IDENTIFICAZIONE PARAMETRICA

G. De Nicolao, R. Scattolini
Dipartimento di Informatica e Sistemistica
Università degli Studi di Pavia
Via Ferrata 1
27100 Pavia

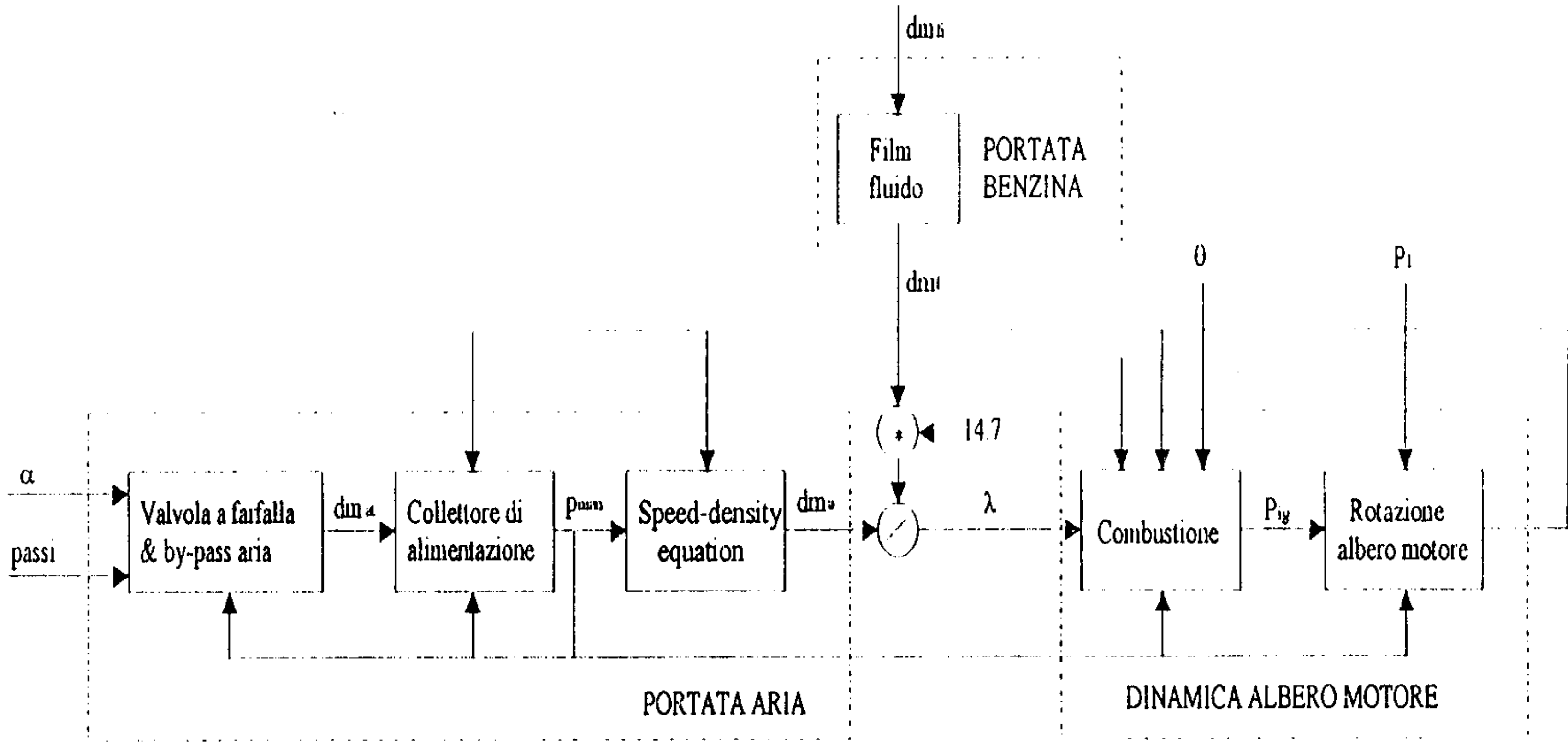
INTRODUZIONE

- Cos'è l'identificazione? L'arte di ***tarare*** o ***ricavare*** modelli a partire dai dati sperimentali

- Anche nei modelli ricavati a partire da considerazioni fisiche vi sono parametri incogniti, il cui valore deve essere dedotto dai dati. In questo caso si parla di modelli ***a scatola grigia***

- Il alcuni casi le equazioni del modello "fisico" non sono note o i loro parametri non possono essere stimati singolarmente. Vale la pena di identificare dei modelli in grado di rappresentare i legami tra le variabili del sistema in esame (ingressi e uscite). La loro struttura è scelta indipendentemente da considerazioni "fisiche". Questi modelli vengono chiamati ***a scatola nera***

- Queste diverse situazioni vengono illustrate con un esempio: la modellistica dei principali fenomeni che avvengono nei motori a combustione interna


Film
fluido
PORTATA
BENZINA
0
Pl
14.7
α
passi
Valvola a farfalla
& by-pass aria
Collettore di
alimentazione
Speed-density
equation
λ
Combustione
Pig
Rotazione
albero motore
PORTATA ARIA
DINAMICA ALBERO MOTORE

2bis

SOTTOSISTEMA "ARIA"

L'andamento della pressione media nel collettore di aspirazione p può essere descritto dalla seguente equazione differenziale

$$\dot{p}(t) = -\frac{V}{120\overline{V}}\eta n(t)p(t) + \frac{RT(t)}{\overline{V}}\dot{m}(t)$$

dove

V = cilindrata del motore

n = numero di giri dell'albero motore

R = costante universale dei gas

T = temperatura media nel collettore

$\dot{m}$ = portata d'aria nel collettore

$\overline{V}$ = volume del collettore

η = rendimento volumetrico

Anche supponendo di conoscere tutte le altre grandezze (ipotesi poco realistica), certamente il rendimento volumetrico non è noto e dipende a sua volta da n, p, T e altre grandezze. Il problema di stima è quindi quello di determinare un'espressione corretta per η in funzione di queste variabili.

Il modello è a scatola grigia

SOTTOSISTEMA "BENZINA"

Un modello fisico che tenga conto del fenomeno dell'evaporazione all'interno del collettore, e quindi della presenza di una miscela bifase, è estremamente difficile da ricavare. D'altra parte, nelle sue caratteristiche essenziali il fenomeno è abbastanza semplice e con dinamica assimilabile a quella di un sistema lineare del primo ordine. Per questo un modello fenomenologico comunemente accettato è il seguente

$$\begin{cases} \dot{m}_{fv} = (1-X)\dot{m}_{fi} \\ \ddot{m}_f = \dfrac{X\dot{m}_{fi} - \dot{m}_{ff}}{\tau} \\ \dot{m}_f = \dot{m}_{fv} + \dot{m}_{ff} \end{cases}$$

dove

$\dot{m}_{fv}$ = portata di combustibile che entra nel collettore sotto forma di vapore

$\dot{m}_{fi}$ = portata di combustibile iniettato

$\dot{m}_{ff}$ = portata di combustibile sotto forma di film fluido

$\dot{m}_f$ = portata di combustibile che entra nei cilindri

La funzione di trasferimento tra $\dot{m}_{fi}$ e $\dot{m}_f$ è

$$G(s) = \frac{1+(1-X)\tau}{1+\tau s}$$

in cui i parametri X e τ devono essere ricavati sperimentalmente con procedure di identificazione.

Il modello è a scatola grigia

DINAMICA DELL'ALBERO MOTORE

Anche in questo caso si può determinare un modello a partire dallo studio dei fenomeni in esame, in particolare

$$\dot{n} = \frac{H\eta\dot{m}_f}{nI} - \frac{P_f + P_p + P_b}{nI}$$

dove

n = numero di giri

I = momento d'inerzia normalizzato all'albero motore

H = potere calorico inferiore del carburante

η = efficienza termica indicata

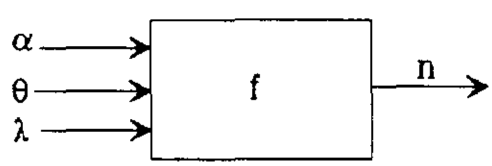
P_f = potenza dissipata per attriti

P_p = potenza utilizzata per il pompaggio

P_b = potenza disponibile all'albero motore

In pratica tutte queste grandezze sono non note e andrebbero stimate singolarmente dai dati, in particolare per quanto riguarda la loro dipendenza dalle variabili esogene del motore, cioè l'angolo di apertura della valvola a farfalla (o di by-pass) α, l'angolo di anticipo accensione θ, il titolo λ. Questo richiederebbe una lunga serie di prove molto focalizzate (e costose).

Vale allora la pena di cercare un modello non basato su considerazioni legate al funzionamento del motore, ma comunque in grado di "spiegare" l'andamento di n a partire dagli ingressi α, θ, λ.



Come è fatta la funzione f? Quanto valgono i suoi parametri?

Il modello è a scatola nera

Programma del corso

- *Variabili Casuali*
  - Funzione di Distribuzione
  - Densità di Probabilità
  - Media e Varianza
  - Gaussianità
  - Variabili Casuali Vettoriali

- *Il Concetto di Stimatore*

- *Modelli Lineari e Non Lineari nei Parametri*

- *Stima ai Minimi Quadrati*
  - Equazioni Normali
  - Identificabilità
  - Interpretazione Probabilistica (Best linear Unbiased Estimator)

- *Validazione e Scelta del Modello*
  - Test χ^2
  - Test F
  - Crossvalidazione
  - Final Prediction Error (FPE)
  - Akaike Information Criterion (AIC)
  - Minimum Description Lenght (MDL)

- *Famiglie di Modelli Lineari*
  - Polinomiali

- *Stepwise Regression*
- *Aspetti Computazionali*
- *Stima Non Lineare ai Minimi Quadrati*
- *Reti Neurali*
  - Reti a Percettrone
  - Reti Neurali a Base Radiale
- *Trasformate Z e di Fourier*
- *Sistemi a segnali campionati*
  - Campionatore, mantenitore e sistema a segnali campionati
  - Stabilità
  - Rappresentazioni di stato e ingresso/uscita
  - Risposta in Frequenza
  - scelta del periodo di campionamento
- *Processi Casuali Stazionari*
  - Realizzazione
  - Medie del I e del II ordine
  - Rumore Bianco
  - Spettro
- *Fattorizzazione Spettrale Canonica*

- *Famiglie di Modelli Dinamici*
    - Modelli a Errore di Equazione
    - Modelli a Errore d'Uscita
    - Predizione ad un Passo
    - Identificazione a Minimizzazione dell'Errore di Predizione
- *Prefiltraggio e Identificazione in Banda*
- *Determinazione dei Segnali di Identificazione*
    - Persistente Eccitazione
    - PRBS
- *Test di Bianchezza*
- *Minimi Quadrati Ricorsivi*
- *Modelli NARX*
- *Cenni a Possibili Estensioni e Approfondimenti*

VARIABILI CASUALI

Contenuti:

- Nozione di variabile casuale
- Funzione di distribuzione
- Densità di probabilità
- Media e varianza
- Gaussianità
- Variabili casuali vettoriali
- Conclusioni

Motivazione: L'identificazione dei modelli si basa su dati sperimentali che sono affetti da errori di misura "casuali" $\Rightarrow$ anche i modelli stimati sono casuali.

NOZIONE DI VARIABILE CASUALE

Variabile Casuale (V.C.) X: un esperimento casuale il cui esito è un numero reale X

Esempio 1: Lancio di un dado. Possibili esiti: $X = 1, 2, 3, 4, 5, 6$. Se il dado è "onesto": $P(X=1) = P(X=2) = P(X=3) = P(X=4) = P(X=5) = P(X=6) = 1/6$

Esempio 2: Fermo per la strada una persona a caso e ne misuro la statura.

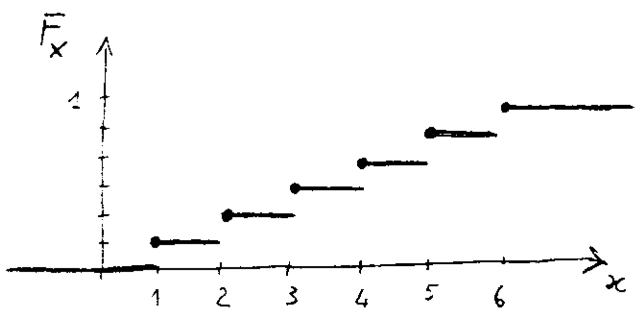
Esempio 3: L'errore di misura compiuto da un sensore in una determinata misurazione.

Nota: Nei tre esempi, ripetendo l'esperimento non ho nessuna garanzia di ottenere lo stesso risultato (casualità).

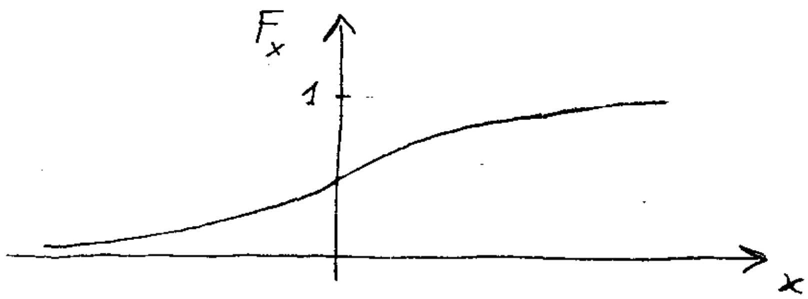
FUNZIONE DI DISTRIBUZIONE

Funzione di distribuzione (f.d.d.): $F_X(x) = P(X \leq x)$

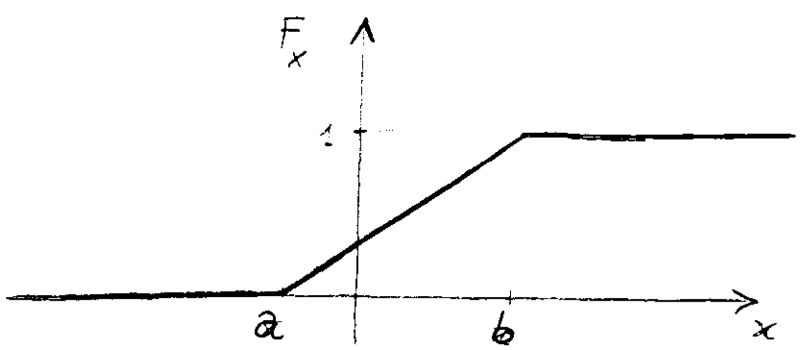
Esempio: Dado onesto



Esempio: Errore commesso da un sensore (analogico)



Esempio: Numero reale $\in [a,b]$ scelto a caso in modo "uniforme" $(P(x_1 \leq X \leq x_2) = |x_1 - x_2|, \forall x_1, x_2 \leq 1)$.



Alcune proprietà della f.d.d.:

- $F_X(x) \leq 1, \ \forall x$

- $F_X(-\infty) = 0, \quad F_X(\infty) = 1$

- $F_X(x)$ è monotona non decrescente

- $P(x_1 < X \leq x_2) = P(X \leq x_2) - P(X \leq x_1) = F_X(x_2) - F_X(x_1)$

- $P(X > x) = 1 - P(X \leq x) = 1 - F_X(x)$

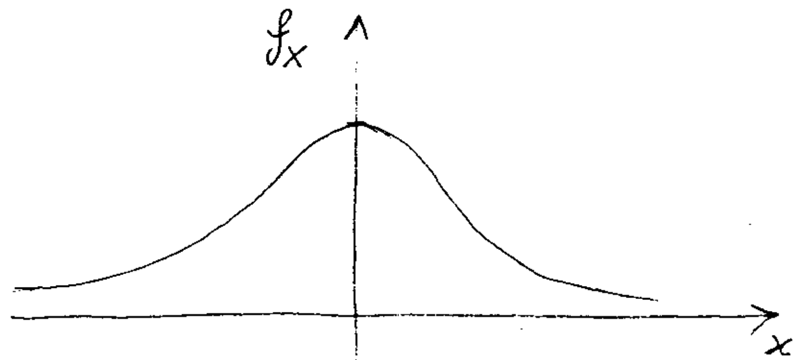
Per molte V.C. (la gaussiana per esempio), non esiste una formula esplicita per la f.d.d. $\Rightarrow$ tabella.

DENSITA' DI PROBABILITA'

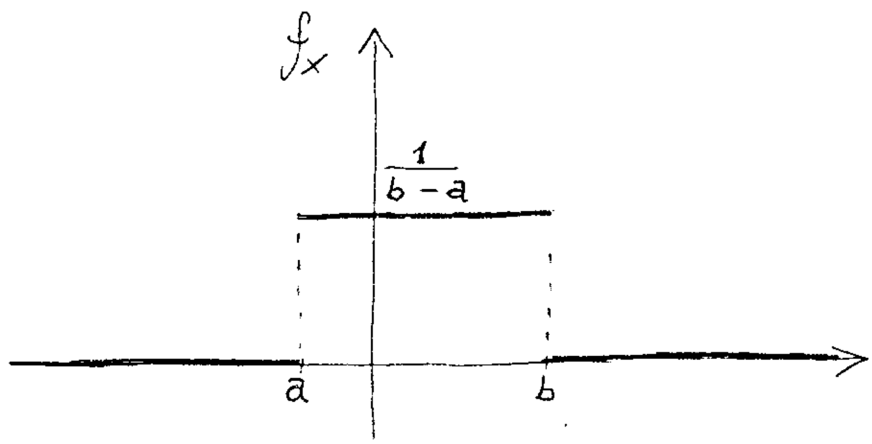
Densità di probabilità (d.d.p.): $f_X(x) = \dfrac{dF_X(x)}{dx}$

Interpretazione: $f_X(x)dx = P(x \leq X < x+dx)$

Esempio: Errore di misura



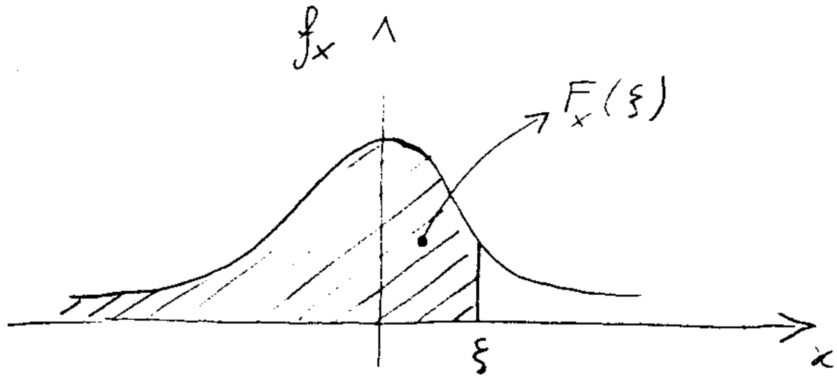
Esempio: Numero reale $\in [a, b]$ scelto a caso in modo "uniforme"



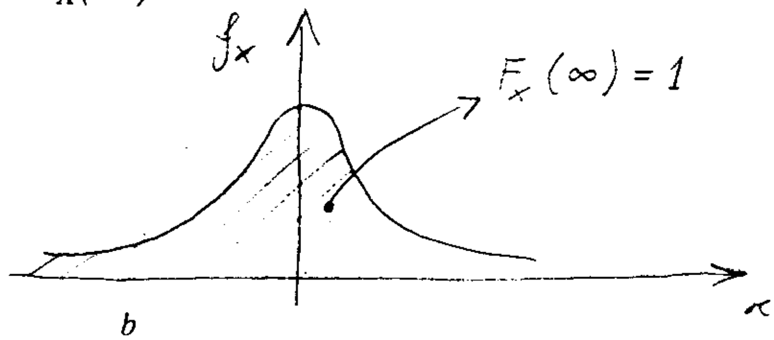
Alcune proprietà della d.d.p.:

- $f_X(x) \geq 0, \ \forall x$ (infatti $F_X(x)$ è monotona)

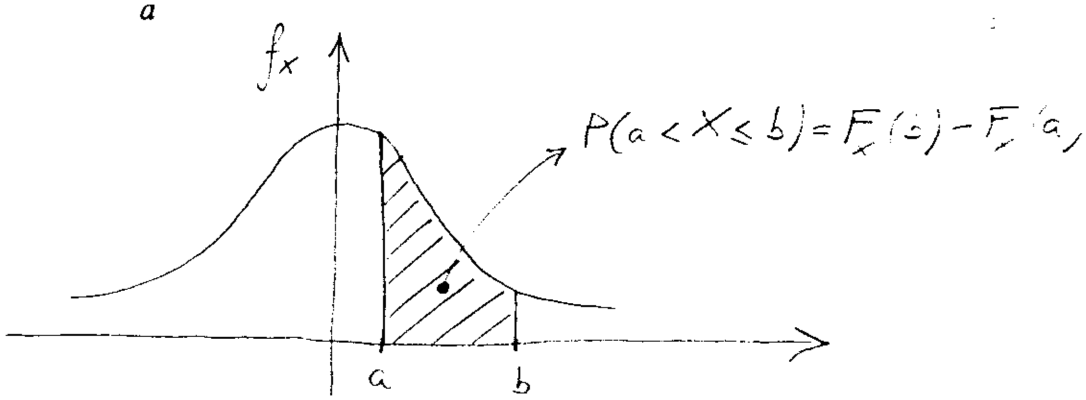
- $\displaystyle\int_{-\infty}^{\xi} f_X(x)dx = F_X(\xi)$



- $\displaystyle\int_{-\infty}^{+\infty} f_X(x)dx = F_X(\infty) = 1$



- $\displaystyle P(a < X \leq b) = \int_{a}^{b} f_X(x)dx$



MEDIA E VARIANZA

Media (o valore atteso): $E[X] = m_X = \int_{-\infty}^{+\infty} x\, f_X(x)dx$

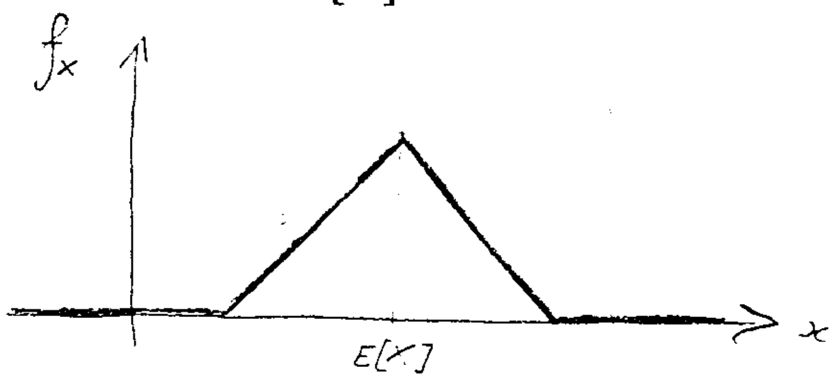
Interpretazione: "Baricentro" della d.d.p. $f_X(x)$

Proprietà:

- d.d.p. simmetrica rispetto a $x = \bar{x}$

$$\Downarrow$$

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



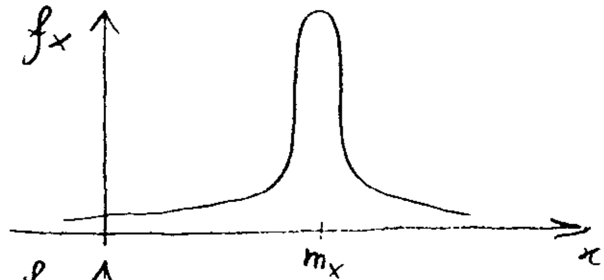
- $E[\alpha+\beta X] = \alpha + \beta E[X]$ (linearità)

Varianza: $Var[X] = \sigma_X{}^2 = \int_{-\infty}^{+\infty} (x-m_X)^2 f_X(x)dx$

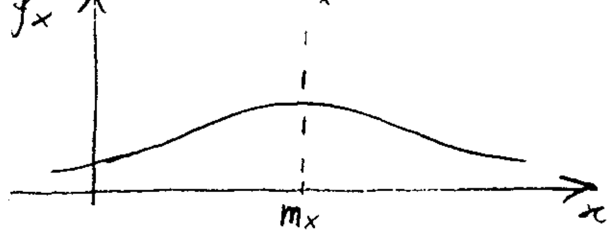
Deviazione standard: $\sqrt{Var[X]} = \sigma_X$

Interpretazione: "Momento di inerzia" della d.d.p. $f_X(x)$ attorno al baricentro m_X.

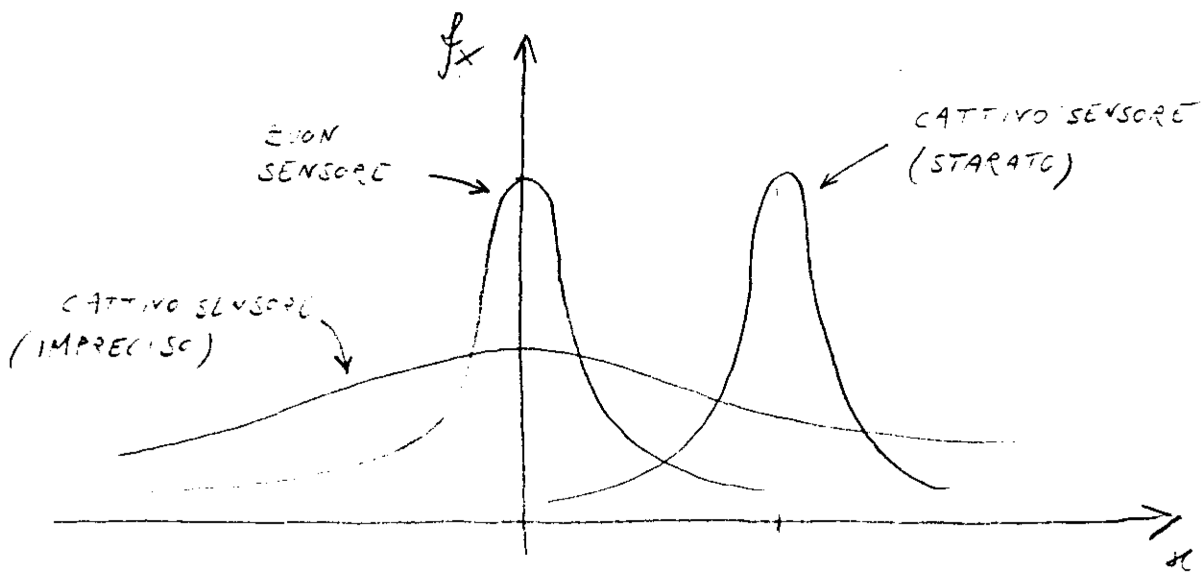
$Var[X]$ "piccola":



$Var[X]$ "grande":



L'errore di misura X di un buon sensore ha $E[X] \cong 0$ e $Var[X]$ "piccola".



Proprietà della varianza:

- $Var[X] = E[(X-m_X)^2]$
- $Var[X] = 0 \Rightarrow P(X = E[X]) = 1$
- $Var[\alpha+\beta X] = \beta^2 Var[X]$

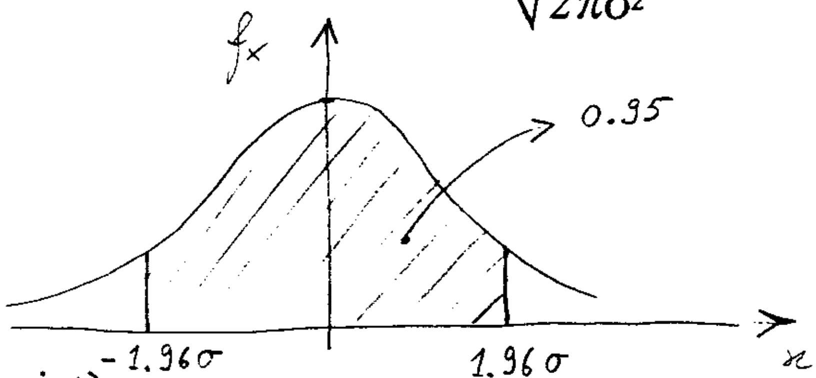
Standardizzazione: Data una V.C. X, definiamo

$$Y = \frac{X-E[X]}{\sqrt{Var[X]}}$$

Y è una nuova V.C. con $E[Y] = 0$, $Var[Y] = 1$

GAUSSIANITA'

d.d.p. gaussiana (normale): $f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-m)^2}{2\sigma^2}}$



Alcune proprietà:

- $E[X] = m$
- $Var[X] = \sigma^2$
- E' completamente caratterizzata da media e varianza

Notazione: $X \sim G(m,\sigma^2)$ (spesso anche $X \sim N(m,\sigma^2)$)

Appendix Table 2 Distribution function of the normal distribution

The table shows the area under the curve $y = (2\pi)^{-\frac{1}{2}}e^{-\frac{1}{2}x^2}$ lying to the left of specified deviates x; e.g. the area corresponding to a deviate 1·86 (= 1·5 + 0·36) is 0·9686.

Deviate	0·0 +	0·5 +	1·0 +	1·5 +	2·0 +	2·5 +	3·0 +	3·5 +
0·00	5000	6915	8413	9332	9772	9^2379	9^2865	9^377
0·01	5040	6950	8438	9345	9778	9^2396	9^2869	9^378
0·02	5080	6985	8461	9357	9783	9^2413	9^2874	9^378
0·03	5120	7019	8485	9370	9788	9^2430	9^2878	9^379
0·04	5160	7054	8508	9382	9793	9^2446	9^2882	9^380
0·05	5199	7088	8531	9394	9798	9^2461	9^2886	9^381
0·06	5239	7123	8554	9406	9803	9^2477	9^2889	9^381
0·07	5279	7157	8577	9418	9808	9^2492	9^2893	9^382
0·08	5319	7190	8599	9429	9812	9^2506	9^2897	9^383
0·09	5359	7224	8621	9441	9817	9^2520	9^2900	9^383
0·10	5398	7257	8643	9452	9821	9^2534	9^303	9^384
0·11	5438	7291	8665	9463	9826	9^2547	9^306	9^385
0·12	5478	7324	8686	9474	9830	9^2560	9^310	9^385
0·13	5517	7357	8708	9484	9834	9^2573	9^313	9^386
0·14	5557	7389	8729	9495	9838	9^2585	9^316	9^386
0·15	5596	7422	8749	9505	9842	9^2598	9^318	9^387
0·16	5636	7454	8770	9515	9846	9^2609	9^321	9^387
0·17	5675	7486	8790	9525	9850	9^2621	9^324	9^388
0·18	5714	7517	8810	9535	9854	9^2632	9^326	9^388
0·19	5753	7549	8830	9545	9857	9^2643	9^329	9^389
0·20	5793	7580	8849	9554	9861	9^2653	9^331	9^389
0·21	5832	7611	8869	9564	9864	9^2664	9^334	9^390
0·22	5871	7642	8888	9573	9868	9^2674	9^336	9^390
0·23	5910	7673	8907	9582	9871	9^2683	9^338	9^404
0·24	5948	7704	8925	9591	9875	9^2693	9^340	9^408
0·25	5987	7738	8944	9599	9878	9^2702	9^342	9^412
0·26	6026	7764	8962	9608	9881	9^2711	9^344	9^415
0·27	6064	7794	8980	9616	9884	9^2720	9^346	9^418
0·28	6103	7823	8997	9625	9887	9^2728	9^348	9^422
0·29	6141	7852	9015	9633	9890	9^2736	9^350	9^425
0·30	6179	7881	9032	9641	9893	9^2744	9^352	9^428
0·31	6217	7910	9049	9649	9896	9^2752	9^353	9^431
0·32	6255	7939	9066	9656	9898	9^2760	9^355	9^433
0·33	6293	7967	9082	9664	9901	9^2767	9^357	9^436
0·34	6331	7995	9099	9671	9904	9^2774	9^358	9^439
0·35	6368	8023	9115	9678	9906	9^2781	9^360	9^441
0·36	6406	8051	9131	9686	9909	9^2788	9^361	9^443
0·37	6443	8078	9147	9693	9911	9^2795	9^362	9^446
0·38	6480	8106	9162	9699	9913	9^2801	9^364	9^448
0·39	6517	8133	9177	9706	9916	9^2807	9^365	9^450
0·40	6554	8159	9192	9713	9918	9^2813	9^366	9^452
0·41	6591	8186	9207	9719	9920	9^2819	9^368	9^454
0·42	6628	8212	9222	9726	9922	9^2825	9^369	9^456
0·43	6664	8238	9236	9732	9925	9^2831	9^370	9^458
0·44	6700	8264	9251	9738	9927	9^2836	9^371	9^459
0·45	6736	8289	9265	9744	9929	9^2841	9^372	9^461
0·46	6772	8315	9279	9750	9931	9^2846	9^373	9^463
0·47	6808	8340	9292	9756	9932	9^2851	9^374	9^464
0·48	6844	8365	9306	9761	9934	9^2856	9^375	9^466
0·49	6879	8389	9319	9767	9936	9^2861	9^376	9^467

Note—Decimal points in the body of the table are omitted. Repeated 9's are indicated by powers, e.g. 9^371 stands for 0·99971.

V.C. VETTORIALI

Variabile Casuale Vettoriale X: un esperimento casuale il cui esito è un vettore $X = [X_1\ X_2\ \dots\ X_n]'$ di numeri reali

Esempio 1: Lancio di due dadi. X_1 = *risultato del primo dado,* X_2 = *risultato del secondo dado.*

Esempio 2: Fermo per la strada una persona a caso e ne misuro la statura X_1 ed il peso X_2.

Esempio 3: Gli errori di misura $X_1, X_2, \dots, X_n$ compiuti in n misurazioni.

Funzione di distribuzione:

$$F_X(x) = F_X(x_1, x_2, \ldots, x_n) = P(X_1 \leq x_1, X_2 \leq x_2, \ldots, X_n \leq x_n)$$

Densità di probabilità: $f_X(x) = \dfrac{\partial^n F_X(x)}{\partial x_1 \partial x_2 \ldots \partial x_n}$

Indipendenza: X_1 e X_2 si dicono indipendenti se $F_X(x_1, x_2) = F_{X_1}(x_1) F_{X_2}(x_2)$

Interpretazione: Conoscere il valore di X_1 non mi è di nessun aiuto nel prevedere il valore assunto da X_2.

Esempio 1 (2 dadi): X_1 e X_2 indipendenti

Esempio 2 (statura e peso): X_1 e X_2 non indipendenti

Esempio 3 (errori di misura): in un buon sistema di misura, X_1, X_2 ,..., X_n dovrebbero essere indipendenti

In pratica, si assume che X_1 e X_2 siano indipendenti tutte le volte che sono generate da fenomeni fisici "non interagenti".

Media: $E[X] = [\ E[X_1]\ \ E[X_2]\ \dots\ E[X_n]\ ]'$

Matrice varianza: $Var[X] = \mathrm{E}[(X-E[X])(X-E[X])']$

$$Var[X] = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \dots & \sigma_{1n} \\ \sigma_{21} & \sigma_{22} & \dots & \sigma_{2n} \\ \dots & \dots & \dots & \dots \\ \sigma_{n1} & \sigma_{n2} & \dots & \sigma_{nn} \end{bmatrix}$$

$$\sigma_{ii} = Var[X_i]$$
$$\sigma_{ij} = \mathrm{E}[(X_i-E[X_i])(X_j-E[X_j])] = Cov[X_i,X_j]$$

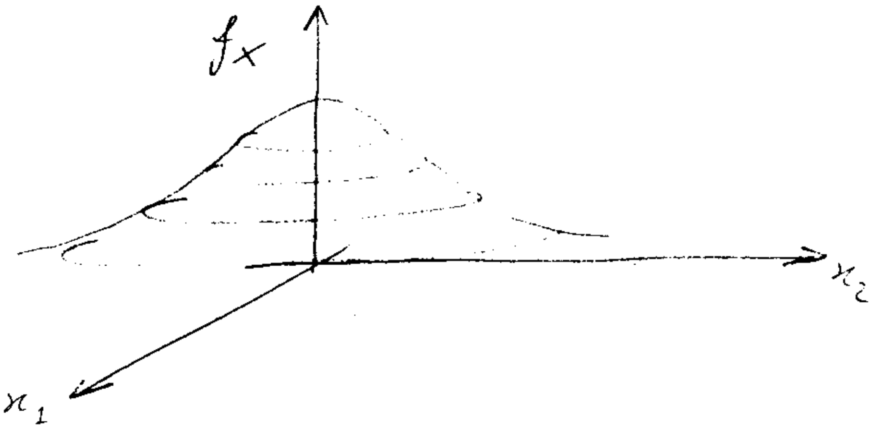
Incorrelazione: X_1 e X_2 si dicono incorrelate se $Cov[X_1,X_2] = 0$

Proprietà di media e varianza:

- $E[AX] = AE[X]$ (A matrice reale $m \times n$)
- $Var[AX] = AVar[X]A'$
- X_1 e X_2 indipendenti $\Rightarrow$ X_1 e X_2 incorrelate
  (in generale, X_1 e X_2 incorrelate $\not\Rightarrow$ X_1 e X_2 indipendenti)

V.C. gaussiana vettoriale: $f_X(x) = \dfrac{1}{\sqrt{(2\pi)^n det(\Sigma)}} e^{-\frac{(x-m)'\Sigma^{-1}(x-m)}{2}}$

Alcune proprietà:



- $E[X] = m$
- $Var[X] = \Sigma$
- E' completamente caratterizzata da media e matrice varianza
- Le singole V.C. $X_1, X_2, \dots, X_n$ sono gaussiane
- Data una matrice A ed un vettore B, se definisco $Y = AX + B$ la nuova V.C. vettoriale Y è ancora gaussiana con $E[Y] = AE[X]+B$, $Var[Y] = AVar[X]A'$
- X_1, X_2 incorrelate $\Rightarrow X_1, X_2$ indipendenti
  (per V.C. gaussiane indipendenza e incorrelazione sono nozioni equivalenti)

Teorema centrale del limite: Siano X_i V.C. indip.. Allora, sotto larghe ipotesi, la V.C. $Y = X_1+X_2+\dots+X_i$ tende, per $i \rightarrow \infty$, ad una V.C. con f.d.d. gaussiana.

(questo risultato spiega l'enorme diffusione "in natura" delle V.C. gaussiane)

CONCLUSIONI

- V.C.: strumento comodo per descrivere fenomeni casuali.

- Caratterizzazione completa di una V.C.: ci vuole la f.d.d. o la d.d.p.

- Caratterizzazione sintetica: media (baricentro) e varianza (indice della dispersione intorno al baricentro).

- V.C. gaussiane: sono completamente caratterizzate da media e varianza.

- Delle buone misure dovrebbero avere errori con media nulla, varianza piccola e indipendenti tra loro.

IL PROBLEMA DELLA STIMA

Contenuti:

- La nozione di stimatore
- Media e varianza campionaria
- Proprietà degli stimatori
- Intervalli di confidenza
- Conclusioni

Motivazione: Impostare il problema della stima di parametri incogniti a partire da dati affetti da errori casuali.

LA NOZIONE DI STIMATORE

Problema: stimare una o più grandezze incognite usando misure affette da "rumore" (descritto come V.C.)

Esempio 1: $Y_i = \theta + V_i \,, \qquad i = 1, 2, \ldots, N$

θ : parametro incognito
V_i : errori di misura
Y_i : misure "rumorose"

Stimatore "ovvio": $\hat{\theta} = \frac{1}{N} \sum_{i=1}^{N} Y_i$

Esempio 2: $I = V/R$ (Legge di Ohm). Applico diversi valori di tensione $V_1, V_2, \ldots, V_N$, e misuro le relative correnti $I_i = V_i/R$. Parametro incognito: $\theta = 1/R$.

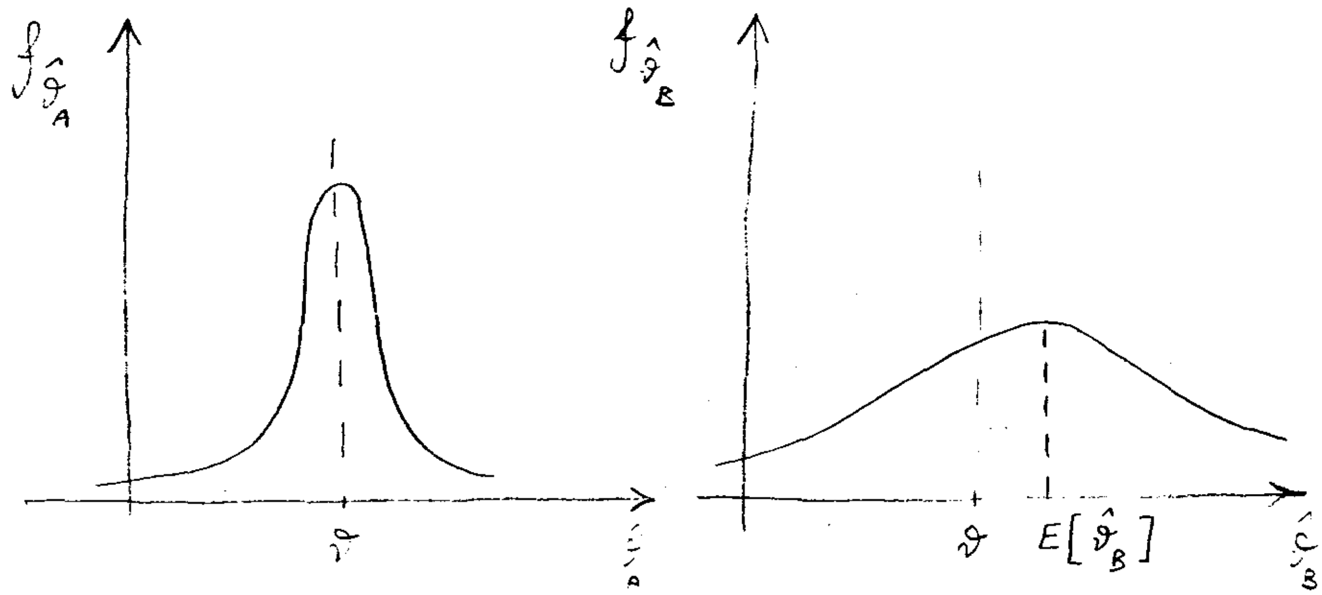
$$Y_i = V_i\theta + V_i \,, \qquad i = 1, 2, \ldots, N$$

Soluzione: ...

Osservazione fondamentale: $\hat{\theta} = \hat{\theta}(Y_1,\ldots,Y_N) \Rightarrow$ lo stimatore $\hat{\theta}$ è una funzione dei dati $Y_1,\ldots,Y_N$ che sono V.C. $\Rightarrow$ *anche lo stimatore è una V.C.*

$$\Downarrow$$

Per giudicare uno stimatore non basta vedere se "ci azzecca" in uno o due casi (sono risultati casuali!) ma bisogna analizzare la sua d.d.p.



Lo stimatore A è migliore di B

(una volta ogni tanto può però capitare che A commetta un errore maggiore di quello commesso da B)

MEDIA E VARIANZA CAMPIONARIA

Si supponga che siano disponibili N valori Y_i indipendenti e identicamente distribuiti con d.d.p. f_Y

Problema (classico): Stimare $E[Y_i]$ e $Var[Y_i]$

Media campionaria: $M(Y_1,\dots,Y_N) = \frac{1}{N}\sum_{i=1}^{N} Y_i$

Nota: Se f_Y è gaussiana, M è una V.C. gaussiana (infatti, è somma di V.C. gaussiane). Per N "grande", M è gaussiana anche se f_Y non lo è (per il teorema centrale del limite).

Proprietà fondamentale (Legge dei grandi numeri):

$\lim_{N\to\infty} M = E[Y_i]$, nel senso che $\lim_{N\to\infty} E[(M-E[Y_i])^2] = 0$
(convergenza in media quadratica: E[(errore di stima)^2] $\to 0$).

Varianza campionaria: $S^2(Y_1,\dots,Y_N) = \frac{1}{N}\sum_{i=1}^{N} (Y_i-M)^2$

PROPRIETA' DEGLI STIMATORI

Non polarizzazione: Uno stimatore si dice *non-polarizzato* (unbiased) quando $E[\hat{\theta}] = \theta$.

Esempio 1: Media campionaria ($\theta = E[Y_i]$, $\hat{\theta} = M$).

$$E[M] = \frac{1}{N}\sum_{i=1}^{N} E[Y_i] = E[Y_i] \Rightarrow \text{stimatore non polarizzato}$$

Esempio 2: Varianza campionaria ($\theta = Var[Y_i]$, $\hat{\theta} = S^2$). Con alcuni calcoli:

$$E[S^2] = \frac{N-1}{N} Var[Y_i] \neq Var[Y_i] \Rightarrow \text{stimatore polarizzato}$$

Uno stimatore non polarizzato della varianza è fornito dalla

Varianza campionaria "corretta": $S_c^2 = \frac{1}{N-1}\sum_{i=1}^{N} (Y_i-M)^2$

$$E[S_c^2] = \frac{N}{N-1} E[S^2] = Var[Y_i] \Rightarrow \text{stimatore non polarizzato}$$

Consistenza: Uno stimatore si dice *consistente* quando

$$\lim_{N \to \infty} \hat{\theta} = \theta$$

(dove la convergenza è da intendersi in media quadratica, per es.)
Interpretazione: al crescere di N, l'errore di stima tende a zero.

Esempio: M, S^2, S_c^2 sono tutti consistenti (lo si dimostra con la legge dei grandi numeri).

Asintotica normalità: Per $N \to \infty$, la f.d.d. di $\hat{\theta}$ tende a diventare gaussiana.

Esempio: M è asintoticamente normale (per il teorema centrale del limite). Si può dimostrare che vale lo stesso per S^2, S_c^2.

Stimatore a minima varianza: Per gli stimatori *non polarizzati*, $Var[\hat{\theta}]$ costituisce un indice della precisione dello stimatore.
Lo stimatore $\hat{\theta}^m$ è detto *a minima varianza* se $\mathrm{Var}[\hat{\theta}^m] < \mathrm{Var}[\hat{\theta}]$, dove $\hat{\theta}$ è un qualsiasi altro stimatore ($\hat{\theta}^m$ è il miglior stimatore possibile).

Esempio: Se i dati Y_i sono distribuiti gaussianamente, la media campionaria è lo stimatore a minima varianza.

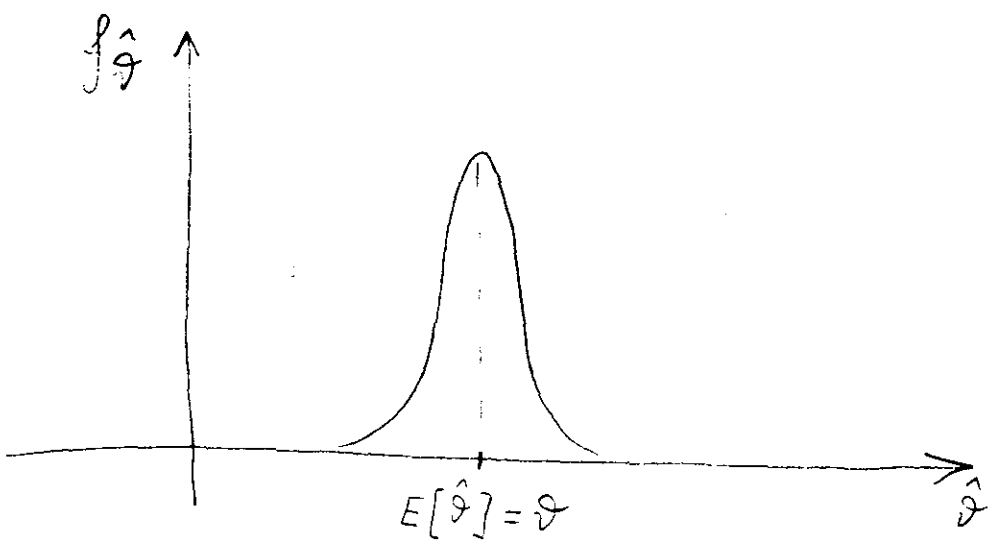
INTERVALLI DI CONFIDENZA

Dato che $\hat{\theta}(Y_1,\ldots,Y_N)$ è una V.C. conoscere $\hat{\theta}$ non mi dà garanzie sul valore effettivo di θ.

$$\Downarrow$$

E' privo di senso fornire stime $\hat{\theta}$ senza fornire indicazioni sull'ampiezza dell'errore di stima θ-$\hat{\theta}$

Se $\hat{\theta}$ è non polarizzato, può essere sufficiente conoscere $Var[\hat{\theta}]$ (se è piccola, la stima è precisa)

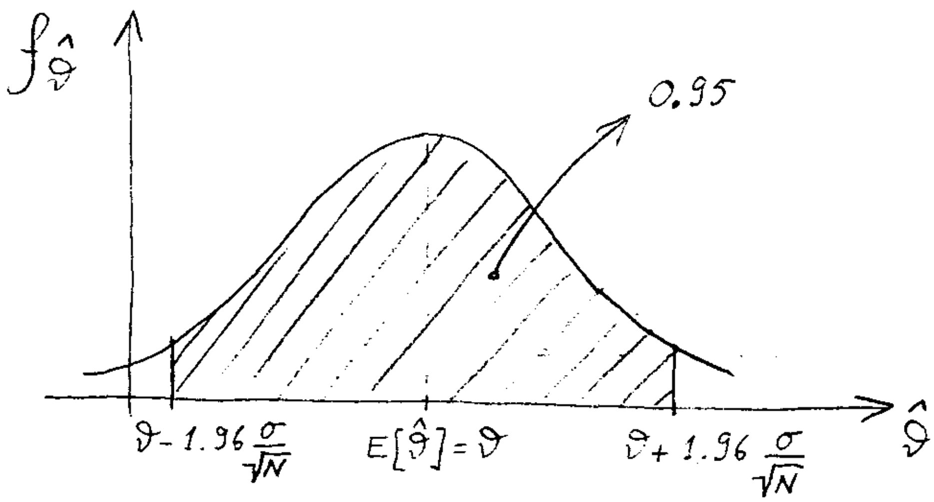


Se conosco la d.d.p. di $\hat{\theta}$ posso dare indicazioni più precise

Intervallo di confidenza al 95%: $I_{95} = [a_{95}, b_{95}]$ tale che

$$P(\theta \in I_{95}) = 0.95$$

Esempio: Media campionaria di V.C. gaussiane i.i.d. Y_i con $E[Y_i] = \theta$, $Var[Y_i] = \sigma^2$



$$I_{95} = [\hat{\theta} - 1.96 \frac{\sigma}{\sqrt{N}}, \hat{\theta} + 1.96 \frac{\sigma}{\sqrt{N}}]$$

In pratica, se risulta $I_{95} = [52.26 - 0.03, \ 52.26 + 0.03]$, scriveremo $\theta = 52.26 \pm 0.03$

Osservazione: per dimezzare I_{95} bisogna quadruplicare il numero N dei dati.

CONCLUSIONI

- Essendo basato su dati casuali, lo stimatore è una V.C.

- Criteri fondamentali per valutare la bontà di uno stimatore: media nulla (non polarizzazione) e varianza piccola (minima varianza).

- E' fondamentale fornire insieme alle stime un indice della loro affidabilità

$$\Downarrow$$

intervalli di confidenza