambda e^{-\lambda x} \lambda e^{-\lambda y} \, dy \, dx =$$

$$= \int_0^{\infty} \lambda e^{-\lambda x} (1 - e^{-\lambda z x}) \, dx = \int_0^{\infty} \lambda e^{-\lambda x} dx - \lambda \int_0^{\infty} e^{-\lambda x (1+z)} dx =$$

$\int$ di una densità exp = 1

$= 1 - \frac{1}{1+z} = \frac{z}{1+z}$ $z \ge 0$

Derivando : $f_Z(z) = \frac{1}{(1+z)^2}$ $z \ge 0.$



<u>Idea</u> : Uso $E\left[\frac{Y}{X}\right]$ come una misura della sfortuna di tizio che aspetta l'autobus per Y istanti di tempo rispetto a Caio che aspetta X istanti di tempo.

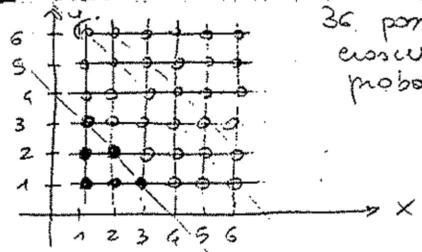
$\int z f_Z(z) \, dz \sim \frac{1}{z}\Big|_0^{\infty}$ $E[Z] = \infty$ NON è un buon indice per def la sfortuna

<u>Esercizio sul dado</u> (Soluzione) dato che sono indipendenti $\rightarrow$ convoluzione ? difficile

ddp della somma di 2 dadi

$P(S \le k) = P(X + Y \le k)$
generico

36 possibili esiti ciascuno con probabilità 1/36



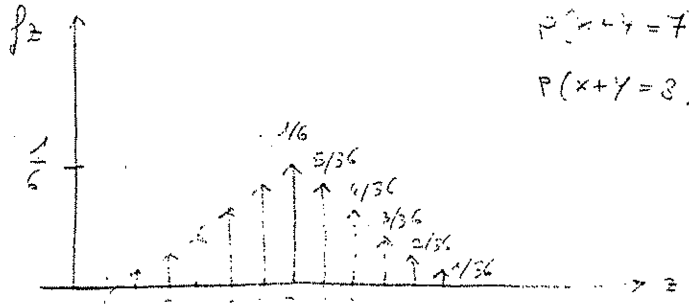
cerco la retta $y + x = 4$

$P(X + Y \le 4) = \frac{6}{36} = \frac{1}{6}$

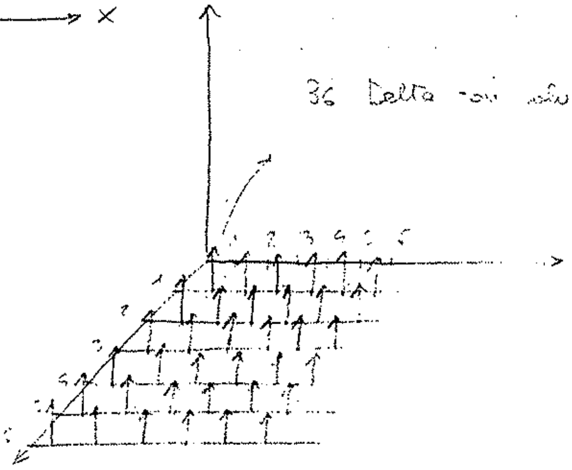
$P(X + Y \le k) = (k-1) + (k-2) + \ldots + 2 + 1$

$P(X+Y=7) = 1/6$

$P(X+Y=8) = 5/36$



36 Delta con altezza $\frac{1}{36}$



Es 2 del 7-4-2000

X, Y indip $\quad E[X]=E[Y] \quad Var[X]=Var[Y]=1$

$V = X - Y$

$W = 2X + Y$

$Var\left\{\begin{bmatrix} V \\ W \end{bmatrix}\right\} = ?$ → è fatta dalla $Var[V]$, $Var[W]$ e Var. di V e W

$$\begin{bmatrix} V \\ W \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & -1 \\ 2 & 1 \end{bmatrix}}_{A} \begin{bmatrix} X \\ Y \end{bmatrix} \rightarrow Var\begin{bmatrix} V \\ W \end{bmatrix} = A \, Var\left[\begin{bmatrix} X \\ Y \end{bmatrix}\right] A^T$$

xché X e Y sono indipendenti $\Rightarrow \sigma_{XY} = 0$

$$\begin{bmatrix} \sigma_V^2 & \sigma_{VW} \\ \sigma_{VW} & \sigma_W^2 \end{bmatrix} = \begin{bmatrix} 1 & -1 \\ 2 & 1 \end{bmatrix} Var\left[\begin{bmatrix} X \\ Y \end{bmatrix}\right] \begin{bmatrix} 1 & -1 \\ 2 & 1 \end{bmatrix}^T = \begin{bmatrix} 1 & -1 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ -1 & 1 \end{bmatrix}' = \begin{bmatrix} 2 & 1 \\ 1 & 5 \end{bmatrix}$$

Es 3 del 7-4-2000

3.a $\quad A+B = \mathcal{S} \quad A, B$ disgiunti $(AB = \{\emptyset\})$

Dimostrare che $E[Y] = E[Y|A]\,P[A] + E(Y|B)P(B)$ — specie di Teo. della prob. TOT. applicato ai valori medi

$$E[Y] = \int_{-\infty}^{\infty} y\, f_Y(y)\, dy = \int_{-\infty}^{\infty} y \left[f_{Y|A}(y|A)P(A) + f_{Y|B}(y|B)P(B) \right] dy =$$

$$= P(A) \int_{-\infty}^{\infty} y\, f_{Y|A}(y|A)\, dy + P(B) \int_{-\infty}^{\infty} y\, f_{Y|B}(y|B)\, dy =$$

$$= E[Y|A]P(A) + E[Y|B]P(B)$$ cvd

3.b $\quad A$ = bambino sopravvive $\quad B$ = muore

$$E[V] = E[V|A]P(A) + E[V|B]P(B) = 58 \cdot \frac{95}{100} + 5 \cdot \frac{5}{100} = 55{,}35$$

3.c $\quad 55{,}35 - 55 = 0{,}35$ → guadagno 1/3 di anno → ne vale la pena? vd. utilità

27-3-2001

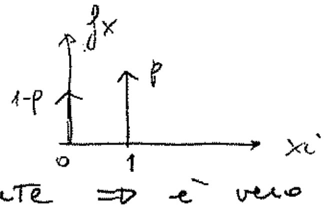
Tema d'esame del 23-4-99

- Per $n \to \infty$ la Binomiale di ordine n converge in distrib. ad una V.C. gaussiana con media np e varianza $np(1-p)$ $\quad$ V ☒ $\quad$ F ☐

  Y = binomiale di ordine n

  $Y = X_1 + X_2 + \dots X_n$, $\quad X_i \sim$ V.C. di Bernoulli

  v.c. i.i.d.

  sono verificate le Hp del Teo. centrale del limite ⇒ è vero che → a una gaussiana.

  BINOMIALE = SOMMA DI VAR. DI BERNOULLI

  $E[Y] = \sum E[X_i] = n\,E[X_i] = np$

  $Var[Y] = n\,Var[X_i] = np(1-p)$

- Dati 3 eventi $A\,B\,C$ con $ABC = \{\emptyset\} \Rightarrow P(A+B+C) = P(A)+P(B)+P(C)$ $\quad$ V ☐ $\quad$ F ☒

  $P(A+B+C) = P((A+B)+C) = P(A+B) + P(C) - P((A+B)C) =$

  $= P(A) + P(B) + P(C) - P(AB) - P(AC+BC) = P(A)+P(B)+P(C) - P(AB) - P(BC)$

  $+ P(ACBC)$!!

  Le intersezioni a 2 potrebbero essere vs le ∩ a 3 sono nulle

- Siano X, Y 2 v.c. congiunt. gaussiane $\Rightarrow Var[X+Y] = Var[X] + Cov(XY) + Var$ manca il 2
$\Rightarrow$ FALSA

- Siano X, Y 2 v.c. scalari congiunt. gaussiane $\Rightarrow Var[X|Y=y] \leq Var[X]$
$Var[X|Y=y] = (1-r^2) Var[X]$ VERA

- $W = g_1(X) + g_2(Y) \Rightarrow E[W] = g_1(E[X]) + g_2(E[Y])$ FALSA

es: $g_1(X) = X^2 \quad g_2(Y) = Y^2$
$E[X^2 + Y^2] = E[X^2] + E[Y^2] \overset{?}{=} E[X]^2 + E[Y]^2$
NO perché $E[X^2] \neq E[X]^2$ (in generale)

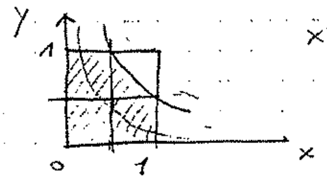
- X_k Tempi di interconnessione tra eventi di Poisson con densità λ ($\Rightarrow$ sono exp)
$\Rightarrow n \to \infty$ $\overline{X}_n$ converge in probabilità a $1/\lambda$ VERA
Tempi sono v. exp $\to$ sono i.i.d. $\Rightarrow$ x la legge dei grandi numeri
$\overline{X}_n \to m_n$ che è, x gli exp., $1/\lambda$

- XY incorrelate $\Rightarrow f_{XY}(x,y) = f_X(x) f_Y(y) \Leftrightarrow$ indipendenti FALSA
xchè indip $\Rightarrow$ incorr.
$\not\Leftarrow$

- XY congiunte con $r_{XY} < 0 \Rightarrow E[XY] < 0$
$r_{XY} < 0 \Rightarrow Cov[XY] < 0 \quad Cov[XY] = \underbrace{E[XY]}_{\text{non posso dire nulla}} - E[X]E[Y] < 0$ FALSA

- XY v.c. indipendenti, exp con $E[X] = \frac{1}{\lambda}$ $E[Y] = \frac{1}{\mu}$ $\Rightarrow W = X - Y$ è exp
con $E[W] = \lambda^{-1} - \mu^{-1}$
NB la differenza tra 2 exp può essere negativa MA una Var. exp NON può mai essere negativa $\Rightarrow$ FALSO

- XY indip., uniformi in $[0\ 1) \Rightarrow W = XY$ è uniforme in $[0\ 1]$
$F_W(w) = P(W \leq w) = P(XY \leq w)$



$XY = W \to$ famiglia di iperboli
$\downarrow$
NON può essere uniforme

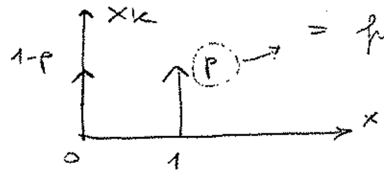
la congiunta è il prodotto delle marginali $\Rightarrow$ è $\neq 0$ solo nel quadrato

FALSO

$XY = 0{,}5$

$F_W(0{,}5) = P(W \leq 0{,}5) = P(XY \leq 0{,}5) > 0{,}5$ se fosse uniforme $F_W(0{,}5) = 0{,}5$

Esercizio 1 23-4-99



p = frazione di elettori che ha intenzione di votare $\Big\}$ $\dfrac{\text{elettori che votano}}{\text{elettori TOT.}}$

p = valore medio della var di Bernoulli

$$\hat{p} = \frac{\sum_{k=1}^{1000} X_k}{1000}$$

1.a $E[X_k] = 0 \cdot (1-p) + 1 \cdot p = p$

$V\ [\ \ ] = E[X^2] - E[\ \ ]^2$

$E[x_k^2] = 0^2 \cdot (1-p) + 1^2 \cdot p = p$

$Var[x_k] = p - p^2 = p(1-p) = pq$

1.b $\quad E[\hat{p}] = \frac{1}{1000} \sum_{k=1}^{1000} E[x_k] = \frac{\cancel{1000}\,p}{\cancel{1000}} = p$

$Var[\hat{p}] = \frac{1}{1000^2} \sum_{1}^{1000} Var[x_k] = \frac{1000\, p(1-p)}{1000^2} = \frac{p(1-p)}{1000}$

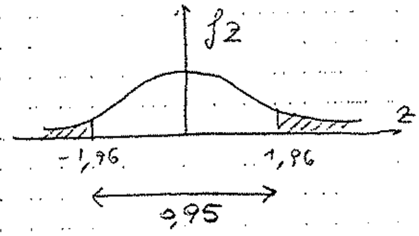
1.c $\quad \hat{p}$ = somma di v.c. i.i.d $\Rightarrow$ uso il teo. centrale del limite

1.d $\quad p = 0,5 \quad$ Trovare δ t.c. $P(p - \delta \leq \hat{p} \leq p + \delta) = 0,95$

↳ margine di errore

$P(-1,96 \leq Z \leq 1,96) = 0,95 \qquad P(|Z| \leq 1,96) = 0,95$

$Z = \frac{\hat{p} - E[\hat{p}]}{\sqrt{Var[\hat{p}]}} = \frac{\hat{p} - 0,5}{0,0158}$



$Var[\hat{p}] = \frac{0,25}{1000}$

$E[\hat{p}] = 0,5$

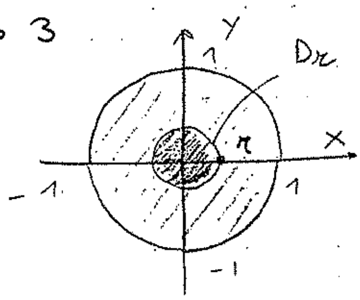
$P(|Z| \leq 1,96) = P\left(\left| \frac{\hat{p} - 0,5}{0,0158} \right| \leq 1,96 \right) \Rightarrow$

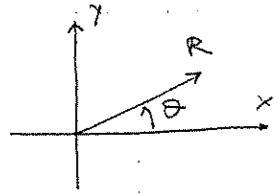
$P(\underbrace{0,5 - 1,96 \cdot 0,0158}_{\delta} \leq \hat{p} \leq 0,5 + \underbrace{1,96 \cdot 0,0158}_{\delta = 0,03}) = 0,95$

1000 interviste → errore del 3%

Esercizio 2 uguale a quello del 2000

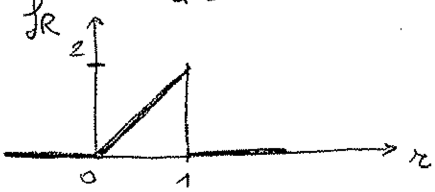
Esercizio 3



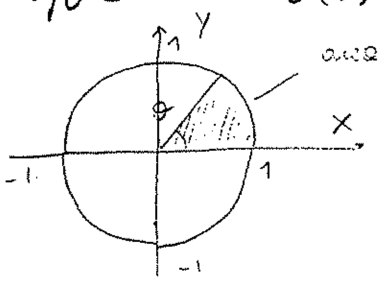


3.a $\quad f_R(r) = ? \qquad F_R(r) = P(R \leq r) = P((X,Y) \in D_r) = \frac{\pi r^2}{\pi} = r^2 \quad 0 \leq r \leq 1$

$f_R(r) = \frac{dF_R(r)}{dr} = 2r$

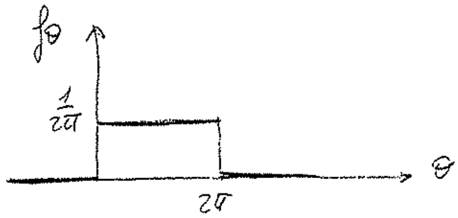


3.b $\quad f_\Theta = ? \qquad F_\Theta(\theta) = P(\Theta \leq \theta)\; P((X,Y) \in D_\theta) = \frac{\theta}{2\pi} \qquad 0 \leq \theta \leq 2\pi$

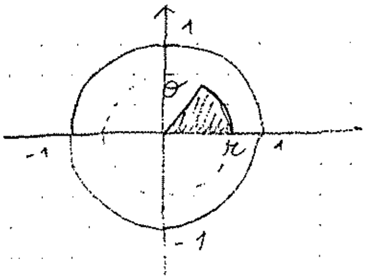


area $= \frac{\theta}{2\pi}$

$f_\Theta(\theta) = \frac{1}{2\pi}$



3.c $\quad f_{R\Theta}(r,\theta) = ? \qquad F_{R\Theta}(r,\theta) = P(R \leq r, \Theta \leq \theta) = P((X,Y) \in D_{r\theta}) =$



$$= \frac{\cancel{\pi} r^2 \cdot \frac{\theta}{2\pi}}{\cancel{\pi}} = \frac{r^2\theta}{2\pi}$$

$$f_{R\Theta}(r,\theta) = \frac{\partial^2 F_{R\Theta}}{\partial r \partial \theta} = \frac{2r}{2\pi} = \frac{r}{\pi}$$

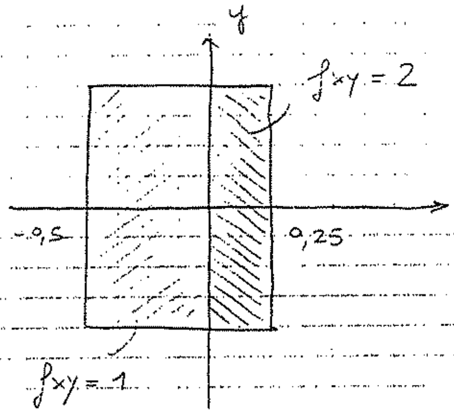
3.d r e θ sono indipendenti? $f_{R\Theta}(r,\theta) \overset{?}{=} f_R(r) f_\Theta(\theta)$

$$\frac{r}{\pi} = \cancel{2}r \cdot \frac{1}{\cancel{2}\pi}$$ Si sono indip.

28-3-2001

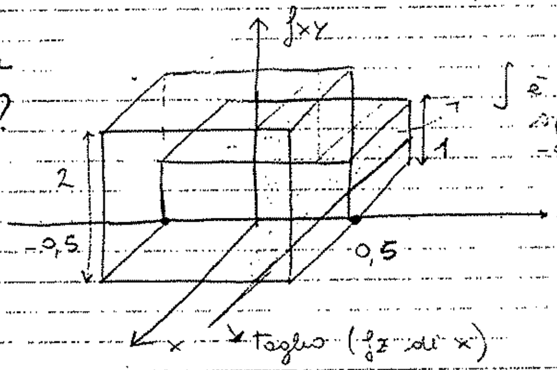
Tema d'esame 18/4/97

1. $f_{XY}(x,y) = \begin{cases} 1 & -0{,}5 \le x < 0 \\ & -0{,}5 \le y \le 0{,}5 \\ 2 & 0 \le x \le 0{,}25 \\ & -0{,}5 \le y \le 0{,}5 \\ 0 & \text{altrove} \end{cases}$



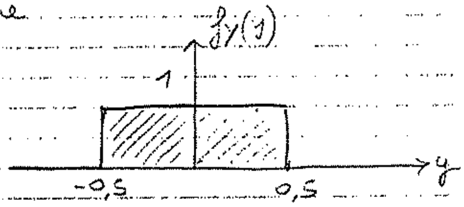
1.a

$f_Y(y) = ?$



$\int$ è uguale spostandomi tra $-0{,}5$ e $0{,}5$ ⇒ densità uniforme

$f_Y(y) = \begin{cases} 1 & |y| \le 0{,}5 \\ 0 & |y| > 0{,}5 \end{cases}$



1.b $f_{X|Y}(x|Y=y) = ?$ $\quad |y| \le 0{,}5$

$f_{X|Y}(x|Y=y) = \dfrac{f_{XY}(x,y)}{f_Y(y)}$ → è uniforme ⇒ cost ($=1$)

↓

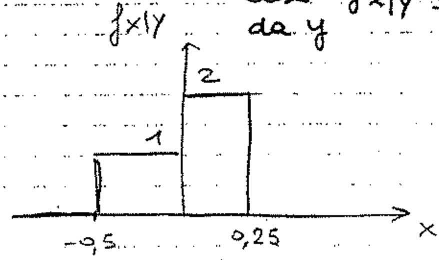
ho tante densità condizionate quante sono le y possibili

Indipendentemente da dove "taglio" ottengo sempre lo stesso risultato

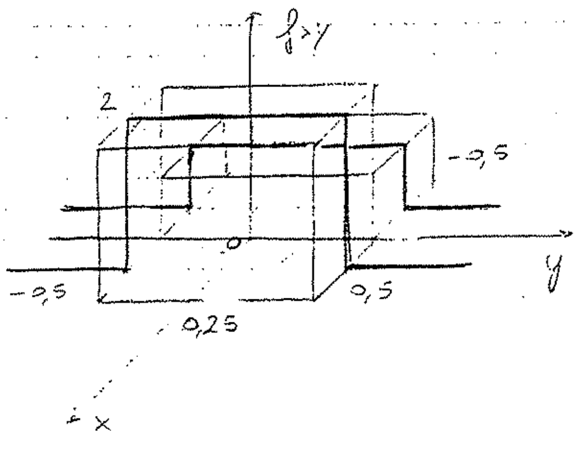
↓

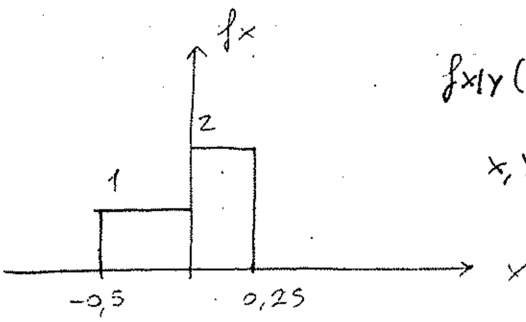
Si vede che in questo caso $f_{X|Y}$ **NON dipende** da y

$$f_{X|Y}(x|Y=y) = \begin{cases} 1 & 0{,}5 \le x < 0 \\ 2 & 0 \le x \le 0{,}25 \\ 0 & \text{altrove} \end{cases}$$



1.c





$f_{X|Y}(x|Y=y) = f_X(x)$

⇓

X, Y indipendenti

2. $Z := aX - bY \qquad \hat{Z} := \alpha X \qquad$ Conosco $E[Y^2], E[X^2], E[XY], a, b$.

Trovare α t.c. minimizza $\underset{\substack{\downarrow \\ \text{errore di stima}}}{E[(Z-\hat{Z})^2]}$

$$J = E[(Z-\hat{Z})^2] = E[(aX - bY - \alpha X)^2] = E[((a-\alpha)X - bY)^2] =$$

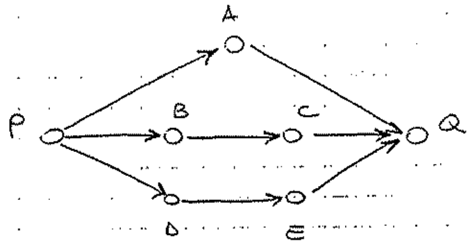
$$= E[(a-\alpha)^2 X^2 - 2b(a-\alpha)XY + b^2Y^2] = (a-\alpha)^2 E[X^2] - 2b(a-\alpha)E[XY] + b^2 E[Y^2]$$

$$\frac{dJ}{d\alpha} = 0 \qquad -\cancel{2}(a-\alpha)E[X^2] + \cancel{2}bE[XY] = 0 \Rightarrow \alpha E[X^2] = aE[X^2] - bE[XY]$$

$$\Rightarrow \alpha = \frac{aE[X^2] - bE(XY)}{E[X^2]}$$ $\rightarrow$ bisogna verificare che NON sia il MAX

$$\frac{d^2J}{d\alpha^2} = 2E[X^2] \geq 0$$ (supposto $E[X^2] > 0$ è un MIN).

4:



4.a $P(PAQ \text{ funziona}) = ? = P(PA \text{ funzione}, AQ \text{ funziona}) = p^2$

4.b $P(PBCQ \text{ funziona}) = p^3$

4.c $P(PAQ, PBCQ, PDEQ \text{ non funzionano tutti e tre}) = (1-p^2)(1-p^3)(1-p^3)$

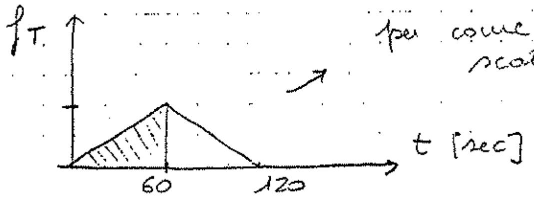
4.d. $P(\underset{\substack{\downarrow \\ \text{almeno 1 dei 3} \\ \text{sia funzionante}}}{P \text{ e } Q \text{ collegati}}) = 1 - \underbrace{(1-p^2)(1-p^3)^2}_{\text{probabilità che funzionino tutti}}$

vd. DOMANDE su fotocopie

Esercizio 2 del 30/6/99

A: 1 scatto al MINUTO + 1 scatto alla risposta   scatto = 220 lire

B: 5 lire al sec.



per come sono fatte le telefonate non avrò MAI 3 scatti

2.a $E[C|B] = E[\underset{\text{↳ 5 lire · sec}}{5T}] = 5E[T] = 300$ £

2.b $P(\underset{\substack{\downarrow \\ \text{solo se la} \\ \text{telefonata dura} \\ - \text{ di 1 MINUTO}}}{1 \text{ scatto}} | A) = P(T < 60) = 0{,}5$ (50% di possibilità)

2.c Per la compagnia B il costo medio è 300 £

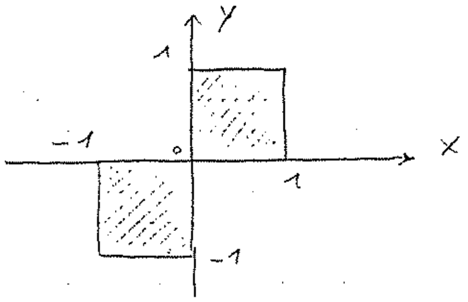
Se scelgo A $\quad E[C] = E[C|1 \text{ scatto}]P(1 \text{ scatto}) + E[C|2 \text{ scatti}]P(2 \text{ scatti})$

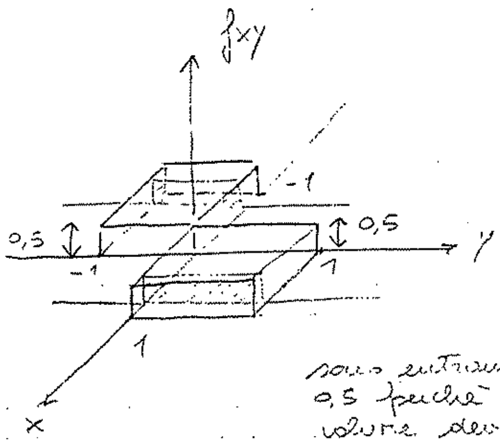
$$= 220 \cdot 0{,}5 + 440 \cdot 0{,}5 = 330 \text{ lire}$$

29-3-2001

Tema d'esame del 27-3-98

1)





sono entrambe alte 0,5 perché il volume deve essere = 1

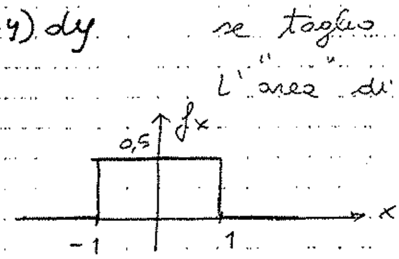
$$f_{XY}(x,y) = \begin{cases} 1/2 & 0<x<1,\ 0<y<1 \\ 1/2 & -1<x<0,\ -1<y<0 \\ 0 & \text{altrove} \end{cases}$$

→ (FB) stabilire quanto vale sui bordi

1.a $f_X(x) = ? = \int_{-\infty}^{\infty} f_{XY}(x,y)\,dy$ se taglio per valori di $x > 1 \to 0$, " " $x < -1 \to 0$

L'area di taglio è sempre uguale $= 1 \cdot \frac{1}{2}$

$$= \begin{cases} 1/2 & -1<x<1 \\ 0 & \text{altrove} \end{cases}$$



1.b $f_{X|Y}(x|Y=y) = ?$

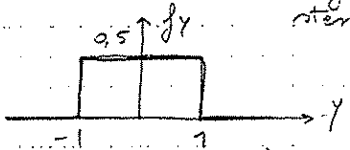
$f_{X|Y}(x|Y=y) = \dfrac{f_{XY}(x,y)}{f_Y(y)}$ → ∄ per $y>1$ o $y<-1$

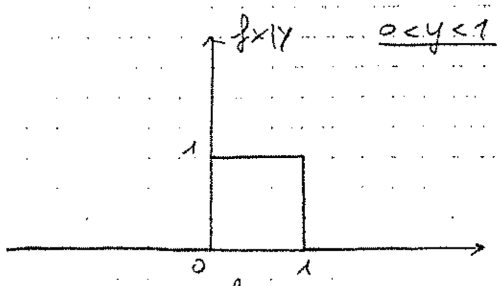
faccio un taglio MA NON integro, poi divido per il profilo stesso

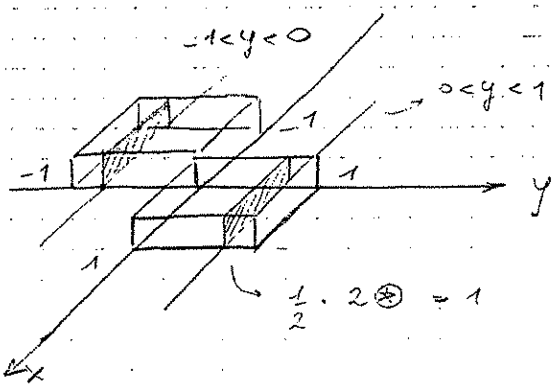
Per motivi di simmetria

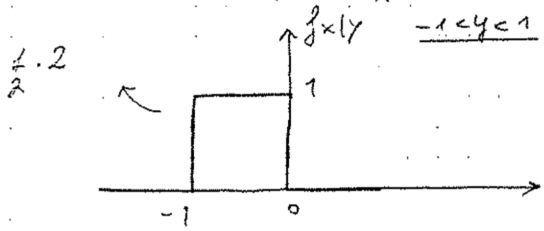
$f_Y(y) = f_X(x)$

vuol dire che dobbiamo dividere per $\frac{1}{2}$ (cioè moltiplicare f_{XY} per 2) ⊛



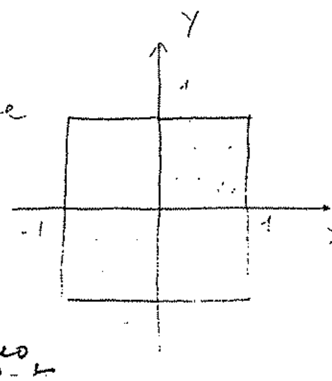






1.c X e Y sono indipendenti? $f_{X|Y} \neq f_X$ ⇒ NO

il prodotto delle marginali dà una "zolletta" con base



1.d indipendenti ⇒ incorrelate

Ma se non sono indipendenti potrebbero comunque essere incorrelate

Se sappiamo che $x > 0$ sappiamo qualcosa sul segno

$\mathrm{cov}[X,Y] = E\big[(X-\underset{0}{E[X]})(Y-\underset{0}{E[Y]})\big] = E[XY] > 0$ (infatti $P(XY>0)=1$)

oppure XY incorrelate $\Leftrightarrow \mathrm{cov}[XY]=0 \rightarrow E[XY] = E[X]\underset{0}{E[Y]}$

$Z := X\cdot Y$ assume solo valori $>0 \Rightarrow$ NON è vero

3) $E[X] = E[Y] = 0 \quad \mathrm{Var}[X] = \mathrm{Var}[Y] = 1 \quad \mathrm{cov}[XY] = 0{,}5$

$Z = X + aY \quad a?$ t.c. $\mathrm{cov}[X,Z] = 0$

$E[(X-E[X])(Z-E[Z])] = E[X(X+aY-\overset{=0}{E[X]}-a\overset{=0}{E[Y]})] = E[X(X+aY)] =$

$= E[X^2] + aE[XY] = 1 + a\cdot 1/2 = 0 \longrightarrow a = -2$

NB $\underline{\mathrm{Var}[X] = E[X^2] - E[X]^2} = E[X^2]$ perché $E[X]=0$

$\underline{\mathrm{cov}[XY] = E[XY] - E[X]E[Y]} = E[XY]$ perché $E[X] = E[Y] = 0$

2) $T > 1 \ (\lambda = 1) \Rightarrow$ B vince S \$ evento $A = \{T \le 1\}$

$T \le 1 \Rightarrow$ A vince T \$ evento $B = \{T > 1\}$

$f_T(t) = \lambda e^{-\lambda t}$

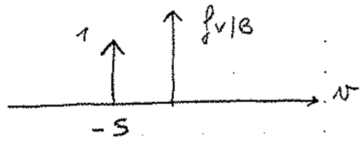
$F_T(t) = 1 - e^{-\lambda t}$

2.a $P(A) = ? \quad P(B) = ?$

$P(A) = P(T \le 1) = F_T(1) = 1 - e^{-1}$

$P(B) = 1 - P(A) = e^{-1}$

2.b $f_{V|B}(v|B) = ? \qquad P(V = -s|B) = 1$ dal pto di vista di B è una perdita



2.c $f_{V|A}(v|A) = ?$ NB $0 \le V \le 1$

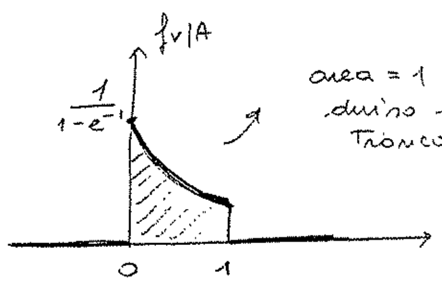
$\longrightarrow f_{V|A}(v|A) = \lim_{\Delta v \to 0} \dfrac{P(v \le V \le v + \Delta v,\ T \le 1)}{P(T \le 1)\Delta v} =$ dato che V è proporzionale a T

$\longrightarrow = \lim_{\Delta v \to 0} \dfrac{P(v \le T \le v+\Delta v, T \le 1)}{(1-e^{-1})\Delta v} = \lim_{\Delta v \to 0} \dfrac{P(v \le T \le v+\Delta v)}{(1-e^{-1})\Delta v} =$

non conta nulla perché $0 \le v \le 1$

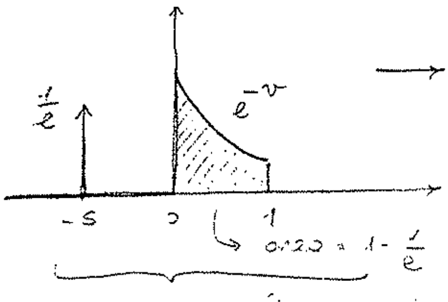
è la ddp dell'esp.

$= \begin{cases} \dfrac{\lambda e^{-\lambda v}}{1-e^{-1}} & 0 \le v \le 1 \\ 0 & \text{altrove} \end{cases} = \begin{cases} \dfrac{e^{-v}}{1-e^{-1}} & 0 \le v \le 1 \\ 0 & \text{altrove} \end{cases}$



area = 1 perché ho diviso l'exp troncata per $1-e^{-1}$

2.d $f_V(v) = ? \qquad f_V(v) = f_{V|A}(v|A)P(A) + f_{V|B}(v|B)P(B) = \dfrac{e^{-v}}{1-e^{-1}}\cdot(1-e^{-1}) + \overset{\delta}{1}\, e^{-1}$



DENSITÀ DI TIPO MISTO

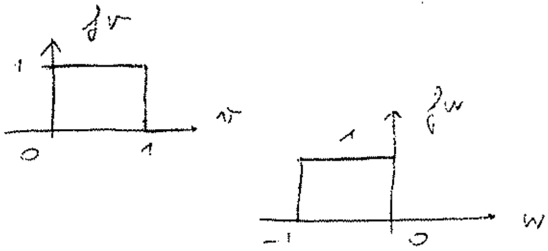
perché ho un impulso e un pezzo di ddp continua

30 - 3 - 2001

Tema d'esame del 18/7/2000

2) A) $p = 0{,}8 \Rightarrow X = V$

B) $p = 0{,}2 \Rightarrow X = W$



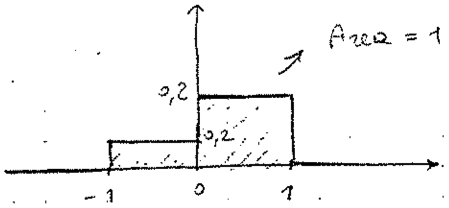
2.a $f_X(x) = ?$ Teo. della prob. Totale

$$F_X(x) = P(X \le x) = P(X \le x | A) P(A) + P(X \le x | B) P(B) =$$

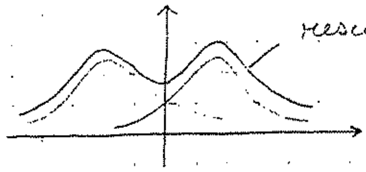
è $F_W(x)$ perché se si verifica l'evento $B \Rightarrow X = W$

$$= F_V(x) P(A) + F_W(x) P(B) = F_V(x) \cdot 0{,}8 + F_W(x) \cdot 0{,}2$$

Deriv.: $f_X(x) = f_V(x) \cdot 0{,}8 + f_W(x) \cdot 0{,}2 \rightarrow$ sovrapposizione delle 2 densità opportunamente scalate



detta **MISTURA** (MIXTURE) delle 2 densità precedenti (si ha quando retta insieme da eterogenei)



rescola o mistura di 2 gaussiane

2.b $E[X] = ?$ $Var[X] = ?$

$E[X] = E[X|A]P(A) + E[X|B]P(B) \longrightarrow$ Somma pesata delle 2 medie ("mescola" delle medie)

$$= 0{,}5 \cdot 0{,}8 - 0{,}5 \cdot 0{,}2 = 0{,}4 - 0{,}1 = 0{,}3$$

$$Var[X] = E\Big[\underbrace{(X - E[X])^2}_{Y}\Big] = E[Y|A]P(A) + E[Y|B]P(B) = \underbrace{Var[X|A]}_{1/12} P(A) + \underbrace{Var[X|B]}_{1/12} P(B)$$

$$= \frac{0{,}8}{12} + \frac{0{,}2}{12} = \frac{1}{12}$$

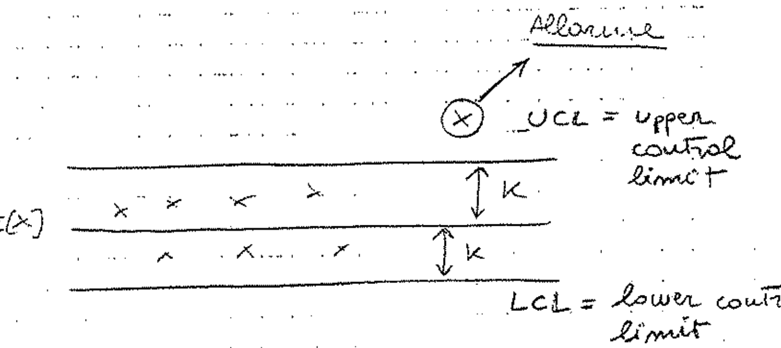
Tema d'esame 6/7/2000

2) $X \sim N(m, \sigma^2)$ $m = 6$ $\sigma^2 = 1$

Allarme quando $|X - E[X]| > k$

È dunque una carta di controllo

se k è troppo piccolo continuo a vedere allarmi per niente $\rightarrow$ k deve essere fissato in modo che la probabilità di falso allarme sia bassa



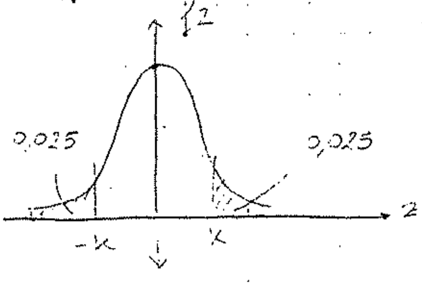
2.a P(falso allarme) = 5% $\Rightarrow$ Trovare k

$$P\left(\Big|\frac{X - E(X)}{\sigma}\Big| > \frac{k}{\sigma}\right) = 0{,}05 \qquad Z := \frac{X - E[X]}{\sigma} \qquad P(|Z| > k) = 0{,}05$$

$\sigma = 1$

5% $\rightarrow \boxed{k = 1{,}96}$ Valore da ricordare !!

(In ambito industriale si prende $k = 3$)



2.b $p = P(\text{falso allarme}) = 0{,}05$ Ho delle prove di Bernulli con $p = 0{,}05$

Per N prove di Bernulli la distribuzione è la BINOMIALE di ordine 1000

2.c $E[N] = ?$ $Var[N] = ?$

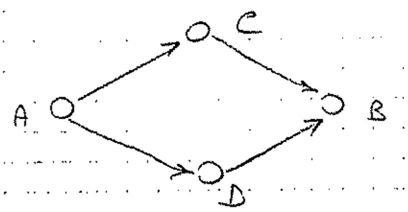
$E[N] = 1000 \cdot p = 50$ $Var[N] = 1000\, p(1-p) = 1000 \cdot 0{,}05\,(0{,}95) = \sigma_N^2$

2.d $P(N > 100) = ?$ Per evitare calcoli troppo complessi usiamo il Teo. centrale del limite (binomiale $\to$ gaussiana).

$Z := \frac{N-50}{\sigma_N}$ $P\left(Z > \frac{100-50}{\sigma_N}\right) = 1 - P\left(Z \le \frac{50}{\sigma_N}\right) = 1 - F_Z\left(\underbrace{\frac{50}{\sigma_N}}_{7{,}25}\right)$

Tema d'esame 2/11/98

2)



ACB : P(ingorgo = 0,1) $\to$ + di 2 ore

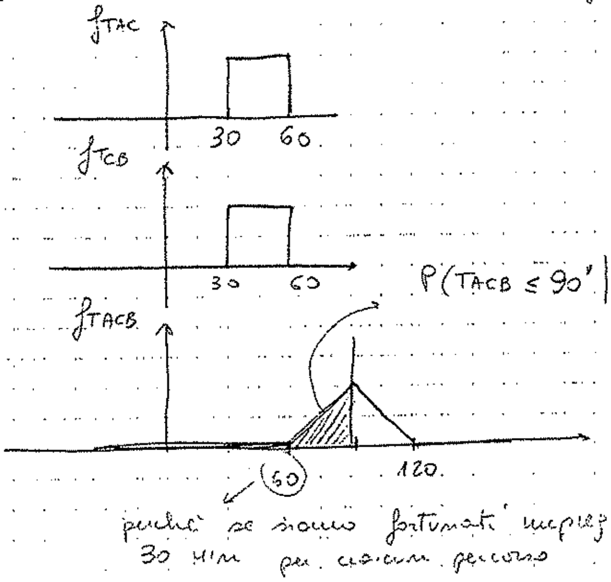
altrimenti T_{AC}, T_{CB} sono indip. e unif. in $[30', 60']$

2.a

$$P(T_{ACB} \le 90') = \underbrace{P(T_{ACB} \le 90' \mid \overset{\text{ingorgo}}{I})}_{0} P(I) + P(T_{ACB} \le 90' \mid \overset{\text{non c'è ingorgo}}{\bar{I}}) P(\bar{I})$$

perchè se c'è l'ingorgo impiego sempre + di 2 ore

$T_{ACB} = T_{AC} + T_{CB}$ somma di 2 uniformi



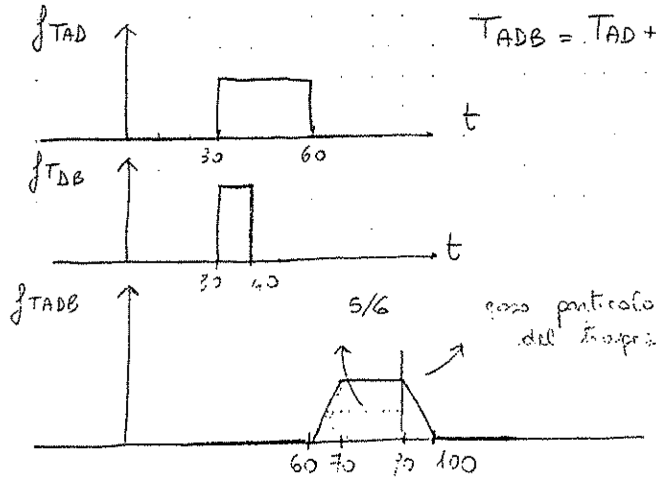
$P(T_{ACB} \le 90' \mid \bar{I}) = 0{,}5$

$P(T_{ACB} \le 90') = P(T_{ACB} \le 90' \mid \bar{I}) P(\bar{I})$

$= 0{,}5 \cdot 0{,}9 = 0{,}45$

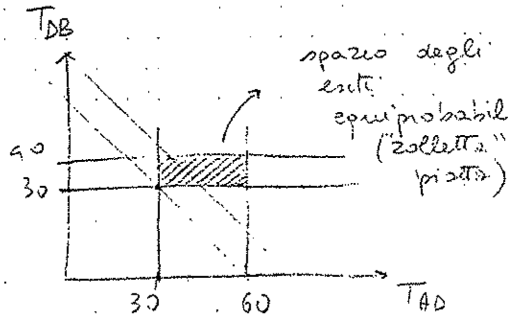
2.b

$T_{ADB} = T_{AD} + T_{DB}$



oppure

$X = T_{AD} + T_{DB}$



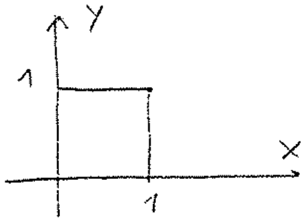
$F_X(x) = P(T_{AD} + T_{DB} \le x)$

faccio i conti con le aree spostando la retta

$P(T_{ADB} \le 90' \mid \bar{I}) = \frac{5}{6}$ $P(T_{ADB} \le 90') = \frac{5}{6} \cdot \underbrace{0{,}8}_{P(\bar{I})} = \frac{4}{6} = \frac{2}{3}$

ES 1/3/99

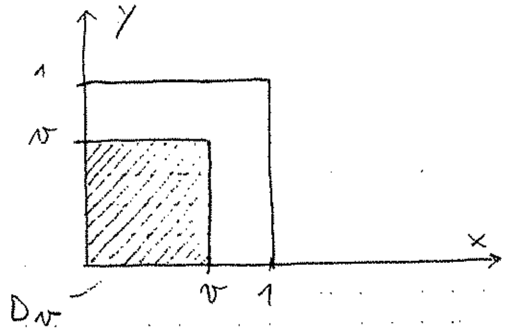
2.a



$V = \max(X,Y) \quad f_V(v) = ?$

$F_V(v) = P(V \le v) = P(\max(X,Y) \le v) = P((x,y) \in D_v)$

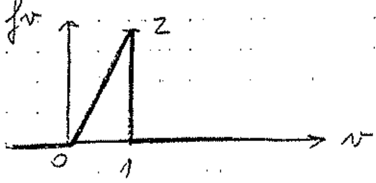
$D_v = \{(x,y) \mid \max(X,Y) \le v\} \qquad 0 < v < 1$



$F_V(V) = \frac{v^2}{1} = v^2$

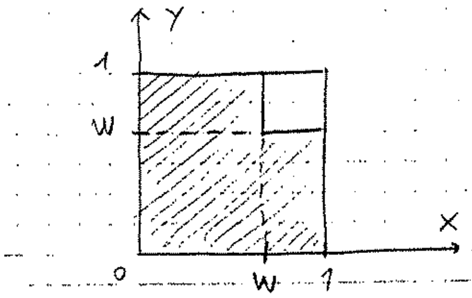
(1) → area del quadrato di partenza

$f_V(v) = \frac{dF_V(v)}{dv} = 2V \qquad 0 \le v \le 1$



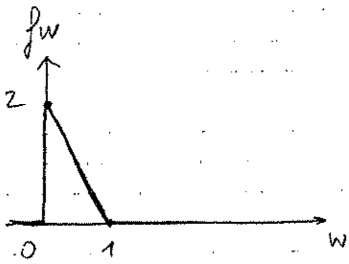
2.b $\quad W = \min(X,Y) \quad f_W(w) = ?$

$F_W = P(W \le w) = P(\min(X,Y) \le w) =$ perché $\min(X,Y) > W$ entrambe le coordinate devono stare sopra W



$= 1 - (1-W)^2 \quad 0 \le W \le 1$

$f_W(w) = \frac{dF_W(w)}{dw} = 2(1-W)$



2.c Modo difficile : $F_{VW}(v,w) = P(V \le v, W \le w)$ .....

Supp. di sapere che il $\max(X,Y) = 0 \Rightarrow \min(X,Y) = 0$ (δ nell'origine)

Se so che il MIN vale 1 anche il MAX vale 1 $\Rightarrow$ ci sono situazioni in cui le 2 variabili sono <u>fortemente</u> legate $\Rightarrow$ <u>NON sono indipendenti</u>

Se $V = 0 \Rightarrow W = 0$ $\left(\underline{f_{W|V}(w|v=0) = \delta(w) \neq f_W(w)} \right)$

ES 1 $\quad$ 17/9/97

1.b $\quad E[X] = 0 \quad V = Var[X]$ è singolare (X vettore)

Dimostrare che $\exists \alpha$ t.c. $P(\alpha^T X = 0) = 1$

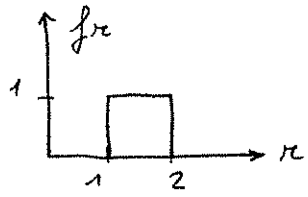
<u>V singolare $\Rightarrow \exists \alpha \neq 0$ t.c. $V\alpha = 0$</u>

$Y = \alpha^T X = \alpha_1 x_1 + \alpha_2 x_2 \dots + \alpha_n x_n$

$Var[Y] = \alpha^T Var[X] \alpha = \alpha^T \underbrace{V\alpha}_{=0} = 0 \Rightarrow P(Y = E[Y]) = 1$ (per Cebicev)

$\Rightarrow P(Y = 0) = 1$

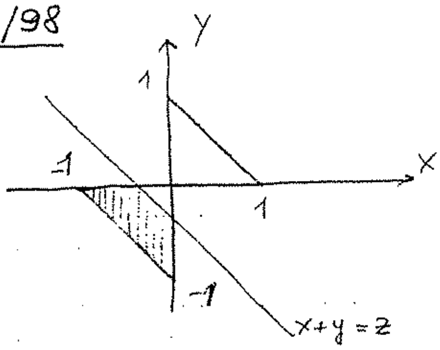
ES 2 del 17/9/97



$$V = \frac{4}{3}\pi r^3 \qquad E[V] = E\left[\frac{4}{3}\pi r^3\right] = \frac{4}{3}\pi \int_{-\infty}^{\infty} r^3 f_r(r)\,dr$$

$$= \frac{4}{3}\pi \int_1^2 r^3 dr = 5\pi$$

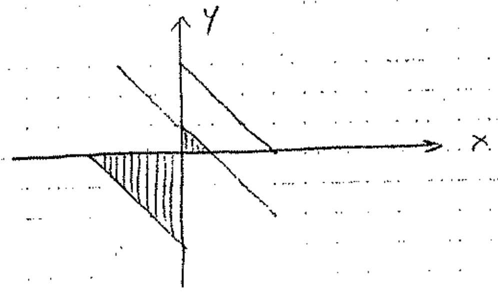
29/9/98



$Z = X+Y \qquad F_Z(z) = ?$ Area Totale = 1

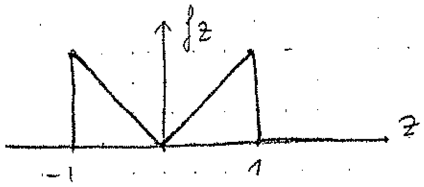
$$P(Z \le z) = P\left((X,Y) \in D_z\right) = P(X+Y \le z)$$

1° caso: intercetta negativa $\rightarrow F_Z = 0{,}5 - \frac{z^2}{2}$



$$F_Z(z) = \begin{cases} \frac{1}{2} - \frac{z^2}{2} & z \le 0 \text{ e } z \ge -1 \\ \frac{1}{2} + \frac{z^2}{2} & z \ge 0 \text{ e } z \le 1 \end{cases}$$

1.b $\quad f_Z(z) = \frac{dF_Z(z)}{dz} = \begin{cases} -z & -1 \le z \le 0 \\ z & 0 \le z \le 1 \end{cases}$



ES 15/9/98

p = probabilità di essere multato $\quad x_k$ = prob. di farla franca k giorni di seguito.

$p = 10^{-3}$

Per quali valori di k si ha $\begin{cases} x_k = 0{,}5 \\ x_k = 0{,}05 \end{cases}$

$x_k = (1-p)^k$ si ricava k per tentativi

13/9/99 $\quad X$ = Tempo per servire un cliente $\quad f_X(x) = \lambda e^{-\lambda x} \quad E[X] = \frac{1}{\lambda} = 5$ min

2.a $\quad P(X > 3) = ?$

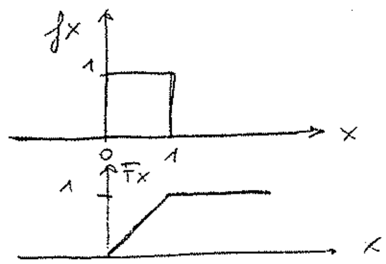
Per la proprietà di NON Memoria dell'exp non devo preoccuparmi di eventuali clienti che sono serviti.

$$P(X > 3) = 1 - F_X(3) = 1 - (1 - e^{-\lambda 3}) = e^{-3/5}$$

2.b $\quad P$(almeno 1 finisca in meno di 3 minuti) $= 1 - P$(Tutti e 2 impieghino + di 3 min).

$$= 1 - (e^{-3/5})^2 = 1 - e^{-\frac{6}{5}}$$

19/7/99 ES 2



$Y = X^2 \quad f_Y(y) = ?$

perché $X > 0$

$$F_Y(y) = P(Y \le y) = P(X^2 \le y) = P(\cancel{-\sqrt{y}} \le X \le \sqrt{y}) =$$

$$P(X \le \sqrt{y}) = F_X(\sqrt{y}) = \sqrt{y} \qquad 0 \le y \le 1$$

$F_X(x) = x \quad 0 < x < 1 \qquad f_Y(y) = \frac{1}{2\sqrt{y}} \quad 0 < y < 1$

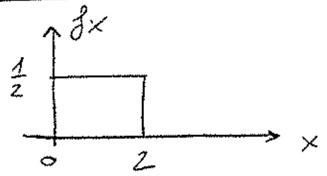
oppure $f_Y(y) = \frac{f_X(x_1)}{|g'(x_1)|} = \frac{1}{2\sqrt{y}}$

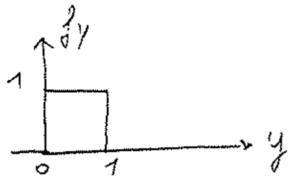
$y = g(x) = x^2$

$g'(x) = 2x \quad x_1 = \sqrt{y}$

$= 2\sqrt{y}$

3/11/2000 ES 2



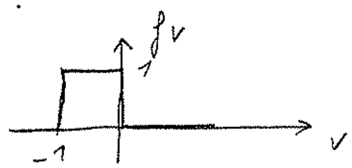


$Z = X - Y$

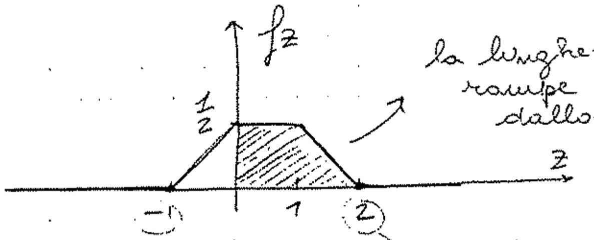
2.a $F_Z = ?$ $f_Z = ?$

2b $P(Z>0) = ?$

$V = -Y$

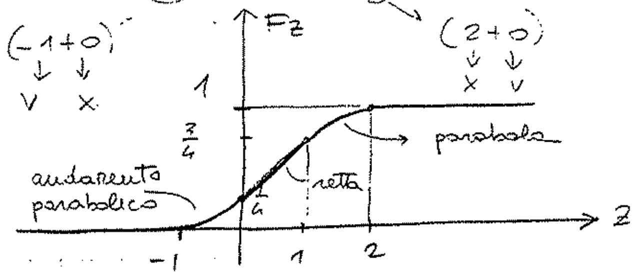


$Z = X + V$



la lunghezza delle rampe è det. dalla uniforme + corta

$P(Z>0) = 3/4$ (area ///)



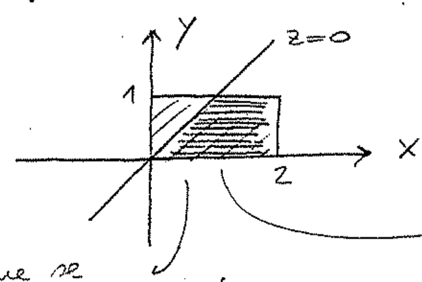
$$F_Z(z) = 0{,}5 \int_{-1}^{z} (\xi + 1)\, d\xi = \frac{1}{2}\left(\frac{\xi^2}{2} + \xi \Big|_{-1}^{z}\right)$$

$-1 \le z \le 0 \rightarrow$

$$= \frac{1}{2}\left(\frac{z^2}{2} + z - \frac{1}{2} + 1\right) = \frac{z^2}{4} + \frac{z}{2} + \frac{1}{4}.$$

oppure

$f_{XY}(xy) = f_X(x) f_Y(y) = \frac{1}{2} \quad 0 \le x \le 2 \quad 0 \le y \le 1$

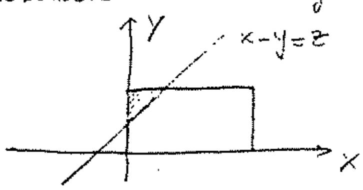


$z = 0 \quad x - y = 0$

$z > 0 \quad x - y > 0 \rightarrow y < x$

$P(Z>0) = \frac{3}{4}$ (area ≡)

è come se avessi preso un pto in modo equiprobabile nel rettangolo



$F_Z(z) = P(Z \le z) = P(X - Y \le z)$ ....

$\downarrow$

$Y \ge X - z$

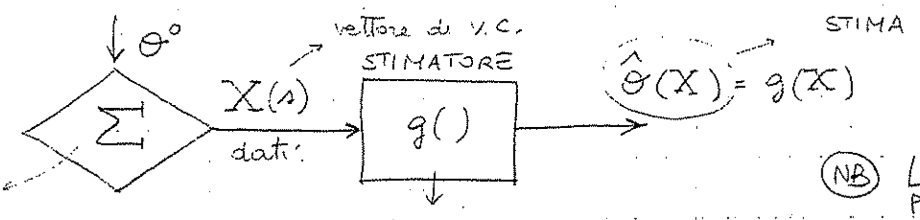
TEORIA DELLA STIMA

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



(NB) Lo stimatore è da PROGETTARE

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

funzione $\downarrow$ elaborazione dei dati

$\hat{\theta}(X)$ è una v.c.!

es $X = \begin{bmatrix} 1 \\ 1 \\ 0 \\ 1 \\ 0 \\ 0 \\ 1 \end{bmatrix}$ Testa o croce $\rightarrow$ $\qquad g() = \dfrac{\sum x_i}{N} \qquad \hat{\theta} = \dfrac{\sum x_i}{N}$ = 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
- $X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix} \qquad f_X^{(\theta^\circ)}(x) = \dfrac{1}{\sqrt{(2\pi)^N}(\theta_2^\circ)^N} \cdot e^{-\frac{1}{2}\sum_1^N \left(\frac{x_i - \theta_1^\circ}{\theta_2^\circ}\right)^2}$

la ddp è f.z. di θ°

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

Stimatore abbastanza buono perché obbedisce alla legge forte dei grandi numeri

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

lo vediamo 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 $\neq$!

MOMENTO CENTRALE CAMPIONARIO DI ORDINE K

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

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

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

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

OSSERVAZIONE: $\quad Z := \frac{M_1 - E[M_1]}{\sqrt{Var[M_1]}} = \frac{M_1 - m}{\sigma/\sqrt{N}} \Rightarrow M_1 = m + \frac{\sigma}{\sqrt{N}} \cdot Z \rightarrow$ 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 $\Rightarrow$ l'intensità del rumore dipende da σ e da N $\Rightarrow$ per N grande il rumore diminuisce $\Rightarrow$ $\to 0$ per $N \to \infty$

$\downarrow$

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. 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

$\downarrow$

$\exists$ un'intera famiglia di v.c. χ^2

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

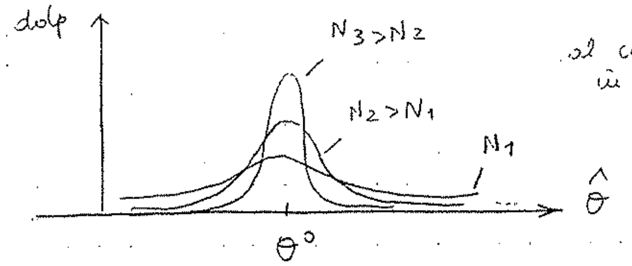
Cosa posso dire per M_2?

$Y := X^2 \Rightarrow M_2$ 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 θ° $\Rightarrow$ per $N \to \infty$ voglio essere certo (probabilità 1) di essere in θ°

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)

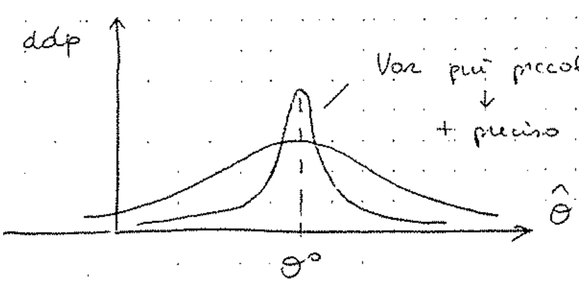
Considero uno stimatore non polarizzato ($E[\hat{\theta}] = \theta_0$) scalare. Allora, sotto condizioni di regolarità, vale

$\hookrightarrow$ ∃ una legge analoga nel caso polarizzato

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

per lo stimatore NON polarizzato la varianza è SINONIMO di PRECISIONE

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



$Var[\hat{\theta}]$ 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

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

è l'inverso della ...

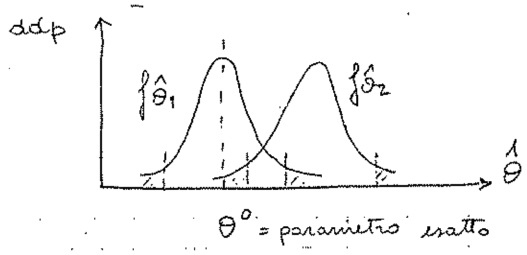
Esempi:

- 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

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



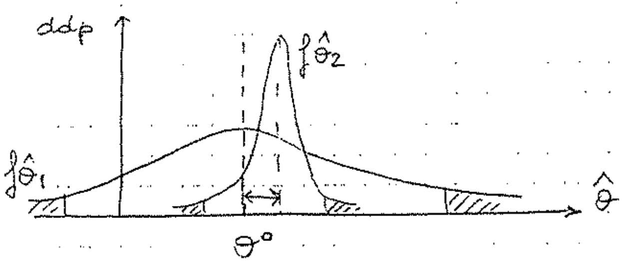
$f\hat{\theta}_1$ è NON polarizzato o CENTRATO perchè il valor medio cade sul parametro vero

$f\hat{\theta}_2$ 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 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, sottraendo, 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 ?

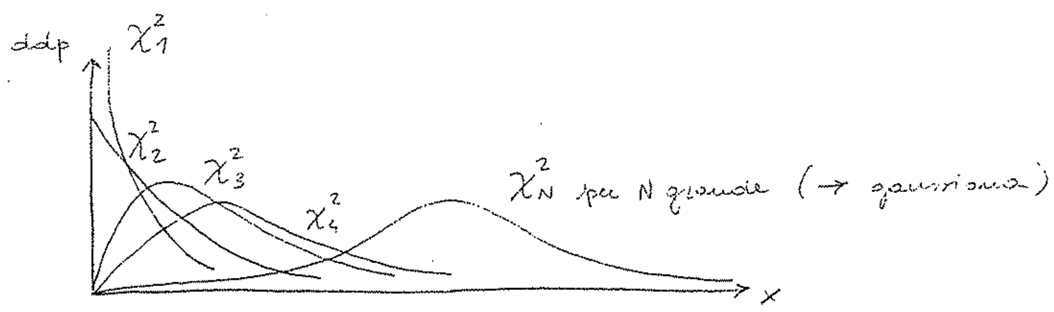
$$E[S^2_c] = E\left[\frac{N}{N-1} S^2\right] = \frac{N}{N-1} E[S^2] = \frac{\cancel{N}}{\cancel{N-1}} \cdot \frac{\cancel{N-1}}{\cancel{N}} \sigma^2 = \sigma^2 \Rightarrow \text{NON POLARIZZATO}$$

(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

[illegible]



Dovrebbe esserci una tabella del χ^2 $\forall$ 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) (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

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

se non conosco il valor medio questo m uso la stima M_1

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

VARIANZA CAMPIONARIA

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

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

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

$$\Rightarrow E[S^2] = E[M_2] - E[M_1^2] = m_2 - (Var[M_1] + E[M_1]^2) =$$

$$= m_2 - \frac{\sigma^2}{N} - m^2 = (m_2 - m^2) - \frac{\sigma^2}{N} = \sigma^2 - \frac{\sigma^2}{N} \qquad \text{cvd}$$

NB la varianza della media campionaria $\neq$ media della varianza campionaria

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

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

N-1 e NON N gradi di libertà perché i tanti addendi condividono M_1 $\Rightarrow$ perdono 1 g.d.l.

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

convergenza meno robusta della media campionaria $\Rightarrow$ N deve essere grande

PROPRIETÀ DEGLI STIMATORI

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

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

- <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 NON ha senso dare un limite inferiore alla varianza</u>

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 $\boxed{E[(\hat{\theta} - \theta^\circ)^2]}$ (<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).

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

MEAN SQUARE ERROR

o

ERRORE QUADRATICO MEDIO

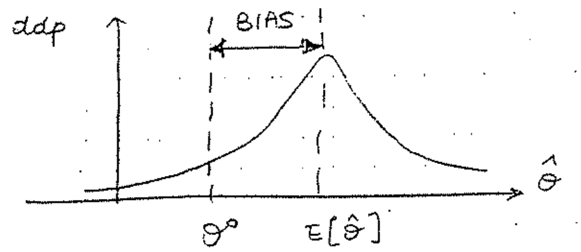
Si noti che $E[(\hat{\theta} - \theta^\circ)^2] = E[((\hat{\theta} - E[\hat{\theta}]) + (E[\hat{\theta}] - \theta^\circ))^2] =$

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

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

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

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



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

$Var[\hat{\theta}^{mv}] \leq Var[\hat{\theta}]$, $\forall \hat{\theta}$ non polarizzato

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

<u>NON</u> 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 [illegible] ui gaussiane (sono tutte =)

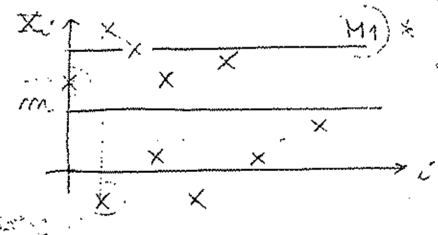
$\boxed{\text{NOTA}}$ $f_x(x) = f_{x_1}(x_1) \cdot f_{x_2}(x_2) \cdots f_{x_N}(x_N) = \prod_{i=1}^{N} f(x_i)$

3-5-2001

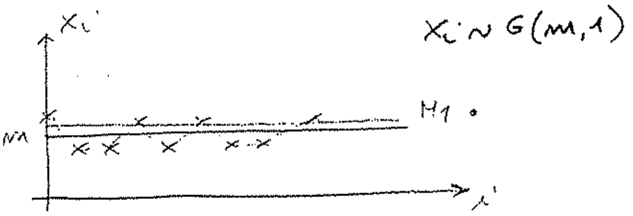
INTERVALLI DI CONFIDENZA

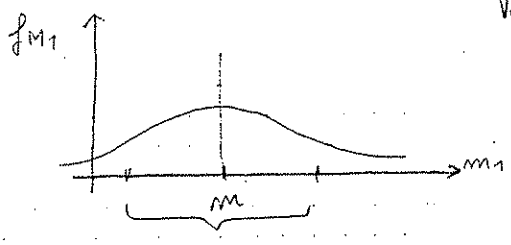
Fornire delle stime senza indicarne la precisione può essere del tutto fuorviante.

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

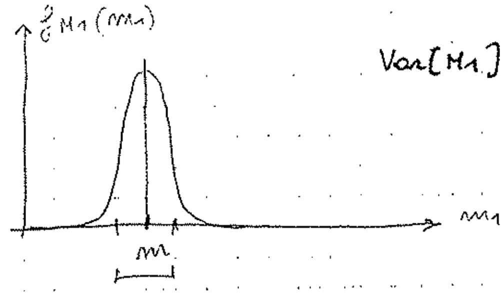


$X_i \sim G(m,100)$ ← distr. gaussiana





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



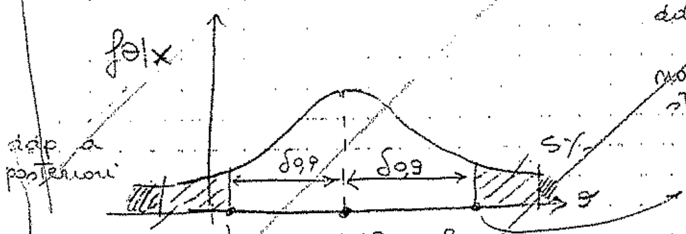
$Var[M_1] = \frac{1}{N}$

* 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

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

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

$\theta^{MAP} = \theta^B$ → 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'al

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)

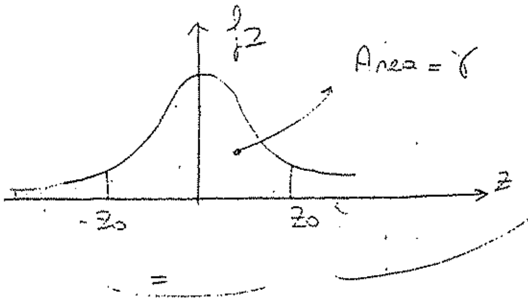
Come fare se non uso la stima a posteriori?

__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_γ per γ fissato

$\hat{\Theta}$ una v.c. Supp ddp gaussiana Supp valore attuale = valore vero

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

Cerco nelle tabelle z_0 t.c. $P(-z_0 \le Z_1 \le z_0) = \gamma \Rightarrow P(|Z| \le z_0) = \gamma$



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

$$= F_Z(z_0) - \left(1 - F_Z(z_0)\right) =$$

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

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

$$\Rightarrow P\left(\left|\frac{\hat{\theta}-\theta}{\sigma_{\hat{\theta}}}\right| \le z_0\right) = \gamma \Rightarrow P\left(\underbrace{|\hat{\theta}-\theta|}_{\text{modulo dell'errore} \Rightarrow 1/2 \text{ forchetta}} \le z_0 \sigma_{\hat{\theta}}\right) = \gamma \Rightarrow$$

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

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

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%.

le Bayesiano dice che tutto ciò è un imbroglio perchè $\hat{\theta}$ NON è una V.C. ma è θ di natura
→ se raccolgo dei dati diventa una V.C.

IMBROGLIO: dire che $\theta \in I_\gamma$ perchè 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: γ 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, $Var[X_i] = \sigma^2$, σ^2 NOTA, N "grande". Trovare I_γ per M_1. Parametro da cercare $\theta = m$

TEO. CENTRALE del LIMITE

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

sono soddisfatte tutte le HP del caso precedente

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

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

$$\boxed{\text{SE: } \frac{\sigma}{\sqrt{N}} \text{ è anche detto STANDARD ERROR}}$$

↓ dev. std della media

Possibili PROBLEMI:

(a) noti $N = 100$ e $\gamma = 0,9$, trovare I_γ

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

(c) noti $\gamma = 0,95$ e $I_\gamma = [M_1 - \delta, M_1 + \delta]$ Trovare N (es: Istituto Demoscopico)

↳ Quale N serve x essere sicuro con prob γ di cadere dentro I_γ

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

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

(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 forcella (δ) dobbiamo quadruplicare i dati (N)!

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

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

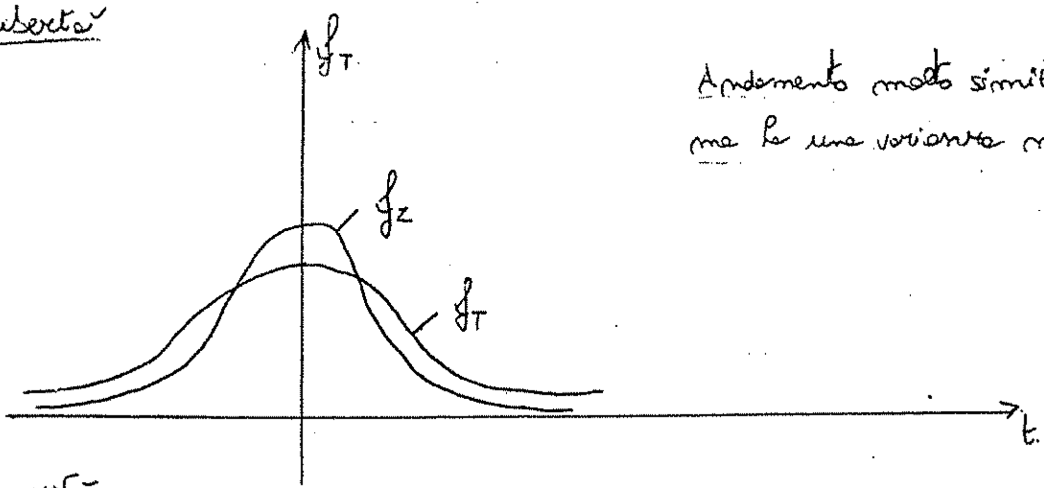
ESEMPIO: X_i i.i.d. gaussiane, σ^2 non nota.

Trovare I_γ 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 $T_N \xrightarrow[\text{distribuzione}]{\text{in}} G(0,1)$ (non per il teorema centrale del limite)

FINE PARENTESI

la varianza è σ^2/N

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

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

effettuando una pseudonormalizzazione

per il Teo di FISHER

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

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

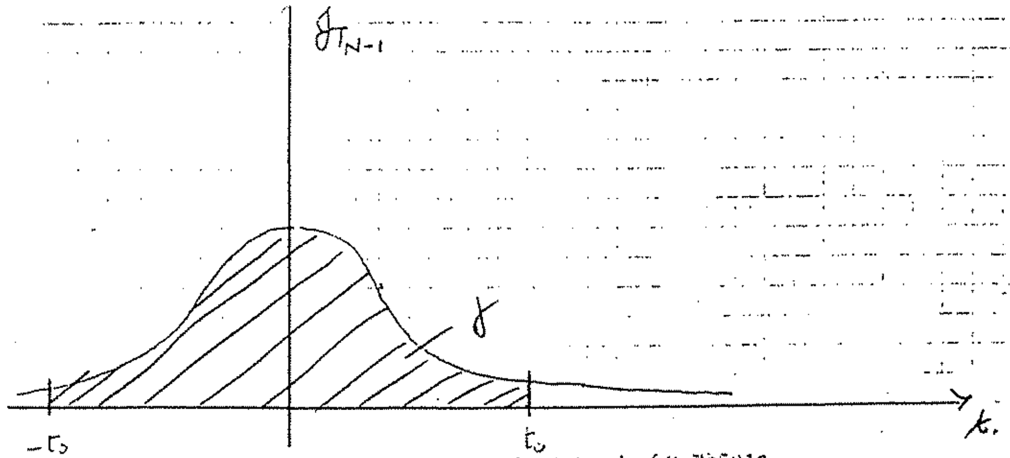
NB $S_c^2 = \frac{N}{N-1} S$ è la varianza campionaria corretta

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



) $P(-t_0 \le T_{N-1} \le t_0) = \gamma$ ← LIVELLO DI CONFIDENZA

t_0: ricavato dalle tabelle della t di Student.

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

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

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

NB

$\left(I_\gamma = \left[M_1 - z_0 \frac{\sigma}{\sqrt{N}} \, ; \, M_1 + z_0 \frac{\sigma}{\sqrt{N}}\right]\right.$, 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 ....

Interv. te: 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%)

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

n° di successi in N prove

$$E[S_N] = Np$$

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

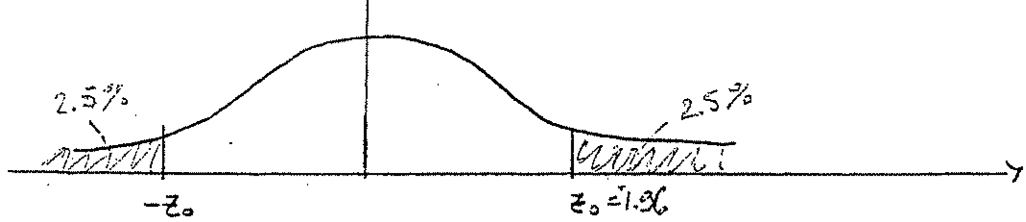
$$\hat{p} := \frac{S_N}{N} \quad \Rightarrow \quad E[\hat{p}] = p \qquad \text{(stimatore non polarizzato)}$$

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

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

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

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

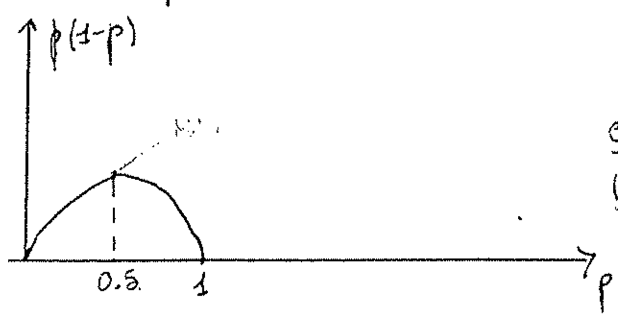


Per N grande, nel 95% dei casi:

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

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

OSSERVAZIONE



Caso peggiore (intervallo più largo) 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 $|p-\hat{p}| \leq \frac{1}{\sqrt{N}}$ ← quanto deve essere numeroso il campione

↳ la precisione costa

Per avere un errore $< 1\%$ $\Rightarrow N \geq 10'000$
(Nel 5% dei casi può ancora andar male)

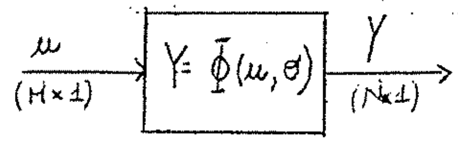
$\gamma = 0.95$	errore %
N=155	± 8%
N=200	± 7%
N=625	± 4%
N=2500	± 2%
N=10'000	± 1%

- 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).

1 Maggio 2001

non dinamici



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

3 insiemi di variabili u, Y, θ.

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

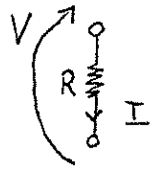
u: variabili indipendenti; variabili misurate oppure note a priori (cause) mediante le quali voglio spiegare gli effetti.

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

θ: parametri incogniti del modello

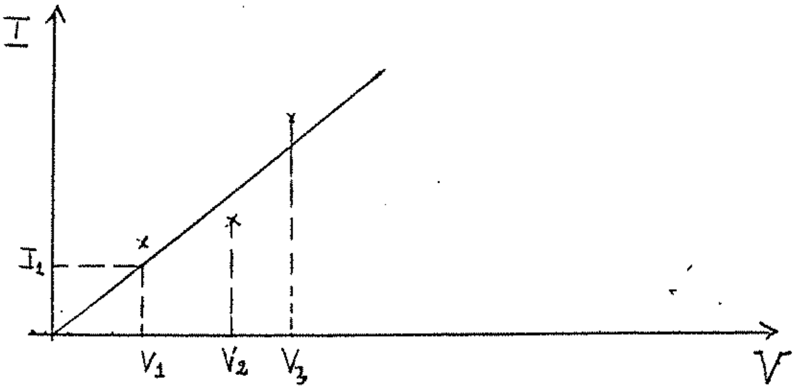
ESEMPIO:

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



$R = ?$

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



× punti reali.

La retta non si verifica nella realtà:
- errori di misura
- la relaz. tra V e I è lineare solo in una resistenza ideale

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

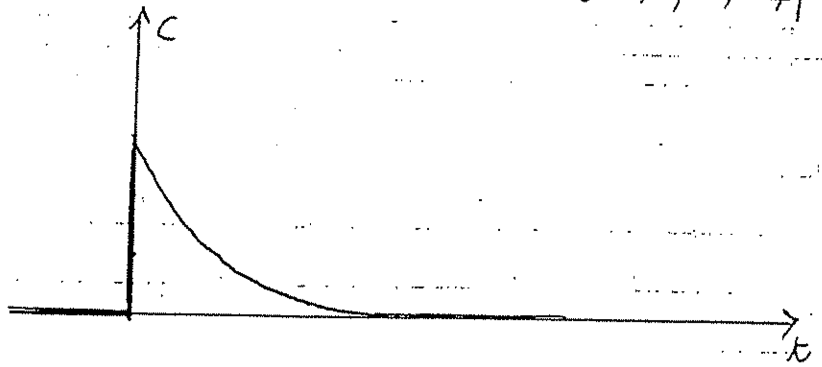
Cause
$$u = \begin{bmatrix} V_1 \\ V_2 \\ \vdots \\ V_N \end{bmatrix}$$

Parametro incognito $\theta = \frac{1}{R}$ $\Rightarrow$ $\underbrace{\Phi(u,\theta) = u\theta}_{\text{modello lineare nel parametro } \theta}$

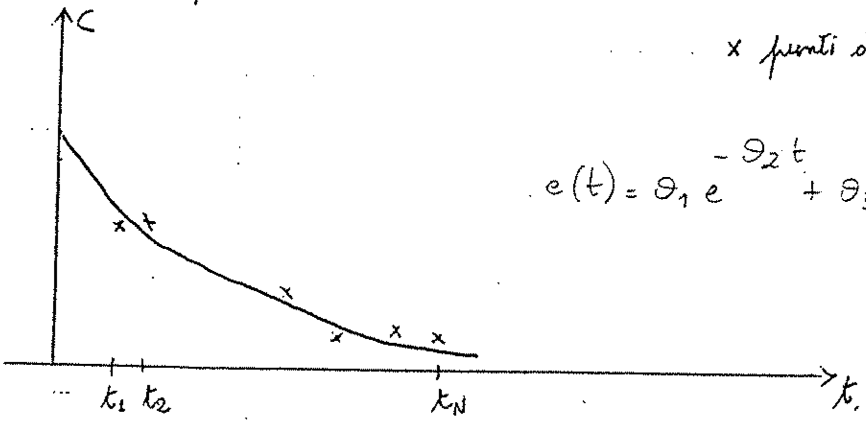
Oppure R $\Rightarrow$ $\Phi(u,\theta) = \frac{1}{\theta} u$ → modello non lineare nel parametro θ

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

$$c(t)=a e^{-\alpha t}+b e^{-\beta t}$$ con a, b, α, β da stimare.

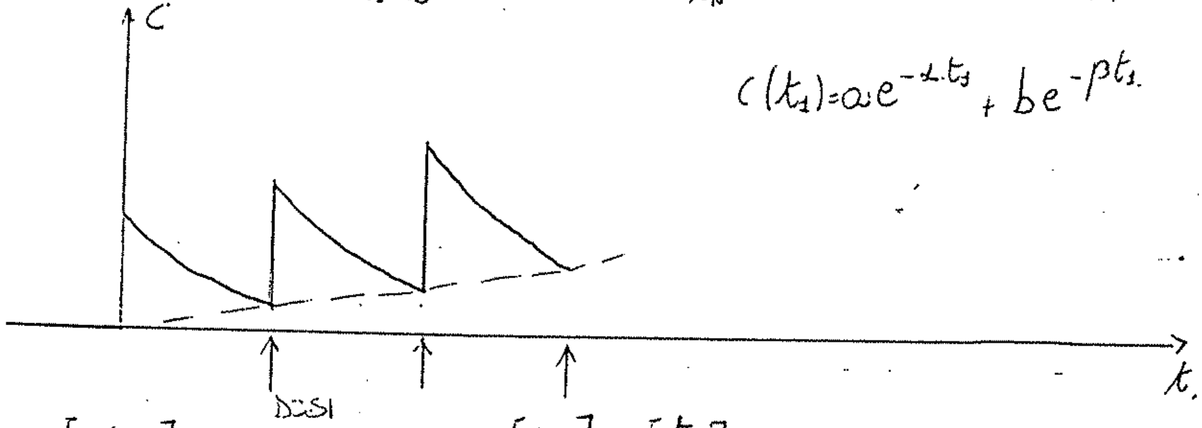


Effettuo dei prelievi ai tempi $t_1, \dots, t_N$



× punti ottenuti sperimentalmente

$$c(t)=\theta_1 e^{-\theta_2 t}+\theta_3 e^{-\theta_4 t}$$



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

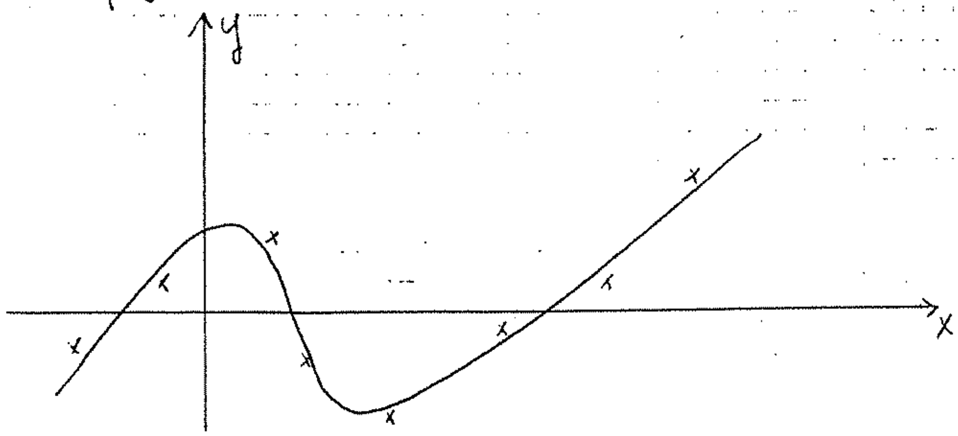
$$Y=\begin{bmatrix} c(t_1) \\ c(t_2) \\ \vdots \\ c(t_N) \end{bmatrix} \qquad u=\begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_N \end{bmatrix}=\begin{bmatrix} t_1 \\ t_2 \\ \vdots \\ t_N \end{bmatrix} \qquad \theta=\begin{array}{l} \left[\begin{matrix} a \\ \alpha \\ b \\ \beta \end{matrix}\right] \begin{matrix} \to \theta_1 \\ \to \theta_2 \\ \to \theta_3 \\ \to \theta_4 \end{matrix} \end{array}$$

$$\phi(u,\theta)=\begin{bmatrix} \theta_1 e^{-\theta_2 u_1}+\theta_3 e^{-\theta_4 u_1} \\ \vdots \\ \theta_1 e^{-\theta_2 u_N}+\theta_3 e^{-\theta_4 u_N} \end{bmatrix}$$

modello non lineare nei parametri
(θ_1 e θ_3 compaiono linearmente mentre θ_2 e θ_4 sono all'esponente)

ESEMPIO : Voglio interpolare (approssimare) delle coppie di punti (X_i, Y_i) con una cubica

$$\left(y = a_0 + a_1 X + a_2 X^2 + a_3 X^3 \right)$$



$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix} \qquad u = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_N \end{bmatrix} = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_N \end{bmatrix} \qquad \theta = \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \end{bmatrix} = \begin{bmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \\ \theta_4 \end{bmatrix}$$

$\phi(u, \theta) =$

$$y_1 = a_0 + a_1 X_1 + a_2 X_1^2 + a_3 X_1^3$$

$$\phi(u, \theta) = \begin{bmatrix} \theta_1 + \theta_2 u_1 + \theta_3 u_1^2 + \theta_4 u_1^3 \\ \vdots \\ \theta_1 + \theta_2 u_N + \theta_3 u_N^2 + \theta_4 u_N^3 \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & u_1 & u_1^2 & u_1^3 \\ & & & \\ 1 & u_N & u_N^2 & u_N^3 \end{bmatrix}}_{\Phi(u)} \underbrace{\begin{bmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \\ \theta_4 \end{bmatrix}}_{\theta}$$

modello lineare nei parametri

ESEMPIO : regressione lineare

Disponendo delle misure di $Q+1$ variabili $y(t), u_1(t), \ldots, u_q(t)$ per $t = 1, \ldots, N$

voglio spiegare Y mediante una funzione lineare delle u_j

$$Y(t) = \theta_1 u_1 + \theta_2 u_2 + \ldots \theta_q u_q$$

Regressione lineare della variabile y sulle variabili u_j

$$Y = \begin{bmatrix} y(1) \\ y(2) \\ \vdots \\ y(N) \end{bmatrix} \qquad u = \begin{bmatrix} u_1(1) \\ u_1(2) \\ \vdots \\ u_1(N) \\ u_2(1) \\ \vdots \\ u_2(N) \\ \vdots \\ u_q(1) \\ \vdots \\ u_q(N) \end{bmatrix} \qquad \theta = \begin{bmatrix} \theta_1 \\ \vdots \\ \theta_q \end{bmatrix}$$

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

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

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

$\Phi = \Phi(u)$ è detta matrice di sensitività

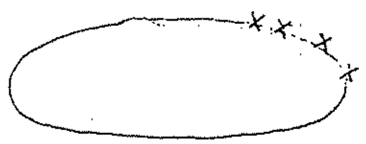
Minimi quadrati (LEAST SQUARES (LS))

Cenni storici:

LS nel 1735.

Gauss (1777 - 1855)

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)

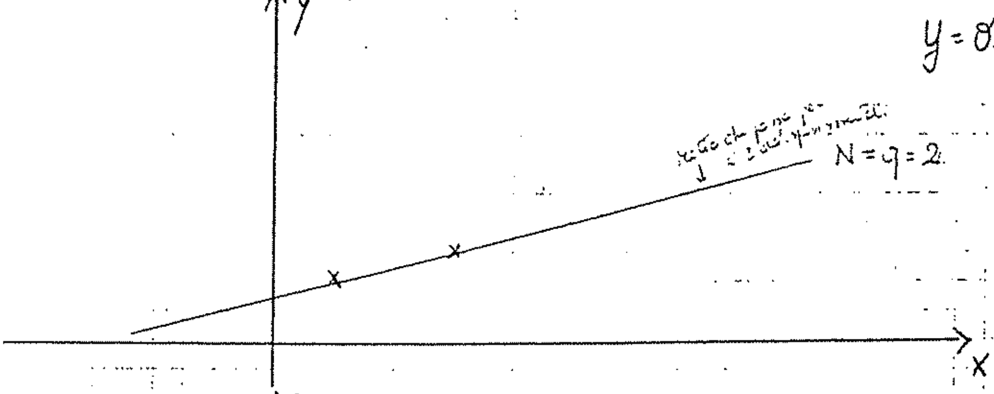
Fine dei cenni storici.

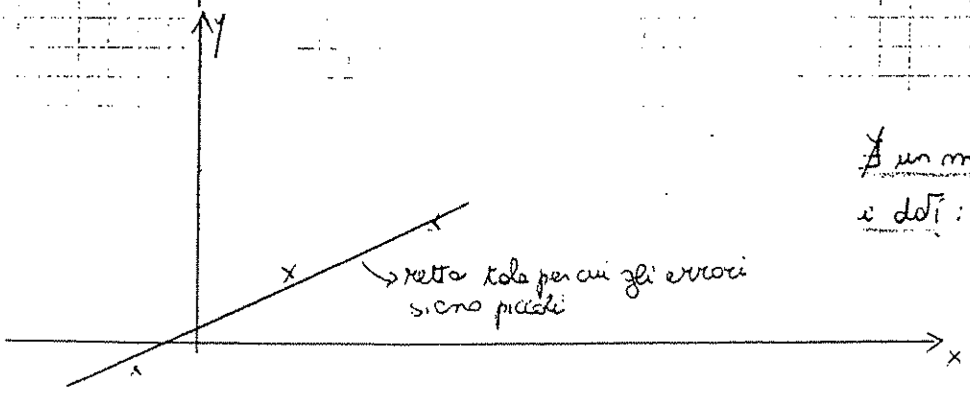
D'ora in poi: $\boxed{q \leq N}$ ← n° delle osservazioni disponibili

n° dei parametri del modello

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

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





$N = 3 > q = 2$

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

Allora, mi accontento che l'errore

$\boxed{\varepsilon := Y - \Phi\theta}$ sia piccolo $\qquad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_N \end{bmatrix}$

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

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

(n° di colonne di Φ linearmente indipendenti è uguale a q)

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

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

Dim:

PARENTESI: Funzioni matriciali

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

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

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

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

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

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

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

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

$$\left[\frac{d^2 f(x)}{dx^2}\right]_{ij} := \left[\frac{\partial^2 f}{\partial x_i \partial x_j}\right] \qquad f(x): \mathbb{R}^n \to \mathbb{R}^1$$

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

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

Sviluppo di Taylor

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

Def: D é detto <u>NADE</u> se $D = A^T A$

Proprietà $\boxed{D = D^T \geq 0}$

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

10 Maggio 2001

STIMA LS.

$$\theta^{LS} = \arg\min_{\theta} (Y-\Phi\theta)^T(Y-\Phi\theta)$$

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

Dim: $\operatorname{rank}(\Phi) = q \Rightarrow \det(\Phi^T\Phi) \neq 0$.

Infatti, se per assurdo fosse $\det(\Phi^T\Phi)=0 \Rightarrow \exists\, x \neq 0$ t.c. $\Phi^T\Phi x = 0 \Rightarrow$ $x^T\Phi^T\Phi x = 0 \Rightarrow ((\Phi x)^T\Phi x) = \|\Phi x\|^2 \Rightarrow \Phi x = 0 \Rightarrow \underset{N\times q}{\Phi}$ ha colonne linearmente dipendenti

$\operatorname{rank}(\Phi) < q$ (assurdo) l'inverso esiste.

Calcolo il gradiente di $J(\theta) = (Y-\Phi\theta)^T(Y-\Phi\theta)$

$$\frac{dJ(\theta)}{d\theta} = -2(Y-\Phi\theta)^T\Phi$$

$$\frac{dJ(\theta)}{d\theta} = 0 \qquad (Y^T - \theta^T\Phi^T)\Phi = 0 \Rightarrow Y^T\Phi = \theta^T\Phi^T\Phi \Rightarrow \underbrace{\boxed{\Phi^T\Phi\theta = \Phi^T Y}}_{\text{equaz. normali}}$$

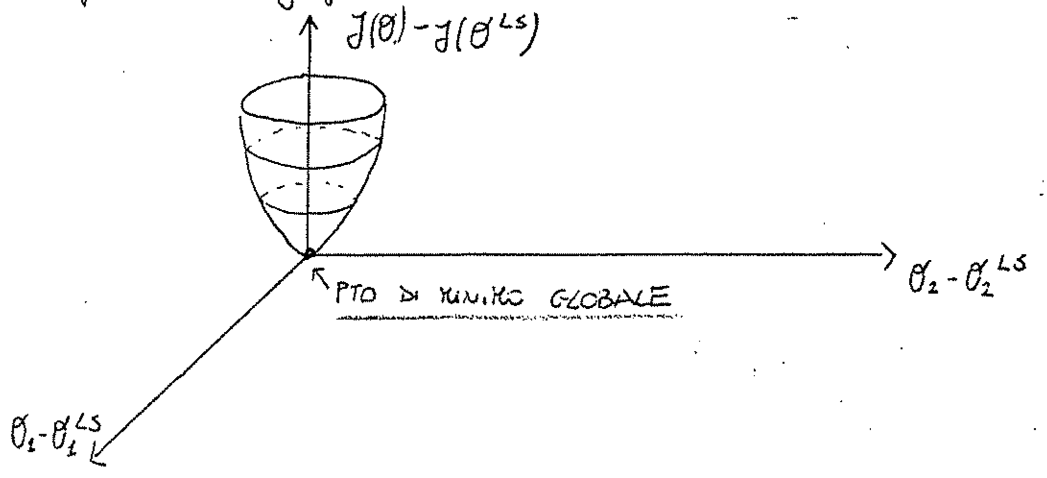
$$\boxed{\theta = (\Phi^T\Phi)^{-1}\Phi^T Y}$$

Hessiano $\qquad \dfrac{d^2J(\theta)}{d\theta^2} = 2\Phi^T\Phi \geq 0$ DIADE semidefinita positiva.

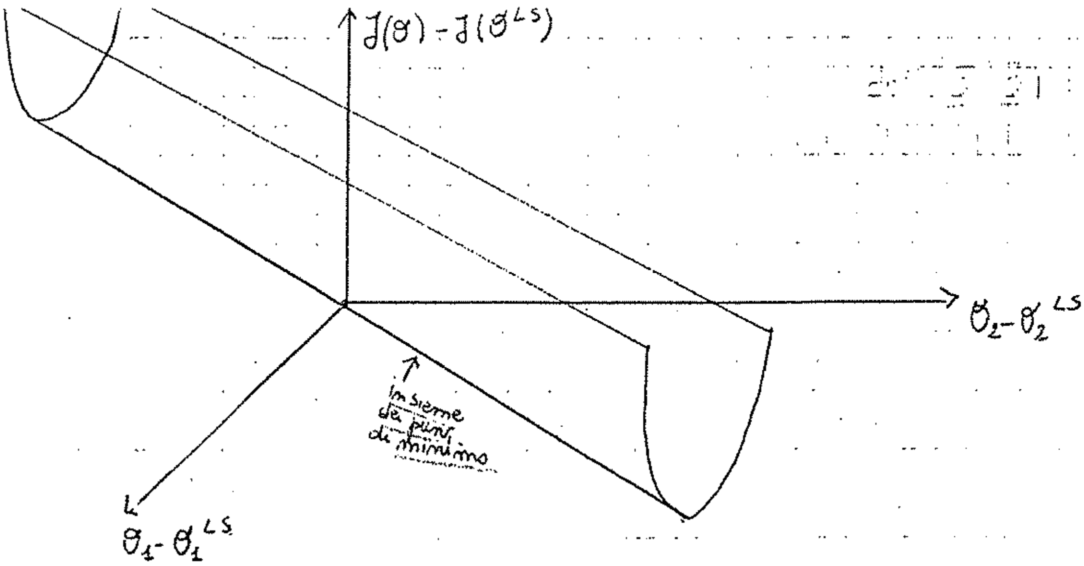
Proprietà $(A \geq 0 \text{ e } \det(A) \neq 0) \Rightarrow A > 0$.

Dato che $\det(\Phi^T\Phi) \neq 0 \Rightarrow \Phi^T\Phi > 0$ Ho un punto di minimo

Interpretazione grafica

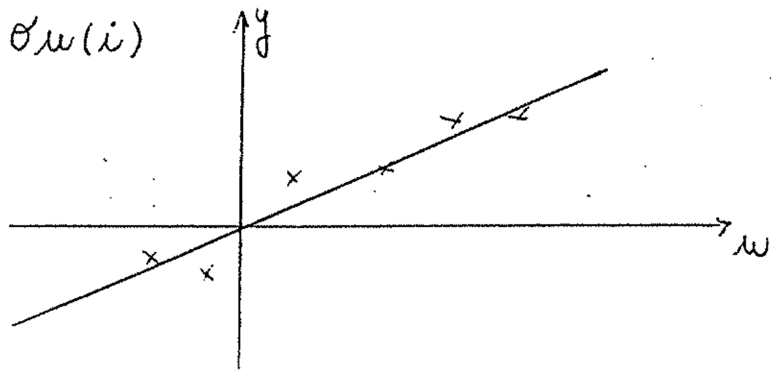


$\det(\Phi^T\Phi) \neq 0$ $\exists$ un'unica soluzione



Esempio : Regressione lineare con $q = 1$.

$$y(i) = \vartheta u(i)$$



$Y = \Phi \vartheta$

$y(1) = u(1)\vartheta$

$y(2) = u(2)\vartheta$

$$Y = \begin{pmatrix} y(1) \\ y(2) \\ \vdots \\ y(N) \end{pmatrix} \qquad \Phi = \begin{pmatrix} u(1) \\ u(2) \\ \vdots \\ u(N) \end{pmatrix}$$

$$\vartheta^{LS} = (\Phi^T \Phi)^{-1} \Phi^T Y = \frac{\sum_{i=1}^{N} u(i) y(i)}{\sum_{i=1}^{N} u(i)^2}$$

Il problema dell'identificabilità

Cosa succede se $\text{rank}(\Phi) < q$?

Per semplicità consideriamo $q = 3$ e un pb di regressione lineare.

$$y(t) = \vartheta_1 u_1(t) + \vartheta_2 u_2(t) + \vartheta_3 u_3(t)$$

$$\Phi = \begin{bmatrix} u_1(1) & u_2(1) & u_3(1) \\ u_1(2) & u_2(2) & u_3(2) \\ \vdots & \vdots & \vdots \\ u_1(N) & u_2(N) & u_3(N) \end{bmatrix}$$

Suppongo che $\text{rank}(\Phi) = 2$. $\Rightarrow$ una delle colonne di Φ è combinazione lineare delle altre, ovvero esistono α, β t.c. $u_3(t) = \alpha u_1(t) + \beta u_2(t)$

(5)

cioè $\quad y(t) = \theta_1 u_1(t) + \theta_2 u_2(t) + \theta_3 u_3(t)$

$= (\theta_1 + \alpha\theta_3) u_1(t) + (\theta_2 + \beta\theta_3) u_2(t)$

$\Rightarrow u_3$ è inutile perché posso ottenere la stessa previsione y usando solo $u_1(t)$, u

Considero un nuovo vettore dei parametri incogniti:

$$\bar{\theta} = \begin{bmatrix} \bar{\theta}_1 \\ \bar{\theta}_2 \end{bmatrix} = \begin{bmatrix} \theta_1 + \alpha\theta_3 \\ \theta_2 + \beta\theta_3 \end{bmatrix}$$

Considero la regressione di Y su u_1 e u_2. Se la condizione di identificabilità è so

Ta trovo un'unica soluzione

$$\bar{\theta}^{LS} = \begin{bmatrix} \bar{\theta}_1^{LS} \\ \bar{\theta}_2^{LS} \end{bmatrix}$$

Se tengo $u_3 \Rightarrow \infty$ soluz.

$\theta^{LS} = [\theta_1^{LS} \ \theta_2^{LS} \ \theta_3^{LS}]^T$ che soddisfo

$$\begin{cases} \bar{\theta}_1^{LS} = \theta_1^{LS} + \alpha\theta_3^{LS} \\ \bar{\theta}_2^{LS} = \theta_2^{LS} + \beta\theta_3^{LS} \end{cases}$$

DI CONDIZIONAMENTO: è un indice del grado di singolarità di una matrice.
Matrice che tende alla singolarità
↳ determinante della matrice che tende a zero

LIMITI DELLA STIMA LS

- affidabilità della stima
- il modello è giusto?
- confronto tra modelli

⇓ per dare delle risposte

Teoria della stima

$\boxed{\text{STIMA ML}}$

Devo fare ipotesi sulla d.d.p dell'errore di misura $V := Y - \Phi\theta$ $\quad V = \begin{bmatrix} V_1 \\ V_2 \\ \vdots \\ V_N \end{bmatrix}$

Ipotesi I 1

$$Y = \Phi\theta + V, \quad V \sim N(0, \Sigma_V), \ \Sigma_V > 0$$

Se gli errori sono indipendenti

$$\Sigma_V = \begin{bmatrix} \sigma_1^2 & & 0 \\ & \sigma_2^2 & \\ 0 & & \sigma_N^2 \end{bmatrix}$$

se inoltre hanno tutti la stessa varianza $\Sigma_V = \sigma^2 I$

Teorema Se vale l'ipotesi I 1 e $\text{rank}(\Phi) = q$, allora

a) $\theta^{ML} = \arg\min_\theta J^{ML}(\theta)$, $\boxed{J^{ML}(\theta) = \varepsilon^T \Sigma_V^{-1} \varepsilon}$ → SOMMA DEI QUADRATI DEI RESIDUI (SSR) PESATI

b) $\theta^{ML} = (\Phi^T \Sigma_V^{-1} \Phi)^{-1} \Phi^T \Sigma_V^{-1} Y$ ← v.c. gaussiana

($\Rightarrow \theta^{ML}$ è gaussiana)

c) $E[\theta^{ML}] = \theta$ (stimatore non polarizzato)

d) $\underline{Var\left[\theta^{ML}\right] = \left(\Phi^T \Sigma_V^{-1} \Phi\right)^{-1}}$

Dim. . . . . .

Commento

Ipotesi forti (V gaussiano)

Si verifica che $Var\left[\theta^{ML}\right] = S^{-1}$

(S: matrice di informazione di Fisher) → si è raggiunto il limite di Cramer-Rao

⇒ θ^{ML} è lo stimatore a minima varianza

→ Relazione con L.S.

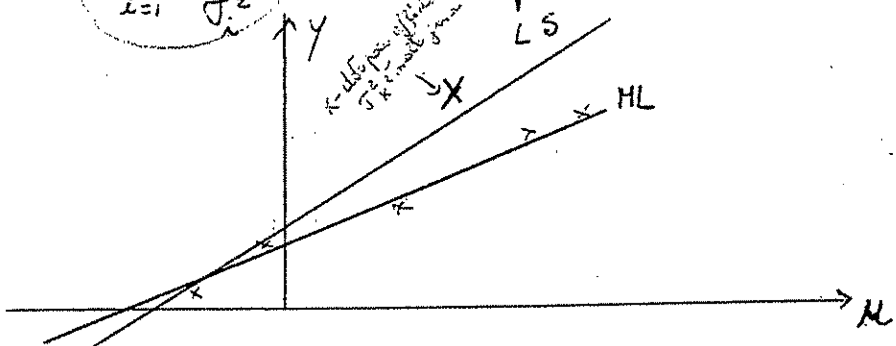
Se $\Sigma_V = \sigma^2 I \Rightarrow \theta^{ML} = \theta^{LS}$

$$\left(\theta^{ML} = \left(\frac{\Phi^T\Phi}{\cancel{\sigma^2}}\right)^{-1} \frac{\Phi^T}{\cancel{\sigma^2}} Y = \theta^{LS}\right)$$

Se $\Sigma_V = \begin{bmatrix} \sigma_1^2 & & 0 \\ & \sigma_2^2 & \\ 0 & \ddots & \sigma_N^2 \end{bmatrix}$

$$J^{ML}(\theta) = \varepsilon^T \Sigma_V^{-1} \varepsilon = [\varepsilon_1 \varepsilon_2 \dots \varepsilon_N] \begin{bmatrix} \frac{1}{\sigma_1^2} & & 0 \\ & \ddots & \\ 0 & & \frac{1}{\sigma_N^2} \end{bmatrix} \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_N \end{bmatrix} =$$

$$= \sum_{i=1}^{N} \frac{\varepsilon_i^2}{\sigma_i^2}$$ (Somma pesata dei quadrati degli errori residui) ←



Vogliamo trovare gli intervalli di confidenza

STIMA DELLA VARIANZA DEL DISTURBO

Spesso Σ_V è noto a meno di un coefficiente moltiplicativo ($\Sigma_V = \sigma^2 I$)

Teorema Vale I1 con $\Sigma_V = \sigma^2 \Psi$ con Ψ matrice nota e σ^2 scalare incognito

Allora

(i) $\theta^{ML} = \left(\Phi^T \Psi^{-1} \Phi\right)^{-1} \Phi^T \Psi^{-1} Y$

(ii) $(\sigma^2)^{ML} = \dfrac{\left(Y - \Phi\theta^{ML}\right)^T \Psi^{-1} \left(Y - \Phi\theta^{ML}\right)}{N} = \dfrac{J_\Psi^{ML}(\theta^{ML})}{N}$

$$J_\psi^{ML} = \varepsilon^T(\theta)\,\psi^{-1}\,\varepsilon(\theta)$$

NOTA: $(\sigma^2)^{ML}$ è polarizzato

Per avere una stima non polarizzata

$$\hat{\sigma}^2 = \frac{J_\psi^{ML}(\theta^{ML})}{N-q}$$

STIMA ML: INTERVALLI DI CONFIDENZA

(a) Σ_V nota

è congiunta gaussiana

$$\theta^{ML} \sim N(\theta, \Sigma_{\theta^{ML}}), \quad \Sigma_{\theta^{ML}} = (\Phi^T \Sigma_V^{-1} \Phi)^{-1}$$

anche le marginali sono gaussiane

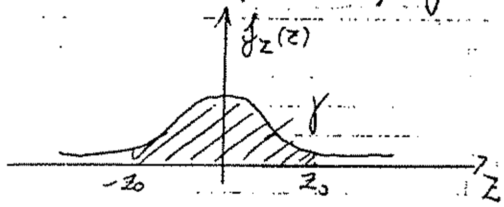
$\Downarrow$

intervalli di confidenza per uno stimatore gaussiano con varianza nota.

$$I_\gamma(\theta_i) = \left[\theta_i^{ML} - z_0\,\sigma_{\theta_i^{ML}},\ \theta_i^{ML} + z_0\,\sigma_{\theta_i^{ML}}\right]$$

$$\sigma^2_{\theta_i^{ML}} = \left[\Sigma_{\theta^{ML}}\right]_{ii}$$

z_0 è t.c. $P(|Z| \le z_0) = \gamma$



11 maggio 2001

(b) $\Sigma_V = \sigma^2 \psi$ (ψ matrice nota, σ^2 scalare incognito)

Stimatore

$$\hat{\Sigma}_{\theta^{ML}} = \hat{\sigma}^2 (\Phi^T \psi^{-1} \Phi)^{-1}, \quad \hat{\sigma}^2_{\theta_i^{ML}} = \left[\hat{\Sigma}_{\theta^{ML}}\right]_{ii}$$

Proprietà: Risulta che

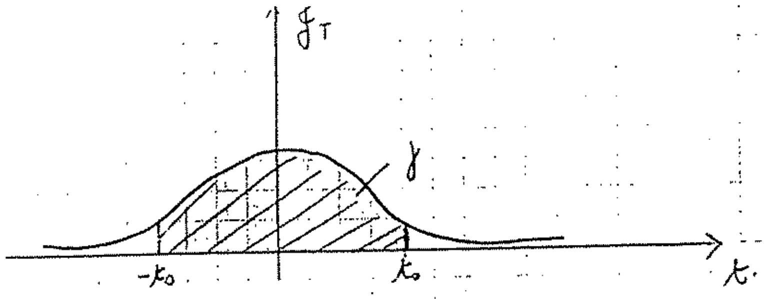
$$\frac{\theta_i^{ML} - \theta_i}{\hat{\sigma}_{\theta_i^{ML}}} \sim T_{N-q} \quad (t \text{ di Student})$$

N: n° dei dati

q: n° dei parametri

$$I_\gamma(\theta_i) = \left[\theta_i^{ML} - t_0\,\hat{\sigma}_{\theta_i^{ML}},\ \theta_i^{ML} + t_0\,\hat{\sigma}_{\theta_i^{ML}}\right]$$

t_0 è t.c. $P(|T_{N-q}| \le t_0) = \gamma$



APPLICAZIONE: IL PROBLEMA DELLA REGRESSIONE LINEARE

$$y(t) = \theta_1 u_1(t) + \theta_2 u_2(t) + \dots + \theta_q u_q(t) + v(t) \quad t=1,\dots,N$$

$$\mathrm{Var}[V] = \sigma^2 I \qquad V = \begin{bmatrix} v(1) \\ \vdots \\ v(N) \end{bmatrix}$$

Se inoltre V è gaussiano
$\Downarrow$

$$\theta^{ML} = \theta^{LS}$$

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

Ponendo $\varphi(t) = \underbrace{[u_1(t) \ u_2(t) \ \dots \ u_q(t)]^T}_{\text{vettore dei regressori}}$, si vede che $\theta^{LS} = (\Phi^T\Phi)^{-1}\Phi^T Y = \left[\sum_{t=1}^{N} \varphi(t)\varphi(t)^T\right]^{-1}\left[\sum_{t=1}^{N} \varphi(t)y(t)\right]$

(se $\mathrm{Rank}(\Phi) = q$ CONDIZ. DI IDENTIFICABILITÀ)

$$\hat{\sigma}^2 = \frac{J(\theta^{ML})}{N-q} = \frac{\sum_{t=1}^{N} \varepsilon(t)^2}{N-q} = \frac{\sum_{t=1}^{N} \left(y(t) - \varphi(t)^T\theta^{ML}\right)^2}{N-q}$$

$y(t) = \varphi(t)^T\theta + v(t)$

Stimo $\mathrm{Var}[\theta^{ML}]$ come $\hat{\Sigma}_{\theta^{ML}} = \hat{\sigma}^2 \left[\sum_{t=1}^{N} \varphi(t)\varphi(t)^T\right]^{-1}$

Spesso i risultati sono espressi come segue:

$$y(t) = \theta_1^{ML} u_1(t) + \theta_2^{ML} u_2(t) + \dots + \theta_q^{ML} u_q(t) + \varepsilon(t)$$

$\varepsilon(t)$: errore espresso coi parametri stimati

$\varepsilon(t) = y(t) - \varphi(t)^T\theta^{ML}$

$v(t) = y(t) - \varphi(t)^T\theta$

$$\left[\hat{\sigma}_{\theta_1^{ML}}\right] \qquad \left[\hat{\sigma}_{\theta_2^{ML}}\right] \qquad \dots \qquad \left[\hat{\sigma}_{\theta_q^{ML}}\right] \qquad \left[\hat{\sigma}\right]$$

Regole per capire quali regressori vanno tolti dal modello ($\exists$ variabili che non influenzano il mode

Problema inerente: come faccio ad essere sicuro che $\theta_K \neq 0$? (Se non sono sicuro può essere meglio togliere u_K dal modello)

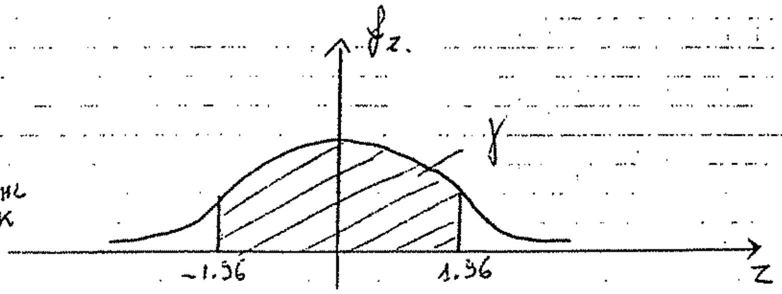
Procedo come "per assurdo", faccio l'ipotesi che $\theta_K = 0$ e controllo se i risultati sperimentali la smentiscono.

$$\theta_K = 0 \Rightarrow \frac{\theta_K^{ML} - E(\theta_K^{ML})}{\sqrt{Var[\theta_K^{ML}]}} \sim Z \Rightarrow \frac{\theta_K^{ML}}{\sigma_K^{ML}} \sim Z$$

$$\left(\frac{\theta_K^{ML}}{\hat{\sigma}_K^{ML}} \sim T_{N-q} \right)$$

Nel 95% dei casi $|Z| \leq 1.96$

$$\left| \frac{\theta_K^{ML}}{\sigma_{\theta_K^{ML}}} \right| \leq 1.96$$



Idea: Se $|\theta_K^{ML}| \leq 1.96\, \sigma_{\theta_K^{ML}}$

non c'è niente di assurdo ⇒ non riesco a smentire l'ipotesi $\theta_K = 0$

Invece se $|\theta_K^{ML}| > 1.96\, \sigma_{\theta_K^{ML}}$

mi trovo in una situazione che si verifica solo nel 5% dei casi (ipotizzando $\theta_K = 0$)

⇒ "ASSURDO" (poco probabile)

⇒ respingo l'ipotesi $\theta_K = 0$ ⇒ "il parametro θ_K è SIGNIFICATIVAMENTE $\neq 0$"

CONCLUSIONE

$\boxed{|\theta_K^{ML}| > 2\, \sigma_{\theta_K^{ML}}}$ sono "sicuro" che $\theta_K \neq 0$

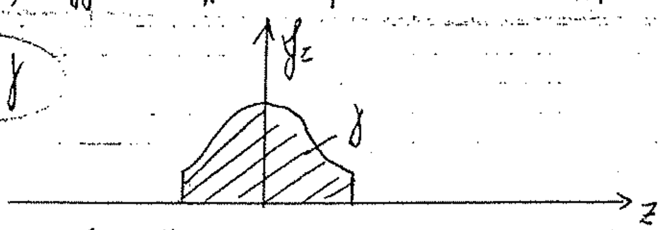
$\boxed{|\theta_K^{ML}| < 2\, \sigma_{\theta_K^{ML}}}$ non so garantire che $\theta_K \neq 0$

Esistono 2 tipi di errore:

F^+ (falso positivo) affermo $\theta_K \neq 0$ mentre in realtà $\theta_K = 0$

F^- (falso negativo) affermo $\theta_K = 0$ quando in realtà $\theta_K \neq 0$

$P(F^+) = 1 - \gamma$



Test di ipotesi (soggettivo: ognuno può scegliere un "suo" γ)

16 - 5 - 2001

pg 30 : Pressinizzazione : numero abbastanza dati

FPE e AIC hanno una prob. > 0 di sovrastimare il modello

⇒ è meglio usare + di un criterio x valutare un modello

APPLICAZIONE : REGRESSIONE LINEARE

$q = 1, 2$

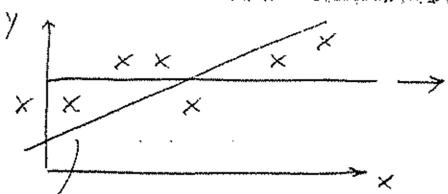
Dati : $y(1), y(2) \dots y(N)$ $\qquad$ Hp : $V \sim N(0, \sigma^2 I)$ $\qquad$ $V = \begin{bmatrix} v(1) \\ v(2) \\ \vdots \\ v(N) \end{bmatrix}$

$x(1), x(2) \dots x(N)$

2 modelli alternativi :

(a) $y(t) = \bar{\theta}_1 + v(t)$

(b) $y(t) = \theta_1 + \theta_2 x(t) + v(t)$



→ modello + semplice = cost (y NON dipende da x)

modello lineare (y è fz lineare di x)

Media campionaria

Stimiamo (a) :

$$Y = \begin{bmatrix} y(1) \\ \vdots \\ y(N) \end{bmatrix} \qquad \Phi = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} \qquad \bar{\theta}_1^{ML} = (\Phi^T \Phi)^{-1} \Phi^T Y = N^{-1} \sum_{i=1}^{N} y(i) = M_y$$

matrice Φ

manca una matrice che in questo caso è $= I$

$$(b) \qquad \Phi = \begin{bmatrix} 1 & x(1) \\ 1 & x(2) \\ \vdots & \vdots \\ 1 & x(N) \end{bmatrix} \qquad \theta^{ML} = \begin{bmatrix} N & \Sigma x(i) \\ \Sigma x(i) & \Sigma x(i)^2 \end{bmatrix}^{-1} \begin{bmatrix} \Sigma y(i) \\ \Sigma x(i) y(i) \end{bmatrix}$$

2 parametri θ_1^{ML} e θ_2^{ML}

$$\theta_1^{ML} = M_y - \theta_2^{ML} M_x$$

$$\theta_2^{ML} = \frac{S_{xy}}{S_{xx}} \quad \text{dove} \quad \boxed{S_{xy} = \frac{\sum_1^N (x(i) - M_x)(y(i) - M_y)}{N} \qquad S_{xx} = \frac{\sum_1^N (x(i) - M_x)^2}{N}}$$

è la COVARIANZA CAMPIONARIA

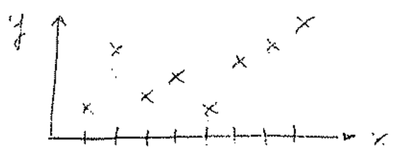
Valore di y predetto

$$\hat{y}(t) := \Phi \theta^{ML} = M_y + \frac{S_{xy}}{S_{xx}} (x(t) - M_x)$$

è come se avessimo sostituito ai valori teorici quelli campionari

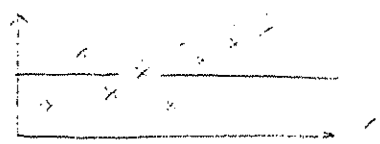
Ricorda $E[Y | X = x] = E[Y] + Cov[X, Y] \, Var[X]^{-1} (x - E[X])$

NB : Dal pto di vista concettuale ↑ immagino che X sia una V.C. Ma nella realtà ($\hat{y}(t)$) potrebbe NON esserlo (es: x potrebbero essere dei tempi → uniforme).



La formula di $\hat{y}(t)$ vale sempre, per VC e NON.

Dobbiamo decidere se il modello è (a) o (b)

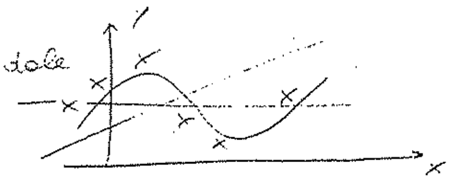


Devo decidere se θ_2 serve o meno: vd. Tab di FISHER

$$f := \frac{J(\bar{\theta}_1^{ML}) - J(\theta^{ML})}{J(\theta^{ML})/(N-2)} \sim F(1, N-2)$$

× fare il test ho bisogno Hp la gaussianità ⇒ ML

Rischioso: la dipendenza potrebbe essere sinusoidale ⇒ tra cost e retta scelgo la retta ma NON è quella vera



Introduciamo un indice normalizzato che misura in che grado il modello (b) è superiore ad (a)

Coeff. di Determinazione (Multipla): $R^2 := \dfrac{\bar{J}(\bar{\theta}_1^{LS}) - J(\theta^{LS})}{\bar{J}(\bar{\theta}_1^{LS})} = 1 - \dfrac{J(\theta^{LS})}{\bar{J}(\bar{\theta}_1^{LS})}$

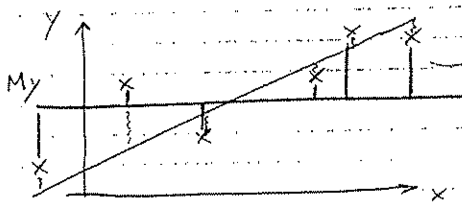
Scarti attorno alla retta di regressione

Scarti attorno a My

in questo caso LS = ML. Tanto non devo fare Hp iniziali

Sempre ≤ 1 (poiché sopra il modello è + complicato).

$$\boxed{0 \le R^2 \le 1}$$

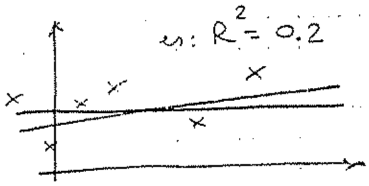


- scarti attorno a My
- = scarti attorno alla retta

- $R^2 \cong 0$ (scarti quasi sovrapponibili).

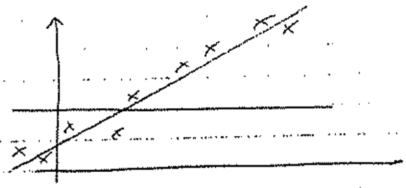
miglioro poco passando da (a) a (b).

es: $R^2 = 0.2$



- $R^2 \cong 1$

es. $R^2 = 0{,}91$



il modello (b) è molto meglio

B est
L inear
U nbiased
E stimator

Se rimuovo l'ipotesi di gaussianità, cosa posso dire?

IPOTESI I2: $Y = \Phi\theta + V$ → ∃ un modello vero

$E[V] = 0$ $\quad Var[V] = \sigma^2 \Psi$

(σ^2 scalare eventualmente incognito) $\quad \Box$

Teorema (Gauss & Markov)

Si consideri la cifra di merito $J^{(M)}(\theta) := \varepsilon^T \Psi^{-1} \varepsilon = (Y - \Phi\theta)^T \Psi^{-1} (Y - \Phi\theta)$ (MARKOV)

che è minimizzata da (*) $\theta^M := (\Phi^T \Psi^{-1} \Phi)^{-1} \Phi^T \Psi^{-1} Y$ (stimatore di MARKOV)

NON posso dire che è lo stimatore ML perché ho Hp I2 (NON ho presupposto la gaussianità).

Allora:

(i) θ^M è, tra tutti gli stimatori lineari e non polarizzati, lo stimatore che minimizza $Var[\hat{\theta} - \theta^o]$ (→ BLUE)

(*) stimatore lineare perché θ^M dipende linearmente da Y

→ NON è il miglior stimatore possibile (potrebbe essere NON lineare) ma è comodo e tra i lineari è il migliore.

(ii) $Var[\theta^M] = \sigma^2 (\Phi^T \Psi^{-1} \Phi)^{-1}$

OSSERVAZIONI

- $Var[V] = \sigma^2 I \Rightarrow \theta^M = \theta^{LS}$

- Si dimostra che (se σ^2 è incognito) $\hat{\sigma}^2 = \dfrac{J^M(\theta^M)}{N-q}$ è uno stimatore NON polarizzato di σ^2
- Usando σ^2 (o $\hat{\sigma}^2$) posso stimare anche $Var[\theta^M] = \Sigma_{\theta^M} = \hat{\sigma}^2(\Phi^T \psi^{-1} \Phi)^{-1}$
- ATTENZIONE! Dato che non conosco la ddp di V:
  - NON posso dire che $J^M(\theta^M) \sim \chi^2_{N-q}$ (Niente validazione del modello!)
  - NON posso calcolare gli intervalli di confidenza (visto non so se i param. sono gaussiani) (se non con Tchebycheff). poco preciso ⇒ poco usato (intervalli troppo grandi)

Idea: fornire solo le deviazioni standard (SD) dei parametri.

  - Cade il Test F
  - Posso usare FPE (~~anche basato sulla~~ Hp I2) → a volte si usa anche M

STIMA DI BAYES

Ho info. a priori su θ

Ipotesi I3: ∃ modello vero $Y = \Phi \Theta + V$, $V \sim N(0, \Sigma_V)$, $\Sigma_V > 0$ (def. pos.), $\Theta \sim N(m_\Theta, \Sigma_\Theta)$, $\Sigma_\Theta > 0$ e V e θ indipendenti

vettore di VC

TEOREMA: Sotto Hp I3:

(a) $\theta^B := \arg\min_\theta \left\{ \varepsilon^T \Sigma_V^{-1} \varepsilon + (\theta - m_\theta)^T \Sigma_\theta^{-1} (\theta - m_\theta) \right\}$ → somma dei quad. dei residui ($\Sigma_V =$ ...) pesata se $\Sigma_V \neq$ ...

(b) $\theta^B = (\Phi^T \Sigma_V^{-1} \Phi + \Sigma_\theta^{-1})^{-1} (\Phi^T \Sigma_V^{-1} Y + \Sigma_\theta^{-1} m_\theta)$ → valore che ho (es. risultati dei sondaggi telefonici)

(c) $Var[\theta^B] = [\Phi^T \Sigma_V^{-1} \Phi + \Sigma_\theta^{-1}]^{-1}$ → Matrice varianza che misura l'affidabilità dei dati (pochi dati ⇒ Σ_θ grande ⇒ Σ_θ^{-1} è piccola ⇒ i dati pesano poco).

COMMENTI

- $\Sigma_\theta^{-1} \to 0 \Rightarrow \theta^B \to \theta^{ML}$
  - matrice var. delle info a priori
  - se → 0 vuol dire che vale poco ⇒ i dati sperimentali valgono poco
- NON è più necessario ipotizzare $rank(\Phi) = q$ ($\Sigma_\theta > 0 \Rightarrow (\Phi^T \Sigma_V^{-1} \Phi + \Sigma_\theta^{-1}) > 0$,
  - la somma di una matrice semidef. pos. e 1 def. pos. è una matrice def. pos ⇒ Hp $rank(\Phi) = q$ NON serve x dire posso fare l'inversa
  - è comunque ≥ 0
  - se $\Sigma_\theta > 0$ anche $\Sigma_\theta^{-1} > 0$

Sfruttando le info a priori, posso avere $q > N$ (più incognite che dati)

sono comunque dati ("pesati") Φ rappresenta solo i dati nuovi

VARIABILI INDIPENDENTI AFFETTE DA ERRORE

Problema: relazione tra stature dei padri e dei figli.

$x(i)$ = statura i-esimo padre − statura media popolazione (→ non ...) E = 0

$y(i)$ = statura i-esimo figlio − statura media pop.

Idea: ∃ una relazione del tipo $y(t) = \theta x(t)$ → scostamento dell'altezza del padre dalla media; scostamento del figlio

↓ GALTON 1822 − 1911

Calcolo la regressione di y su x → $y(t) = \theta_x x(t) + \varepsilon_y(t)$ → residuo

Soluzione: $\theta_x^{LS} := \dfrac{\sum_{t=1}^N x(t) y(t)}{\sum_{t=1}^N x(t)^2}$ (Hp gaussiane ... θ^{ML})

NB ...

VALIDAZIONE TEST χ^2

Avendo a disposizione N dati $y_1, \dots y_N$ come faccio a sapere se il modello

$$Y = \Phi\theta^\circ + V \quad , \quad Var[V] = \sigma^2 I$$

li descrive adeguatamente?

Idea: $\theta^{LS} \cong \theta^\circ \Rightarrow \underline{\varepsilon = Y - \Phi\theta^{LS} \cong V}$ (il vettore dei residui è una "stima" del vettore degli errori di misura). Perciò mi aspetto che

$$\frac{1}{N}\sum_{i=1}^{N} \varepsilon_i^2 \cong \sigma^2$$

Se conosco σ^2 è bene controllare che σ^2 e la varianza campionaria dei residui abbiano lo stesso ordine di grandezza (idea ovvia: se ho errori di misura dell'ordine di 10^{-3} e i residui sono dell'ordine di 10^{-1}, il modello è sbagliato perché non riesce a spiegare i dati)

Ipotesi I1 $\quad Y = \Phi\theta^\circ + V \quad V \sim N(0, \sigma^2 I)$

Teorema: sotto I1, dividendo per σ^2 la somma dei quadrati dei residui della stima LS si ottiene una V.C. di tipo "χ^2 con N-q g.d.l."
(N n° dei dati $\quad$ q: n° de parametri)

$$\boxed{\frac{\varepsilon^T\varepsilon}{\sigma^2} \sim \chi^2(N-q)}$$

∃ TAB per χ^2

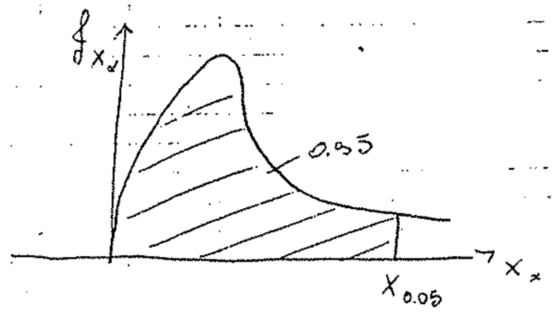
95% dei casi $\quad \frac{\varepsilon^T\varepsilon}{\sigma^2} < 14.07$

Se accade $\quad \frac{\varepsilon^T\varepsilon}{\sigma^2} > 14.07 \Rightarrow$ modello non buono

- Fissare un livello di signif. α
- Cercare su TAB il valore $X_\alpha \Rightarrow$

$$\frac{\varepsilon^T\varepsilon}{\sigma^2} < X_\alpha \Rightarrow \text{non respingo il modello}$$

$$\frac{\varepsilon^T\varepsilon}{\sigma^2} > X_\alpha \Rightarrow \text{respingo il modello}$$



Di sistema $\quad V \sim N(0, \Sigma_V)$

Basta usare $\quad \varepsilon^T \Sigma_V^{-1} \varepsilon$

Punti deboli: $\quad$ 1) può essere difficile capire quali sono i motivi per cui viene scartato il modello:

- $Y = \Phi\theta^{\circ} + V$ spiega male i dati
- V non è gaussiano
- Il valore di σ^2 è sbagliato per difetto
- Gli errori di misura non hanno tutti la stessa varianza

2) Il test si basa sull'ipotesi che ∃ un "modello vero" di tipo lineare che genera i dati e che gli errori sono gaussiani (ipotesi semplificative)

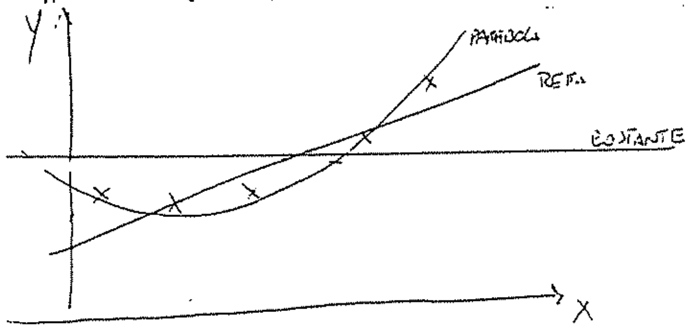
$\boxed{\text{TEST F}}$

Confronto tra modelli

MODELLI "MATRIOSKA" Una sequenza di classi di modelli in cui ciascuna classe comprende al suo interno (come casi particolari) le classi precedenti

Es. M_1 = [rette] M_2 = [parabole] M_3 = [cubiche] . . .

Pb N coppie (x_i, y_i), trovare una curva che approssimi al meglio i dati sperimentali



Idea stupida: considero i diversi modelli (rette, parabole, ...) e ne stimo i parametri con LS

Poi tra i modelli scelgo quello che minimizza SSR (: la somma dei quadrati dei residui)

FATTO : SSR decresce sempre al crescere dell'ordine del modello (dato che LS si basa sulla minimizzazione di SSR, non è possibile che la miglior parabola abbia una SSR di una retta)

⇓

Usare la minimizzazione di SSR conduce a scegliere sempre il modello più complesso (per esempio N=100 ⇒ polinomio di ordine 99) 100 punti con 100 parametri passa per tutti i punti

- Che male c'è nell'eccedere con il n° di parametri?
- Nessun pb se i dati fossero privi di rumore.

Es. Fitto con una parabola dati su una retta → il coeff del termine quadratico risulta = 0 ⇒ non commetto errori

- Se c'è rumore ed ho troppi parametri il modello tende ad essere influenzato dal rumore (riproduce oscillazioni che non hanno significato fisico ma che sono frutto degli errori di misura)

PRINCIPIO DI PARSIMONIA: non usare più parametri di quelli che servono

Idea: M_{k-1} e M_k, sappiamo che $SSR_k < SSR_{k-1}$ (somma dei quadrati dei residui)

Scegliamo M_k solo se SSR_k è "molto più piccolo" di SSR_{k-1}

Teorema Sotto l'ipotesi I1 $\left(Y = \Phi\theta^{\circ} + V \quad V \sim N(0, \sigma^2 I) \right)$

$$f = (N-K)\,\frac{SSR_{K-1} - SSR_K}{SSR_K} =$$ è una F di Fisher con $(1, N-K)$ g. di l.

dati · parametri

- Per $N-K$ grande $F(1, N-K) \underset{\text{tende}}{=} \chi^2(1)$

TAB F di Fisher.

Test. F — Fisso un livello di significatività α (ipotesi 0.05) cerco f_α t.c. $P(F(1, N-K) \ldots$ = 0.95

$f < f_\alpha$ — scelgo H_{K-1}

$f > f_\alpha$ — scelgo H_K

- Non è necessario conoscere σ^2
- è soggettivo (va scelto il livello di significatività)

LIMITI

$V \sim N(0, \sigma^2 \Psi)$ Ψ matrice nota $\Rightarrow \varepsilon^T \Psi^{-1} \varepsilon$ al posto di $SSR = \varepsilon^T \varepsilon$

- Presuppone l'esistenza di un modello vero
- Si applica solo a modelli matrioska.

16 Maggio 2001.

APPLICAZIONE: regressione lineare.. → è utile per fare delle previsioni

$q = 1, 2$

Dati $\quad y(1), y(2), \ldots, y(N)$

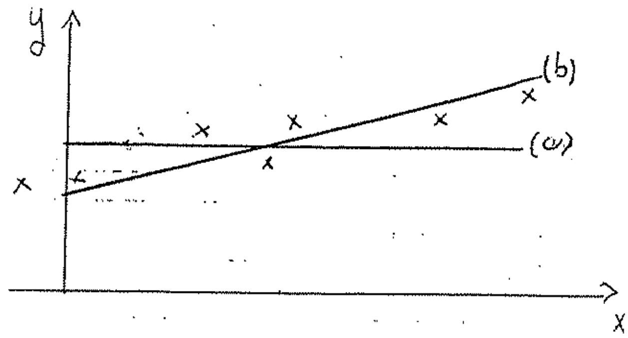
$\quad\quad x(1), x(2), \quad\quad , x(N)$

Ipotesi: $V \sim N(0, \sigma^2 I)$ $\quad V = \begin{bmatrix} v(1) \\ v(2) \\ \vdots \\ v(N) \end{bmatrix}$

2 modelli alternativi

(a) $\quad y(t) = \bar{\theta}_1 + v(t)$

(b) $\quad y(t) = \theta_1 + \theta_2 x(t) + v(t)$



(a) $\quad Y = \begin{bmatrix} y(1) \\ y(2) \\ \vdots \\ y(N) \end{bmatrix} \quad \bar{\Phi} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}$

Se i dati avessero un andamento sinusoidale andrebbe meglio il modello della costante

$$\bar{\theta}_1^{ML} = (\bar{\Phi}^T \bar{\Phi})^{-1} \bar{\Phi}^T Y = N^{-1} \sum_{i=1}^{N} y(i) = M_y \quad \text{media campionaria}$$

(b) $\quad \bar{\Phi} = \begin{bmatrix} 1 & x(1) \\ 1 & x(2) \\ \vdots & \vdots \\ 1 & x(N) \end{bmatrix} \quad \theta^{ML} = \begin{bmatrix} N & \sum x(i) \\ \sum x(i) & \sum x(i)^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum y(i) \\ \sum x(i) y(i) \end{bmatrix}$

$\theta_1^{ML} = M_y - \theta_2^{ML} M_x$

$\theta_2^{ML} = S_{xy} / S_{xx}$

$$S_{xy} = \frac{\sum_{i=1}^{N} [(x(i) - M_x)(y(i) - M_y)]}{N} \quad \text{covarianza campionaria}$$

$$S_{xx} = \frac{\sum_{i=1}^{N} (x(i) - M_x)^2}{N} \quad \text{varianza campionaria}$$

in questo caso x non è più funzione di t, poiché rappresenta il tempo

$$\hat{y}(t) = \bar{\Phi}\, \theta^{ML} = M_y + \frac{S_{xy}}{S_{xx}} (x(t) - M_x)$$

Apparentemente ex. prima da la statistica e dopo campionaria c'è comunque una differenza concettuale tra le due.

$$E[Y/X = x] = E[Y] + Cov(X, Y)\, Var^{-1}(X)\, (X - E[X])$$

← Formula del valore atteso condizionato per la gaussiana

in questo caso X v.c

devo decidere se θ_2 serve o meno

TEST F

$$f = \frac{\overbrace{J(\bar{\theta}_1^{ML}) - J(\theta^{ML})}^{SSR}}{J(\theta^{ML}) / (N-2)} \sim \overset{\text{F di Fisher}}{F}(1, N-2)$$

$f > f_\alpha$: prendiamo la retta

$f < f_\alpha$: prendiamo la costante

Va indicato il livello di significatività scelto per il Test F

Introduciamo un indice normalizzato che misura in che grado il modello (b) è superiore al modello (a).

scarti attorno alla retta di regressione

Nelle nostre ipotesi $ML = LS$

Si può usare anche senza ipotesi probabilistiche →

⇒ metto LS per evidenziare ciò

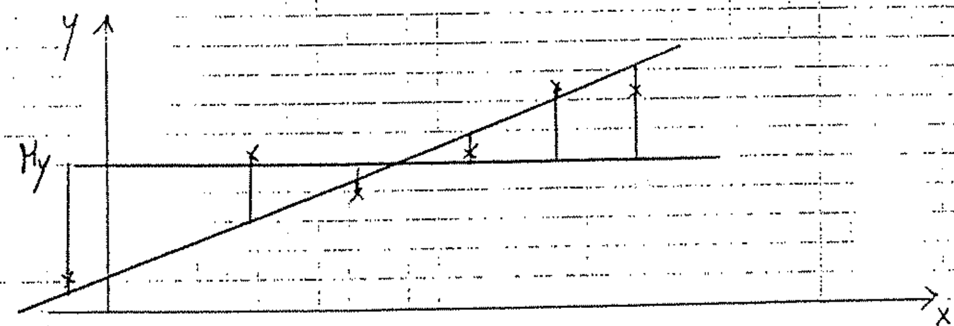
$$R^2 := \frac{\bar{J}(\bar{\theta}_1^{LS}) - J(\theta^{LS})}{\bar{J}(\bar{\theta}_1^{LS})} = 1 - \frac{J(\theta^{LS})}{\bar{J}(\bar{\theta}_1^{LS})}$$

(coeff. di determinazione multiplo)

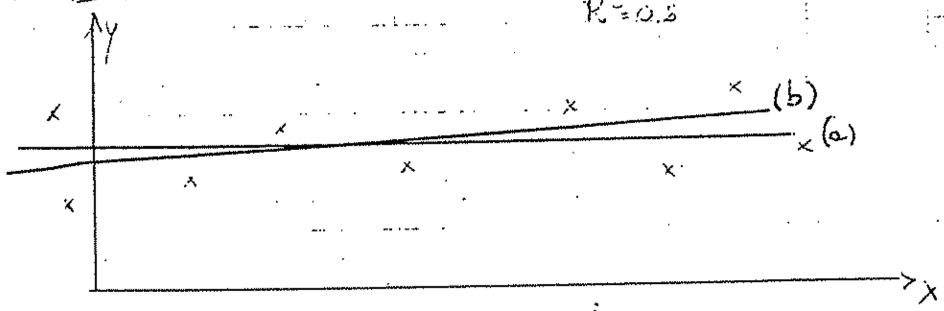
$\bar{J}(\bar{\theta}_1^{LS})$ ← scarti attorno a M_y

$\frac{J(\theta^{LS})}{\bar{J}(\bar{\theta}_1^{LS})} \leq 1$

$0 \leq R \leq 1$

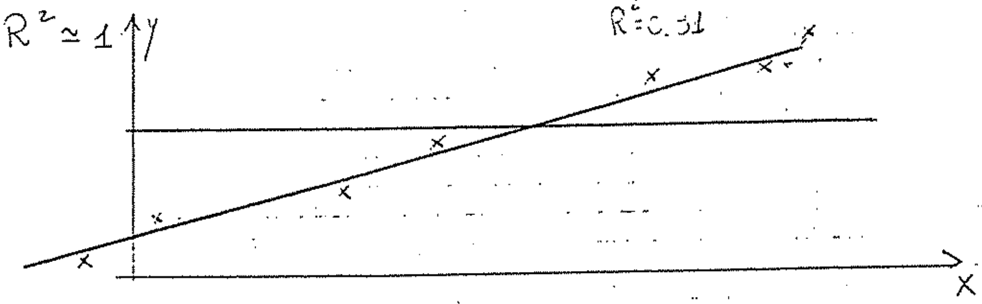


Se $R^2 \simeq 0$



Migliora poco passando da (a) a (b)

Se $R^2 \simeq 1$



il modello (b) è molto meglio

Best
Linear
Unbiased
Estimator

Se rimuovo l'ipotesi di gaussianità, cosa posso dire?

IPOTESI I2): $Y = \Phi\theta + V$ (∃ un modello vero) $E[V] = 0$ $Var[V] = \sigma^2\Psi$ (σ^2 scalare eventualmente incognito

<u>**Teorema (Gauss & Markov)**</u>

Si consideri la cifra di merito

$$J^M(\theta) = \varepsilon^T \Psi^{-1} \varepsilon = (Y - \Phi\theta)^T \Psi^{-1} (Y - \Phi\theta)$$

che è minimizzata da

$$\theta^M = (\Phi^T \Psi^{-1} \Phi)^{-1} \Phi^T \Psi^{-1} Y \quad \text{[illegible]}$$

Allora:

(i) θ^M è tra tutti gli stimatori lineari e non polarizzati, lo stimatore che minimizza $Var[\hat{\theta}-\theta^o]$ (BLUE)

→ il vettore dei parametri dipende linearmente da Y.

(ii) $Var[\theta^M] = \sigma^2 (\Phi^T \Psi^{-1} \Phi)^{-1}$

Osservazioni

- $Var[V] = \sigma^2 I \Rightarrow \theta^M = \theta^{LS}$
- Si dimostra che (se σ^2 è incognito)

$$\hat{\sigma}^2 = \frac{J^M(\theta^M)}{N-q}$$ è uno stimatore non polarizzato di σ^2

- Usando σ^2 (o $\hat{\sigma}^2$) posso stimare anche $Var[\theta^M] = \Sigma_{\theta^M} = \hat{\sigma}^2 (\Phi^T \Psi^{-1} \Phi)^{-1}$
- ATTENZIONE! Dato che non conosco la d d p di $V \Rightarrow$
  - non posso più dire che $J^M(\theta^M) \sim \chi^2_{N-q}$ (niente più validazione del modello)
  - non posso calcolare gli intervalli di confidenza (se non con Tchebycheff, che però dà intervalli troppo conservativi)

Idea: fornire solo le SD dei parametri
- Cade il Test F
- Posso usare FPE

STIMA DI BAYES

→ informazioni a priori su θ

Ipotesi I3 : $Y = \Phi \underset{\substack{| \\ \text{parametro: vettore di V.C}}}{\Theta} + V \qquad V \sim N(0, \Sigma_V), \ \Sigma_V > 0, \ \Theta \sim N(m_\theta, \Sigma_\Theta)$

$\Sigma_\Theta > 0 \qquad V, \Theta$ indip.

Teorema Sotto I3

a) $\theta^B := \arg\min_\theta \{ \overbrace{\varepsilon^T \Sigma_V^{-1} \varepsilon}^{\text{somma dei quadrati dei residui opportunamente pesati}} + (\theta - \underset{\uparrow \text{ informazioni a priori}}{m_\Theta})^T \Sigma_\Theta^{-1} (\theta - m_\Theta) \}$

misura l'affidabilità dei dati di cui dispongo a priori

b) $\theta^B = (\Phi^T \Sigma_V^{-1} \Phi + \Sigma_\Theta^{-1})^{-1} (\Phi^T \Sigma_V^{-1} Y + \Sigma_\Theta^{-1} m_\Theta)$

c) $Var[\theta^B] = [\Phi^T \Sigma_V^{-1} \Phi + \Sigma_\Theta^{-1}]^{-1}$

Commenti:

- $\Sigma_\Theta^{-1} \to 0 \Rightarrow \theta^B \to \theta^{ML}$

L'informazione a priori vale molto poco

↖ si basa interamente sui dati sperimentali

- Non è più necessario ipotizzare $rank(\Phi) = q$ $\quad (\Sigma_\Theta > 0 \Rightarrow (\Phi^T \Sigma_V^{-1} \Phi + \Sigma_\Theta^{-1} > 0))$

Sfruttando le informazioni a priori posso avere $q > N$ (più incognite dei dati)

Introduciamo un indice normalizzato che misura in che grado il modello (b) è superiore al modello (a). *nella nostra ipotesi* $HL = LS$

Si può usare anche senza ipotesi probabilistiche

⇒ metto LS per evidenziare ciò

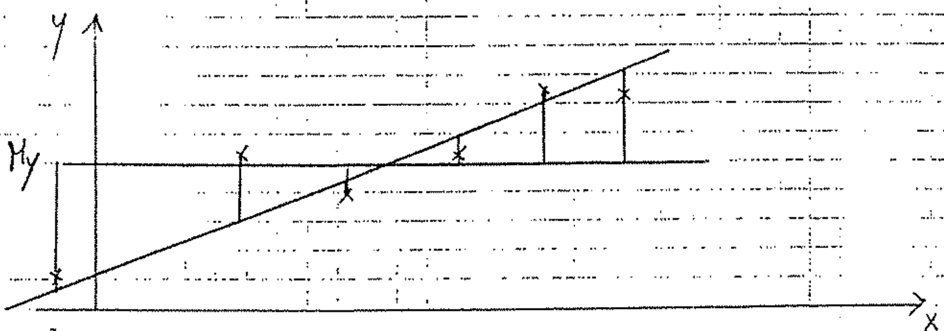
$$R^2 := \frac{\bar{J}(\bar{\theta}_1^{LS}) - J(\theta^{LS})}{\bar{J}(\bar{\theta}_1^{LS})} = 1 - \frac{J(\theta^{LS})}{\bar{J}(\bar{\theta}_1^{LS})} \leq 1$$

coeff. di determinazione multipla

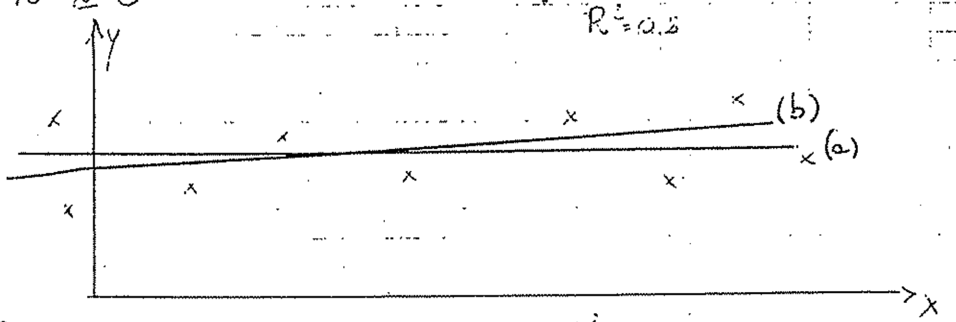
scarti attorno alla retta di regressione

scarti attorno a M_y

$$0 \leq R \leq 1$$

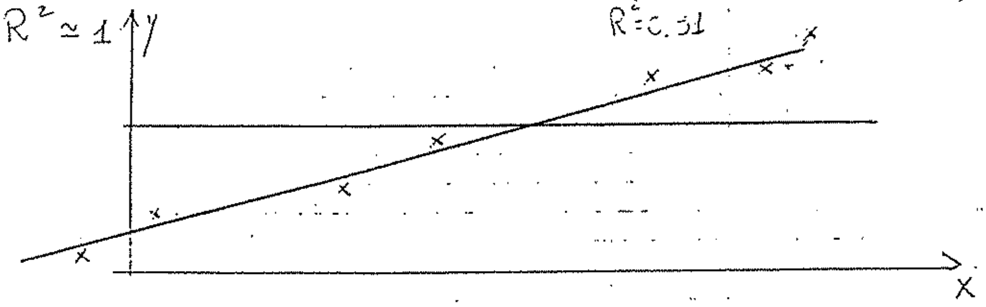


Se $R^2 \simeq 0$



Miglioro poco passando da (a) a (b)

Se $R^2 \simeq 1$



il modello (b) è molto meglio

Best

Linear

Unbiased

Estimator

Se rimuovo l'ipotesi di gaussianità, cosa posso dire?

IPOTESI I2) $Y = \Phi\theta + V$ ($\exists$ un modello vero) $E[V] = 0$ $Var[V] = \sigma^2 \Psi$ (σ^2 scalare eventualmente incognito)

<u>Teorema (Gauss & Markov)</u>

Si consideri la cifra di merito

$$J^M(\theta) := \varepsilon^T \Psi^{-1} \varepsilon = (Y - \Phi\theta)^T \Psi^{-1} (Y - \Phi\theta)$$

che è minimizzata da

$$\theta^M = (\Phi^T \Psi^{-1} \Phi)^{-1} \Phi^T \Psi^{-1} Y$$ [illegible]

Allora:

(i) θ^M è tra tutti gli stimatori lineari e non polarizzati (→ il vettore dei parametri dipende linearmente da Y), lo stimatore che minimizza $Var[\hat{\theta}-\theta^0]$ (BLUE)

(ii) $Var[\theta^M] = \sigma^2 (\Phi^T \Psi^{-1} \Phi)^{-1}$

Osservazioni

- $Var[V] = \sigma^2 I \Rightarrow \theta^M = \theta^{LS}$
- Si dimostra che (se σ^2 è incognito)

$$\hat{\sigma}^2 = \frac{J^M(\theta^M)}{N-q}$$ è uno stimatore non polarizzato di σ^2

- Usando σ^2 (o $\hat{\sigma}^2$) posso stimare anche $Var[\theta^M] = \Sigma_{\theta^M} = \hat{\sigma}^2 (\Phi^T \Psi^{-1} \Phi)^{-1}$
- ATTENZIONE! Dato che non conosco la d d p di $V \Rightarrow$
  - non posso più dire che $J^M(\theta^M) \sim \chi^2_{N-q}$ (niente più validazione del modello)
  - non posso calcolare gli intervalli di confidenza (se non con Tchebycheff, che però dà intervalli troppo conservativi)

Idea: fornire solo le SD dei parametri
- Cade il Test F
- Posso usare FPE

STIMA DI BAYES

→ informazioni a priori su θ

Ipotesi I3 : $Y = \Phi \underset{\text{parametri vettore di V.C}}{\Theta} + V$ $\quad V \sim N(0, \Sigma_V)$, $\Sigma_V > 0$, $\Theta \sim N(m_\Theta, \Sigma_\Theta)$

$\Sigma_\Theta > 0$ $\quad V, \Theta$ indip.

Teorema Sotto I3

a) $\theta^B := \arg\min_\theta \{ \underbrace{\varepsilon^T \Sigma_V^{-1} \varepsilon}_{\text{somma dei pesi dei residui sperimentali}} + (\theta - \underset{\text{informazioni a priori}}{m_\Theta})^T \underset{\text{misura l'affidabilità dei dati di cui dispongo a priori}}{\Sigma_\Theta^{-1}} (\theta - m_\Theta) \}$

b) $\theta^B = (\Phi^T \Sigma_V^{-1} \Phi + \Sigma_\Theta^{-1})^{-1} (\Phi^T \Sigma_V^{-1} Y + \Sigma_\Theta^{-1} m_\Theta)$

c) $Var[\theta^B] = [\Phi^T \Sigma_V^{-1} \Phi + \Sigma_\Theta^{-1}]^{-1}$

Commenti:

- $\Sigma_\Theta^{-1} \to 0 \Rightarrow \theta^B \to \theta^{ML}$

L'informazione a priori vale molto poco — si basa interamente sui dati sperimentali

- Non è più necessario ipotizzare $rank(\Phi) = q$ $\quad (\Sigma_\Theta > 0 \Rightarrow (\Phi^T \Sigma_V^{-1} \Phi + \Sigma_\Theta^{-1}) > 0)$

<u>Sfruttando le informazioni a priori posso avere $q > N$</u> (più incognite dei dati)