

UNIVERSITA' DEGLI STUDI DI GENOVA



FACOLTA' DI INGEGNERIA

Corso di Laurea Specialistica in Ingegneria Elettronica

Dispense per il corso di

Microsistemi, misure e segnali 1

prof. Giacomo Mario Bisio

prof. Silvio Paolo Sabatini

Appunti a cura di Mauro Gaggero e Orlando Mocci

Indice

1	Misure, incertezze ed errori	11
1.1	Significato del termine misura	11
1.2	Il concetto di incertezza	12
1.3	Sul concetto di errore; errori casuali e sistematici	13
1.3.1	Analisi statistica degli errori casuali	15
1.4	Propagazione degli errori e dell'incertezza	18
1.4.1	Grandezza misurata sommata a una costante numerica	19
1.4.2	Grandezza misurata moltiplicata per una costante numerica	20
1.4.3	Somma di due grandezze misurate	20
1.4.4	Caso generale	22
1.5	Deviazione standard della media	23
1.6	Propagazione dell'incertezza (senza distribuzione normale)	24
1.7	Valutazione dell'incertezza tipo	27
1.7.1	Incetezza tipo composta	28
1.7.2	Incetezza estesa	29
1.8	Correlazione tra le grandezze in ingresso	29
1.9	Esempi	31
1.9.1	Realizzazione di una resistenza serie	31
1.9.2	Misura di impedenza	32
1.9.3	Misura di una barretta metallica per confronto	34
1.10	Combinazione e confronto dei risultati di misure separate	36

1.10.1	Calcolo dell'incertezza da associare al valor medio	37
1.11	Compatibilit� tra due misure	38
1.12	Accettabilit� (confidenza) di due misure	39
1.12.1	Esempio	40
1.12.2	Osservazione sulle grandezze in gioco	41
2	Distribuzioni di probabilit� e modelli di rumore	43
2.1	Variabile di Bernoulli e distribuzione binomiale (*)	43
2.1.1	Caratteristiche della distribuzione binomiale (*)	45
2.2	Distribuzione di Gauss	47
2.2.1	La gaussiana come approssimazione della binomiale	47
2.2.2	Modello degli errori casuali	49
2.3	Distribuzione di Poisson	50
2.3.1	Caratteristiche della distribuzione di Poisson	51
2.3.2	La poissoniana come approssimazione della binomiale	51
2.4	Il rumore nel dominio delle frequenze e del tempo	55
2.5	Tipologie di rumore	58
2.5.1	Rumore shot	59
2.5.2	Rumore termico	62
3	Il problema della stima	69
3.1	Stima di parametri	69
3.1.1	Principi fondamentali della teoria della stima	70
3.1.2	Criteri di bont� della stima	70
3.2	Stima secondo il criterio del minimo errore quadratico medio	72
3.2.1	Stima lineare (caso monodimensionale)	72
3.2.2	Stima non lineare (caso monodimensionale)	75
3.2.3	Principio di ortogonalit�	77
3.2.4	Stima lineare (caso a pi� dimensioni)	79

3.2.5	Stima non lineare (caso a piú dimensioni)	82
3.2.6	Determinazione di densità di probabilità condizionali	83
3.3	Stima Bayesiana	83
3.3.1	Funzioni di perdita tipiche	84
3.3.2	Lo stimatore ottimo bayesiano	85
3.4	Stima a massima verosimiglianza	88
3.5	Stima di parametri con rumore additivo	89
3.6	Quadro complessivo delle varie tipologie di stima	94
3.6.1	Stima bayesiana	95
3.6.2	Stima MAP	96
3.6.3	Stima ML	96
3.6.4	Stima lineare	97
3.7	Applicazioni dei concetti esposti in precedenza	97
3.7.1	Esempio A	99
3.7.2	Esempio B	100
3.7.3	Esempio C	102
3.8	Rappresentazioni distribuite di una misura	104
3.8.1	Utilizzo di rivelatori per la misura di un parametro	107
4	Stima dinamica	111
4.1	Stima dello stato	111
4.1.1	Principi fondamentali della teoria della stima dinamica	111
4.2	Qualificazione statistica di un processo stocastico	112
4.2.1	Criteri di bontà della stima	112
4.2.2	Stima della funzione di autocorrelazione	113
4.2.3	Stima della densità spettrale di potenza	118
4.2.4	Campionamento di segnali aleatori	128
4.2.5	Formulazione numerica delle stime	132
4.2.6	Scelta della finestra temporale	134

4.3	Stima di un processo stocastico	136
4.3.1	Filtro di Wiener–Hopf	137
4.3.2	Tipologie del processo di stima	142
4.3.3	Il problema di smoothing in dettaglio	144
4.3.4	Il problema di predizione in dettaglio	146
4.3.5	Inconvenienti del filtro di Wiener	146
4.4	Filtro di Kalman (cenni)	147
4.4.1	Soluzione del problema di Kalman (cenni)	149
4.4.2	Struttura ricorsiva del filtro	150
5	Microsistemi elettromeccanici	153
5.1	Definizione di microsistema	153
5.2	Tecnologie alla base dei microsistemi	154
5.2.1	Microsistemi in silicio	155
5.3	Tipologie di microsistemi	157
5.3.1	Dispositivi microfluidodinamici	157
5.3.2	Dispositivi inerziali	158
5.3.3	Dispositivi a radiofrequenza	158
5.3.4	Dispositivi microattuatori	160
5.3.5	Dispositivi ottici	160
5.4	Aspetti economici che influenzano la realizzazione di microsistemi	160
A	Richiami di calcolo delle probabilità	163
A.1	Concetto di variabile aleatoria	163
A.2	Expectation, varianza e scarto tipo	164
A.3	Teorema del limite centrale	166
B	La disuguaglianza di Schwartz	167
C	Il limite di Cramer-Rao	171

D Il teorema di De Moivre–Laplace	173
Bibliografia	177

Prefazione

Questi appunti costituiscono una traccia alle lezioni del corso di “Microsistemi, misure e segnali 1”, che copre tutti gli argomenti affrontati dal corso nell’Anno Accademico 2003/2004. I Capitoli 1, 2 e 5 riguardano argomenti spiegati a lezione dal prof. G. M. Bisio, mentre i Capitoli 3 e 4 riguardano concetti esposti dal prof. S. P. Sabatini. Le Appendici contengono invece approfondimenti di quanto visto a lezione o richiami di concetti di particolare importanza utili per la comprensione degli altri argomenti. I paragrafi asteriscati () sono da considerarsi di sola lettura e non sono indispensabili al fine della comprensione degli argomenti trattati, in quanto costituiscono un approfondimento delle lezioni.*

Capitolo 1

Misure, incertezze ed errori

1.1 Significato del termine misura

Una misura é il risultato di una misurazione, ossia una serie di operazioni aventi lo scopo di determinare il valore di una quantità, in genere non noto, che prende il nome di misurando.

Il ruolo di un *sistema di misura* può essere definito considerando che il valore del misurando può non essere direttamente osservabile. Ciò si verifica per le *misure indirette*¹. Anche nel caso di *misure dirette*², effettuate comparando il misurando a un insieme di grandezze campione, non si osserva direttamente il valore del misurando ma piuttosto il valore di quel campione che é identificato essere equivalente al misurando attraverso una qualche specie di comparazione. Pertanto in generale possiamo definire la funzione del sistema di misura come

¹*misure indirette* = misure ottenute con metodo di misurazione detto indiretto o assoluto o derivato, nel quale il valore del misurando é determinato sulla base di una sua relazione funzionale nota con altre grandezze fisiche non della stessa specie del misurando; in questo caso non occorre disporre di campioni. Questo é un grande vantaggio; in compenso le misure assolute sono di regola piú difficili di quelle relative.

²*misure dirette* = misure ottenute con metodo di misurazione detto diretto o di confronto o relativo, nel quale il valore del misurando é determinato rispetto a una grandezza fisica della stessa specie di valore noto o assegnato: campione di misura.

quella di fornire *effetti osservabili* o uscite, legate in modo noto al valore del misurando. Si può quindi definire un modello matematico della misurazione, che permette di esprimere la grandezza da misurare come funzione di altre grandezze

$$y = f(x_1, x_2, \dots, x_n; \theta_1, \theta_2, \dots, \theta_p) \quad (1.1)$$

dove y rappresenta la grandezza da misurare, x_1, x_2, \dots, x_n rappresentano le grandezze osservate che consentono di risalire a y e $\theta_1, \theta_2, \dots, \theta_p$ sono le cosiddette grandezze di influenza, quantità che non sono l'oggetto della misura, ma che influenzano il valore del misurando e/o l'indicazione dello strumento di misura.

1.2 Il concetto di incertezza

Ripetendo la misurazione della stessa grandezza x più volte, pur nell'ipotesi che tale grandezza non sia variata, si osservano generalmente valori diversi x_1, x_2, \dots, x_n che tendono a distribuirsi in un certo range attorno a un valore centrale

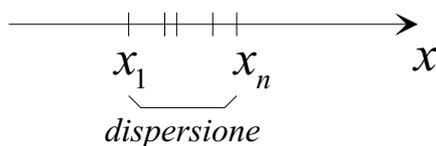


Figura 1.1: Esempio di dispersione dei valori di una misurazione.

Nessuna misura, per quanto fatta con cura, può essere completamente libera da incertezze. E' quindi di fondamentale importanza essere capaci di calcolare queste incertezze e di ridurle al minimo. L'*incertezza di misura* è il parametro, associato al risultato di una misura, che caratterizza la dispersione dei valori che potrebbero essere attribuiti al misurando. Questo parametro può essere specificato come scarto tipo (deviazione standard) oppure come intervallo (o semi-intervallo)

di valori con un livello di fiducia (confidence) specificato

$$x_{misurato} = x \pm dx \quad (1.2)$$

Il valore del misurando sta con una specificata percentuale di probabilità all'interno dell'intervallo $[x - dx, x + dx]$.

I motivi per cui l'incertezza di una qualsiasi misura non può essere eliminata sono molteplici. In generale le componenti dell'incertezza sono classificabili in accordo al metodo usato per valutarla:

- *Valutazione di tipo A*: si tratta di una valutazione empirica, statistica; richiede la ripetizione della misura. Si effettuano più misurazioni della stessa grandezza; è un metodo di valutazione dell'incertezza mediante analisi statistica dell'insieme di osservazioni;
- *Valutazione di tipo B*: è basata su una conoscenza a priori. Si ipotizza una certa distribuzione di probabilità a priori dei valori osservabili, basata sulla conoscenza posseduta, sia del metodo di misurazione, sia degli strumenti impiegati, sia delle grandezze di riferimento (standard). In questo metodo non vengono effettuate misurazioni ripetute della grandezza, ma l'incertezza viene specificata su base probabilistica.

1.3 Sul concetto di errore; errori casuali e sistematici

Indichiamo con V il valore “vero” del misurando e con A la stima di V . Il valore vero non è mai noto, mentre se ne conosce la stima. Si definisce *errore assoluto* la differenza tra il valore misurato V_m e il valore vero V del misurando:

$$E = V - V_m \quad (1.3)$$

L'errore assoluto commesso potrebbe essere valutato esattamente solamente se conoscessimo il valore vero del misurando. Poiché si conosce solo la stima del misurando, l'errore assoluto non potrà mai essere calcolato esattamente. Questo é il motivo per cui una misura sar  sempre affetta da incertezza. Errore e incertezza non sono sinonimi, in quanto l'errore non   noto con esattezza, mentre l'incertezza   nota, ed   un numero positivo K che permette di maggiorare l'errore assoluto

$$|E| \leq K \quad (1.4)$$

Inoltre, mentre gli errori sono quantit  con segno, le incertezze sono sempre quantit  positive. Riassumendo i concetti esposti, ogni volta che si effettua una misurazione si compie anche un certo errore, che non   possibile determinare con esattezza. E' possibile stimare questo errore tramite un suo maggiorante, l'incertezza. Le cause degli errori sono da imputare a fattori casuali, il cui effetto pu  essere ridotto mediante osservazioni ripetute della grandezza da misurare, e ad *effetti sistematici*, dovuti a grandezze di influenza che non siano state prese in considerazione, e il cui effetto pu  essere ridotto utilizzando modelli opportuni che li considerino (mentre la ripetizione della misura non permette di ridurre l'errore sistematico, in quanto esso possiede sempre lo stesso segno). Consideriamo al riguardo il seguente esempio.

Esempio. Si vuole misurare la lunghezza di una sbarra di metallo. Se si effettua la misurazione senza considerare la temperatura, la misura sar  affetta da errore sistematico. E' possibile, conoscendo il coefficiente di dilatazione della sbarra, correggere la misurazione e rimuovere, o almeno ridurre, l'errore sistematico dovuto a questa grandezza di influenza. Il modello di misurazione potrebbe essere il seguente:

$$l = l_0 (1 + \alpha(T - T_0)) \quad (1.5)$$

dove

- l   la grandezza osservata,

- l_0 é il valore ricercato, la lunghezza della sbarra alla temperatura di riferimento T_0 ,
- T é la temperatura della sbarra a cui si effettua la misurazione,
- α é il coefficiente di dilatazione lineare della sbarra.

Si ricava

$$l_0 = \frac{l}{1 + \alpha(T - T_0)} = l \cdot f(l; \alpha, T) \quad (1.6)$$

dove la quantità $f(l; \alpha, T) = \frac{1}{1 + \alpha(T - T_0)}$ é chiamata *fattore di correzione*.

La rimozione dell'errore sistematico puó non essere completa per la limitatezza del modello. E' necessario prevenire le possibili sorgenti di errore sistematico e accertarsi che tutti gli errori sistematici siano molto minori della precisione richiesta. Nel caso dell'esempio tale limite deriva dall'aver assunto lineare la dipendenza della lunghezza della sbarra dalla temperatura.

1.3.1 Analisi statistica degli errori casuali

La trattazione degli errori casuali é del tutto diversa da quella degli errori sistematici. Si vogliono prendere in esame le metodologie che consentono di ridurre l'incertezza del risultato della misurazione. Ci concentreremo solo sugli errori casuali, supponendo trascurabili le sorgenti di errore sistematico. Sia data una grandezza osservata X e una serie di n osservazioni X_1, X_2, \dots, X_n . Per le ipotesi di errore fatte, X puó essere considerata a tutti gli effetti una variabile aleatoria (vedere l'Appendice A per un breve richiamo sui concetti di variabile aleatoria).

Utilizzando n osservazioni indipendenti X_1, X_2, \dots, X_n della grandezza X si vogliono ricavare le informazioni sulla distribuzione statistica, ovvero si vuole dare una stima del valore medio e della deviazione standard (scarto tipo) della distribuzione. Si assume che la distribuzione statistica dovuta agli errori casuali

sia *gaussiana* (o *normale*), ossia caratterizzata da una densità di probabilità del tipo

$$g_{\bar{X},\sigma}(X) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(X - \bar{X})^2}{2\sigma^2}\right] \quad (1.7)$$

dove abbiamo indicato con X , \bar{X} e σ rispettivamente la grandezza osservata, il suo valor medio e il suo scarto tipo.

Nota questa distribuzione di probabilità, la probabilità di osservare X compresa tra X_1 e $X_1 + dX_1$ risulta pari a

$$\begin{aligned} \Pr\{X_1 \leq X \leq X_1 + dX_1\} &= \int_{X_1}^{X_1+dX_1} \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(X_1 - \bar{X})^2}{2\sigma^2}\right] dX \simeq \\ &\simeq \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(X_1 - \bar{X})^2}{2\sigma^2}\right] dX_1 \end{aligned} \quad (1.8)$$

in quanto l'intervallo di integrazione è infinitesimo.

Poiché non è di nostro interesse la dimensione dell'intervallo dX_1 , né il fattore $\frac{1}{\sqrt{2\pi}}$, l'espressione può essere abbreviata come segue

$$\Pr\{X_1 \leq X \leq X_1 + dX_1\} \propto \frac{1}{\sigma} \exp\left[-\frac{(X_1 - \bar{X})^2}{2\sigma^2}\right] \quad (1.9)$$

Sinteticamente possiamo indicare la probabilità di osservare X_1 come risultato della misurazione nel seguente modo:

$$\Pr\{X_1\} = \Pr\{X_1 \leq X \leq X_1 + dX_1\} \propto \frac{1}{\sigma} \exp\left[-\frac{(X_1 - \bar{X})^2}{2\sigma^2}\right] \quad (1.10)$$

La probabilità di ottenere una osservazione X_2 è

$$\Pr\{X_2\} \propto \frac{1}{\sigma} \exp\left[-\frac{(X_2 - \bar{X})^2}{2\sigma^2}\right] \quad (1.11)$$

In modo del tutto analogo possiamo scrivere tutte le probabilità, sino a

$$\Pr\{X_n\} \propto \frac{1}{\sigma} \exp\left[-\frac{(X_n - \bar{X})^2}{2\sigma^2}\right] \quad (1.12)$$

Poiché abbiamo considerato un insieme di osservazioni indipendenti, la probabilità di ottenere l'intero insieme di osservazioni (probabilità congiunta) è dato

dal prodotto delle probabilità marginali

$$\begin{aligned} \Pr\{X_1, X_2, \dots, X_n\} &= \Pr\{X_1\} \cdot \Pr\{X_2\} \cdots \Pr\{X_n\} \\ &\propto \frac{1}{\sigma^n} \exp \left[- \sum_{k=1}^n \frac{(X_k - \bar{X})^2}{2\sigma^2} \right] \end{aligned} \quad (1.13)$$

Vogliamo ora determinare i parametri \bar{X} e σ^2 (valor medio e varianza) a partire dalle n misurazioni X_1, X_2, \dots, X_n . Per trovare le migliori stime per questi parametri statistici usiamo il *principio di massima verosimiglianza*. Questo principio stabilisce che le migliori stime possibili sono quei valori che rendono $\Pr\{X_1, X_2, \dots, X_n\}$ massima. Analizzando (1.13) osserviamo che la probabilità é massima quando la sommatoria all'esponente é minima, ossia dobbiamo risolvere il seguente problema di ottimizzazione:

$$\min_{\bar{X}, \sigma} \sum_{k=1}^n \frac{(X_k - \bar{X})^2}{2\sigma^2} \quad (1.14)$$

Per calcolare questo minimo differenziamo rispetto a \bar{X} e poniamo uguale a zero la derivata, ottenendo

$$\frac{\partial}{\partial \bar{X}} \left(\sum_{k=1}^n \frac{(X_k - \bar{X})^2}{2\sigma^2} \right) = 0 \quad \Rightarrow \quad 2 \sum_{k=1}^n \frac{(X_k - \bar{X})}{2\sigma^2} = 0 \quad (1.15)$$

da cui

$$\text{miglior stima di } \bar{X} = x = \frac{1}{n} \sum_{k=1}^n X_k \quad (1.16)$$

Quindi la miglior stima per il valore medio \bar{X} é la *media campione* delle n osservazioni.

Possiamo ricavare la migliore stima della deviazione standard differenziando $\Pr\{X_1, X_2, \dots, X_n\}$ rispetto a σ e uguagliando il tutto a zero:

$$\begin{aligned} \frac{\partial}{\partial \sigma} \left(\frac{1}{\sigma^n} \exp \left[- \sum_{k=1}^n \frac{(X_k - \bar{X})^2}{2\sigma^2} \right] \right) &= \frac{-n}{\sigma^{n+1}} \exp \left[- \sum_{k=1}^n \frac{(X_k - \bar{X})^2}{2\sigma^2} \right] + \\ &+ \frac{-1}{\sigma^n} \exp \left[- \sum_{k=1}^n \frac{(X_k - \bar{X})^2}{2\sigma^2} \right] (-1) \frac{-2}{2\sigma^3} \sum_{k=1}^n (X_k - \bar{X})^2 = 0 \end{aligned} \quad (1.17)$$

Svolgendo i calcoli necessari si ricava

$$\sigma^2 = \frac{1}{n} \sum_{k=1}^n (X_k - \bar{X})^2 \quad (1.18)$$

Sostituendo ora a \bar{X} la sua stima x otteniamo lo stimatore della varianza

$$s^2 = \frac{1}{n} \sum_{k=1}^n (X_k - x)^2 \quad (1.19)$$

Nel passare dalla (1.18) alla (1.19) abbiamo in realtà sottostimato σ . Questa sottostima può essere corretta sostituendo il denominatore n con $n - 1$. La migliore stima per σ^2 risulta quindi

$$s^2 = \frac{1}{n-1} \sum_{k=1}^n (X_k - x)^2 \quad (1.20)$$

In sintesi abbiamo determinato una specificazione dell'incertezza di tipo A, basata sull'analisi statistica delle osservazioni della grandezza di interesse X . I valori misurati sono dispersi intorno a x con uno scarto tipo pari a s .

1.4 Propagazione degli errori e dell'incertezza

Il problema della propagazione degli errori sorge quando misuriamo una o più grandezze X_1, X_2, \dots, X_n e delle grandezze di influenza $\theta_1, \theta_2, \dots, \theta_p$, tutte con incertezza, e poi utilizziamo i valori misurati per ricavare una grandezza Y che dipende da esse attraverso una certa funzione f :

$$Y = f(X_1, X_2, \dots, X_n; \theta_1, \theta_2, \dots, \theta_p) \quad (1.21)$$

Se le grandezze da misurare sono soggette solo a errori casuali, possiamo, a partire dalla distribuzione delle misure delle grandezze osservate, risalire alla distribuzione di Y . Cominciamo col considerare alcuni casi semplici, al fine di dare in seguito una trattazione del tutto generale.

1.4.1 Grandezza misurata sommata a una costante numerica

Supponiamo di osservare X e di voler calcolare

$$Y = X + A \quad (1.22)$$

dove A é una costante numerica non affetta da incertezza e le grandezze in gioco hanno distribuzione di tipo gaussiano. La probabilità di ottenere un valore X é

$$\Pr\{\text{osservare il valore } X\} \propto \frac{1}{s_X} \exp \left[-\frac{(X - \bar{x})^2}{2s_X^2} \right] \quad (1.23)$$

La probabilità di osservare Y risulta essere

$$\Pr\{\text{osservare il valore } Y\} = \Pr\{\text{osservare il valore } X = Y - A\} \quad (1.24)$$

ossia

$$\Pr\{\text{osservare } Y\} \propto \frac{1}{s_Y} \exp \left[-\frac{(Y - A - \bar{X})^2}{2s_Y^2} \right] = \frac{1}{s_Y} \exp \left[-\frac{(Y - (\bar{X} + A))^2}{2s_X^2} \right] \quad (1.25)$$

Pertanto si ricava $\bar{Y} = \bar{X} + A$ e $s_Y = s_X$.

Il risultato mostra che i valori di Y sono normalmente distribuiti e sono centrati sul valore $\bar{Y} = \bar{X} + A$.

La dispersione intorno al valore medio é la stessa dispersione che si ha sulla grandezza osservata X (vedi figura 1.2).

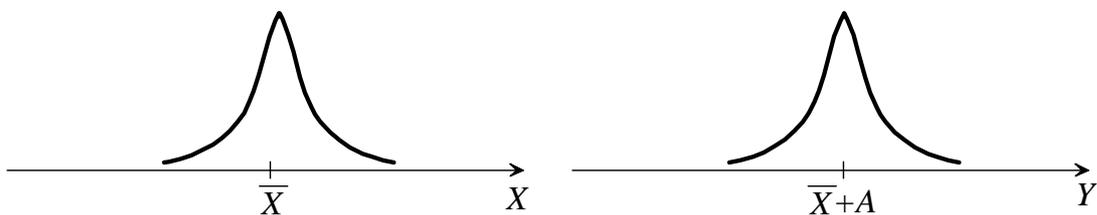


Figura 1.2: Rappresentazione grafica dei risultati ottenuti nel caso di grandezza misurata sommata a una costante numerica.

1.4.2 Grandezza misurata moltiplicata per una costante numerica

Supponiamo di osservare X e di voler calcolare

$$Y = B \cdot X \quad (1.26)$$

dove B é una costante numerica il cui valore é noto senza incertezza. Ripetendo gli stessi ragionamenti del caso precedente e supponendo ancora distribuzioni gaussiane abbiamo

$$\Pr\{Y\} = \Pr\left\{\frac{Y}{B}\right\} \propto \exp\left[-\frac{\left(\frac{Y}{B} - \bar{X}\right)^2}{2s_X^2}\right] = \exp\left[-\frac{(Y - B\bar{X})^2}{2B^2s_X^2}\right] \quad (1.27)$$

Come risultato si ha $\bar{Y} = B\bar{X}$ e $s_Y^2 = Bs_X^2$. In questo caso lo scarto tipo é amplificato o ridotto del valore di B rispetto allo scarto che caratterizza la grandezza osservata X (vedi figura 1.3).

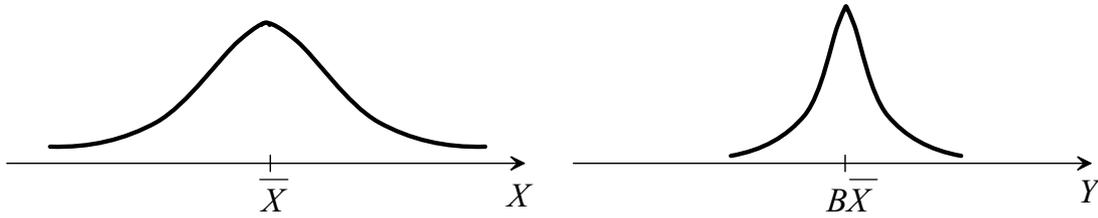


Figura 1.3: Rappresentazione grafica dei risultati ottenuti nel caso di grandezza misurata moltiplicata per una costante numerica.

1.4.3 Somma di due grandezze misurate

Supponiamo di osservare due grandezze indipendenti X_1 e X_2 e di calcolarne la somma

$$Y = X_1 + X_2 \quad (1.28)$$

Supponiamo che le misure di X_1 e X_2 siano normalmente distribuite attorno ai loro valori medi. In un primo momento per semplificare la trattazione si ipotizza

tali valori medi siano nulli. In questo caso la probabilità di ottenere un particolare valore di X_1 é

$$\Pr\{X_1\} \propto \exp \left[-\frac{X_1^2}{2s_{X_1}^2} \right] \quad (1.29)$$

Analogamente la probabilità di ottenere un particolare valore di X_2 é

$$\Pr\{X_2\} \propto \exp \left[-\frac{X_2^2}{2s_{X_2}^2} \right] \quad (1.30)$$

Poiché le due grandezze osservate sono indipendenti, la probabilità di osservare Y é data dal prodotto delle due probabilità marginali

$$\Pr\{Y\} = \Pr\{X_1\} \cdot \Pr\{X_2\} \propto \exp \left[-\left(\frac{X_1^2}{2s_{X_1}^2} + \frac{X_2^2}{2s_{X_2}^2} \right) \right] \quad (1.31)$$

Si può osservare che sussiste la seguente identità

$$\frac{X_1^2}{A} + \frac{X_2^2}{B} = \frac{(X_1 + X_2)^2}{A + B} + \frac{(BX_1 - AX_2)^2}{AB(A + B)} \quad (1.32)$$

Posto

$$z^2 = \frac{(BX_1 - AX_2)^2}{AB(A + B)} \quad (1.33)$$

la $\Pr\{X_1, X_2\}$ si può esprimere nel seguente modo:

$$\Pr\{X_1, X_2\} \propto \exp \left[-\left(\frac{X_1^2}{2s_{X_1}^2} + \frac{X_2^2}{2s_{X_2}^2} \right) \right] = \exp \left[-\left(\frac{(X_1 + X_2)^2}{2(s_{X_1}^2 + s_{X_2}^2)} + \frac{z^2}{2} \right) \right] \quad (1.34)$$

L'ultima espressione é interpretabile come $\Pr\{X_1 + X_2, z\}$, ossia la probabilità di osservare indipendentemente X_1 e X_2 é la stessa di osservare indipendentemente $(X_1 + X_2)$ e z (si veda l'equazione (1.33)).

Quindi possiamo ottenere $\Pr\{X_1 + X_2\}$ integrando $\Pr\{X_1 + X_2, z\}$ su tutti i valori di z , in quanto siamo interessati a $\Pr\{X_1 + X_2\}$ indipendentemente dal valore di z . Integrando si ricava

$$\Pr\{Y\} = \Pr\{X_1 + X_2\} = \int_{-\infty}^{+\infty} \Pr\{X_1 + X_2, z\} dz \propto \exp \left[-\left(\frac{(X_1 + X_2)^2}{2(s_{X_1}^2 + s_{X_2}^2)} \right) \right] \quad (1.35)$$

dove $\bar{Y} = 0$ in quanto per ipotesi $\bar{X}_1 = 0$ e $\bar{X}_2 = 0$, mentre $s_Y^2 = s_{X_1}^2 + s_{X_2}^2$. La dispersione della Y risulta quindi piú ampia di quelle della X_1 e della X_2 .

Consideriamo ora il caso in cui \bar{X}_1 e \bar{X}_2 siano diversi da zero. Possiamo riscrivere la relazione (1.28) nella forma

$$Y = (X_1 - \bar{X}_1) + (X_2 - \bar{X}_2) + (\bar{X}_1 + \bar{X}_2) \quad (1.36)$$

I primi due termini fra parentesi hanno valor medio nullo, mentre l'ultima una quantità costante. Sfruttando i risultati ricavati per il caso

$$Y = X + A \quad (1.37)$$

(vedi paragrafo 1.4.1) si ottiene

$$\bar{Y} = (\bar{X}_1 - \bar{X}_1) + (\bar{X}_2 - \bar{X}_2) + (\bar{X}_1 + \bar{X}_2) = \bar{X}_1 + \bar{X}_2 \quad (1.38)$$

$$s_Y^2 = s_{X_1}^2 + s_{X_2}^2 \quad (1.39)$$

Infatti $X_1 - \bar{X}_1$ ha lo stesso scarto di X_1 , $X_2 - \bar{X}_2$ ha lo stesso scarto di X_2 e lo scarto di $\bar{X}_1 + \bar{X}_2$ è nullo. Per una rappresentazione grafica si veda la figura 1.4.

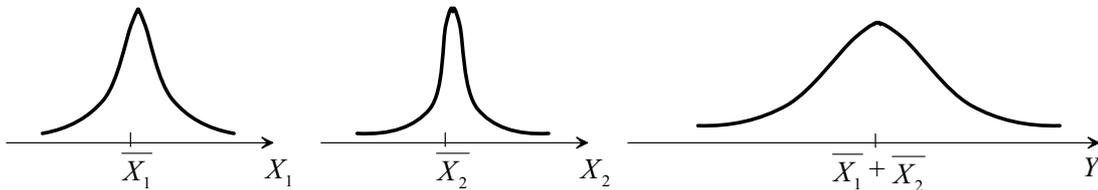


Figura 1.4: Rappresentazione grafica dei risultati ottenuti nel caso della somma di due grandezze misurate.

1.4.4 Caso generale

Supponiamo di misurare due grandezze indipendenti X_1 e X_2 e di voler conoscere una grandezza Y dipendente da X_1 e X_2 attraverso una generica funzione f :

$$Y = f(X_1, X_2) \quad (1.40)$$

Se consideriamo variazioni delle grandezze osservate in un campo di valori abbastanza piccolo, possiamo approssimare la funzione f con uno sviluppo di Taylor arrestato al primo ordine

$$Y = f(X_1, X_2) \simeq f(\bar{X}_1, \bar{X}_2) + \left. \frac{\partial f}{\partial X_1} \right|_{\bar{X}_1, \bar{X}_2} (X_1 - \bar{X}_1) + \left. \frac{\partial f}{\partial X_2} \right|_{\bar{X}_1, \bar{X}_2} (X_2 - \bar{X}_2) \quad (1.41)$$

Siamo in una situazione del tipo

$$Y = A + B_1(X_1 - \bar{X}_1) + B_2(X_2 - \bar{X}_2) \quad (1.42)$$

quindi sfruttando i risultati ricavati per $Y = BX + A$ con $\bar{X} = 0$ si ottiene $\bar{Y} = A$ e $s_Y^2 = Bs_X^2 = B_1s_{X_1}^2 + B_2s_{X_2}^2$. Si può ricavare facilmente

$$\bar{Y} = f(\bar{X}_1, \bar{X}_2) \quad (1.43)$$

$$s_Y^2 = \left. \frac{\partial f}{\partial X_1} \right|_{\bar{X}_1, \bar{X}_2} s_{X_1}^2 + \left. \frac{\partial f}{\partial X_2} \right|_{\bar{X}_1, \bar{X}_2} s_{X_2}^2 \quad (1.44)$$

La trattazione può essere facilmente estesa al caso in cui Y dipenda da più di due variabili.

1.5 Deviazione standard della media

Consideriamo una grandezza X ed effettuiamo su di essa una serie di n osservazioni, indicate con X_1, X_2, \dots, X_n . la miglior stima del valore medio è la media campione di queste osservazioni

$$Y = \text{stima di } \bar{X} = f(X_1, X_2, \dots, X_n) = \frac{X_1 + X_2 + \dots + X_n}{n} \quad (1.45)$$

Le varie osservazioni X_1, X_2, \dots, X_n costituiscono una sequenza di variabili aleatorie che fanno tutte riferimento alla stessa grandezza X , pertanto esse saranno tutte caratterizzate dallo stesso valor medio e dalla stessa varianza. Per ricavare la media e la deviazione standard di questo insieme di esperimenti sfruttiamo

i risultati ottenuti nel caso generale di propagazione degli errori (vedi paragrafo 1.4.4), considerando n variabili

$$y = f(\bar{X}_1, \bar{X}_2, \dots, \bar{X}_n) \quad (1.46)$$

$$s_Y^2 = \frac{\partial f}{\partial X_{i,1}} \Big|_{\bar{X}_1, \dots, \bar{X}_n} s_{X_1}^2 + \frac{\partial f}{\partial X_{i,2}} \Big|_{\bar{X}_1, \dots, \bar{X}_n} s_{X_2}^2 + \dots + \frac{\partial f}{\partial X_{i,n}} \Big|_{\bar{X}_1, \dots, \bar{X}_n} s_{X_n}^2 \quad (1.47)$$

Poiché

$$\frac{\partial f}{\partial X_i} \Big|_{\bar{X}_1, \bar{X}_2, \dots, \bar{X}_n} = \frac{1}{n} \quad i = 1, \dots, n \quad (1.48)$$

le relazioni precedenti diventano

$$y = f(\bar{X}_1, \bar{X}_2, \dots, \bar{X}_n) = \frac{\bar{X}_1 + \dots + \bar{X}_n}{n} \quad (1.49)$$

e

$$s_y^2 = \frac{1}{n^2} s_{X_1}^2 + \frac{1}{n^2} s_{X_2}^2 + \dots + \frac{1}{n^2} s_{X_n}^2 \quad (1.50)$$

Come anticipato X_1, X_2, \dots, X_n si riferiscono tutte alla stessa grandezza, quindi

$$s_{X_1}^2 = s_{X_2}^2 = \dots = s_{X_n}^2 = s_X^2 \quad (1.51)$$

Sostituendo tale risultato nella (1.50) si ottiene

$$s_y^2 = \frac{1}{n} s_X^2 \quad (1.52)$$

1.6 Propagazione dell'incertezza (senza distribuzione normale)

In questo paragrafo vogliamo determinare lo scarto tipo di una grandezza Y senza alcuna ipotesi sulla distribuzione delle grandezze osservate. Esprimiamo ancora Y nel seguente modo:

$$Y = f(X_1, X_2, \dots, X_n) \quad (1.53)$$

dove X_1, X_2, \dots, X_n sono variabili aleatorie (d'ora in avanti per comodità di scrittura ci riferiremo al caso di due sole grandezze X_1 e X_2 , senza perdita di generalità). Se si considera che la dispersione delle grandezze sia in un piccolo intervallo, e' possibile utilizzare l'approssimazione di Taylor della funzione f arrestata al primo ordine, ossia

$$Y = f(X_1, X_2) \simeq f(\bar{X}_1, \bar{X}_2) + \left. \frac{\partial f}{\partial X_1} \right|_{X_1, X_2} (X_1 - \bar{X}_1) + \left. \frac{\partial f}{\partial X_2} \right|_{X_1, X_2} (X_2 - \bar{X}_2) \quad (1.54)$$

Sfruttando la definizione di media campione

$$y = \frac{1}{n} \sum_{i=1}^n Y_i = \frac{1}{n} \sum_{i=1}^n f(X_{1,i}, X_{2,i}) \quad (1.55)$$

e l'approssimazione (1.54) abbiamo

$$\begin{aligned} y &= \frac{1}{n} \sum_{i=1}^n f(X_{1,i}, X_{2,i}) \simeq \\ &\simeq \frac{1}{n} \sum_{i=1}^n \left[f(\bar{X}_1, \bar{X}_2) + \left. \frac{\partial f}{\partial X_1} \right|_{X_1, X_2} (X_{1,i} - \bar{X}_1) + \left. \frac{\partial f}{\partial X_2} \right|_{X_1, X_2} (X_{2,i} - \bar{X}_2) \right] \stackrel{(1.56)}{=} \\ &= f(\bar{X}_1, \bar{X}_2) + 0 + 0 \end{aligned}$$

Quindi alla fine si ottiene lo stesso risultato trovato nell'ipotesi di gaussianità, che risulta quindi valido per una qualunque distribuzione

$$y = f(\bar{X}_1, \bar{X}_2) \quad (1.57)$$

Per quanto riguarda lo scarto tipo si ha

$$\begin{aligned} y &= \frac{1}{n} \sum_{i=1}^n (Y_i - \bar{Y})^2 \simeq \\ &\simeq \frac{1}{n} \sum_{i=1}^n \left[\left. \frac{\partial f}{\partial X_1} \right|_{X_1, X_2} (X_{1,i} - \bar{X}_1) + \left. \frac{\partial f}{\partial X_2} \right|_{X_1, X_2} (X_{2,i} - \bar{X}_2) \right]^2 = \\ &= \frac{1}{n} \left(\left. \frac{\partial f}{\partial X_1} \right|_{X_1, X_2} \right)^2 \sum_{i=1}^n (X_{1,i} - \bar{X}_1)^2 + \frac{1}{n} \left(\left. \frac{\partial f}{\partial X_2} \right|_{X_1, X_2} \right)^2 \sum_{i=1}^n (X_{2,i} - \bar{X}_2)^2 + \\ &+ \frac{2}{n} \left. \frac{\partial f}{\partial X_1} \right|_{X_1, X_2} \left. \frac{\partial f}{\partial X_2} \right|_{X_1, X_2} \sum_{i=1}^n (X_{1,i} - \bar{X}_1)(X_{2,i} - \bar{X}_2) \end{aligned} \quad (1.58)$$

Quindi alla fine

$$s_y^2 = \left(\frac{\partial f}{\partial X_1} \Big|_{X_1, X_2} \right)^2 s_{X_1}^2 + \left(\frac{\partial f}{\partial X_2} \Big|_{X_1, X_2} \right)^2 s_{X_2}^2 + 2 \frac{\partial f}{\partial X_1} \Big|_{X_1, X_2} \frac{\partial f}{\partial X_2} \Big|_{X_1, X_2} s_{X_1 X_2} \quad (1.59)$$

dove la quantità

$$s_{X_1 X_2} = \frac{1}{n} \sum_{i=1}^n (X_{1,i} - \bar{X}_1)(X_{2,i} - \bar{X}_2) \quad (1.60)$$

prende il nome di *covarianza*.

Lo scarto tipo del misurando dipende da due variabili che possono anche non essere scorrelate. La covarianza tiene conto della correlazione tra le due variabili. E' possibile maggiorare s_y^2 sfruttando la *disuguaglianza di Schwartz* (si veda al riguardo l'appendice B):

$$|s_{X_1 X_2}| \leq s_{X_1} s_{X_2} \quad (1.61)$$

Sostituendo tale disuguaglianza nella relazione (1.59) si ottiene

$$s_y^2 \leq \left[\left| \frac{\partial f}{\partial X_1} \Big|_{X_1, X_2} \right| s_{X_1} + \left| \frac{\partial f}{\partial X_2} \Big|_{X_1, X_2} \right| s_{X_1} \right]^2 \quad (1.62)$$

Quindi

$$s_y \leq \left| \frac{\partial f}{\partial X_1} \Big|_{X_1, X_2} \right| s_{X_1} + \left| \frac{\partial f}{\partial X_2} \Big|_{X_1, X_2} \right| s_{X_1} \quad (1.63)$$

Lo scarto tipo é maggiorato dalla somma degli scarti tipo, pesati da opportuni *coefficienti di sensibilità*

$$\frac{\partial f}{\partial X_1} \Big|_{X_1, X_2}, \quad \frac{\partial f}{\partial X_2} \Big|_{X_1, X_2} \quad (1.64)$$

che rappresentano quanto varia la Y in funzione di X_1 e X_2 .

Si noti che il risultato trovato é diverso dal caso in cui si supponga una distribuzione gaussiana, in cui le variabili aleatorie erano state supposte indipendenti. La maggiorazione riflette il caso piú sfavorevole.

1.7 Valutazione dell'incertezza tipo

Consideriamo ancora il modello di misurazione in cui il misurando può essere espresso come funzione di N altre grandezze X_1, X_2, \dots, X_N :

$$Y = f(X_1, X_2, \dots, X_N) \quad (1.65)$$

Si vuole determinare una stima del valor medio del misurando Y in funzione delle X_1, X_2, \dots, X_N e valutare l'incertezza.

Indichiamo con x_i la stima del valor medio \bar{X}_i della variabile aleatoria X_i (calcolato su n osservazioni distinte), ossia

$$x_i = \frac{1}{n} \sum_{k=1}^n X_{i,k} \simeq \bar{X}_i \quad (1.66)$$

e con y la stima del valor medio \bar{Y} del misurando:

$$y = \frac{1}{n} \sum_{k=1}^n Y_k = \frac{1}{n} \sum_{k=1}^n f(X_{1,k}, X_{2,k}, \dots, X_{N,k}) \simeq \bar{Y} \quad (1.67)$$

Lo scarto tipo stimato $u_c(y)$ (o *incertezza tipo composta*) di Y può essere determinata in funzione dello scarto tipo delle X_1, X_2, \dots, X_N . A tal proposito, la varianza sperimentale delle osservazioni può essere scritta come

$$s^2(X_i) = \frac{1}{n-1} \sum_{k=1}^n (X_{i,k} - \bar{X}_i)^2 \quad i = 1, 2, \dots, N \quad (1.68)$$

mentre la varianza sperimentale della media é

$$s^2(\bar{X}_i) = \frac{s^2(X_i)}{n} \quad i = 1, 2, \dots, N \quad (1.69)$$

dove $s(\bar{X}_i)$ prende il nome di scarto tipo sperimentale della media.

Per una grandezza di ingresso X_i determinata mediante n osservazioni ripetute indipendenti $X_{i,k}$, l'incertezza tipo $u(x_i)$ della sua stima $x_i \simeq \bar{X}_i$ é

$$u(x_i) = s(\bar{X}_i) \quad i = 1, 2, \dots, N \quad (1.70)$$

1.7.1 Incertezza tipo composta

E' conveniente separare la trattazione per il caso in cui le grandezze in ingresso non siano correlate e per quello in cui invece esse siano correlate.

Grandezze in ingresso non correlate

Sulla base dei ragionamenti precedenti abbiamo:

$$u_c^2(y) = \sum_{i=1}^N \left(\frac{\partial f}{\partial X_i} \right)^2 u^2(x_i) \quad (1.71)$$

dove

$$c_i = \frac{\partial f}{\partial X_i} \quad i = 1, 2, \dots, N \quad (1.72)$$

prendono il nome di *coefficienti di sensibilità*.

Posto $u_i(y) = c_i u(x_i)$ si ottiene

$$u_c^2(y) = \sum_{i=1}^N u_i^2(y) \quad (1.73)$$

Nel caso in cui il misurando Y possa essere espresso come

$$Y = C X_1^{p_1} X_2^{p_2} \cdots X_N^{p_N} \quad (1.74)$$

e nell'ipotesi che i coefficienti p_i siano noti senza incertezza e l'unica incertezza riguardi i valori X_i , si può ricorrere a una formulazione dell'incertezza in termini relativi:

$$\left(\frac{u_c(y)}{y} \right)^2 = \sum_{i=1}^N \left(\frac{p_i u(x_i)}{x_i} \right)^2 \quad (1.75)$$

Grandezze in ingresso correlate

Nel caso in cui le grandezze X_i siano tra loro correlate abbiamo:

$$u_c^2(y) = \sum_{i=1}^N \left(\frac{\partial f}{\partial X_i} \right)^2 u^2(x_i) + 2 \sum_{i=1}^{N-1} \sum_{j=i+1}^N \left(\frac{\partial f}{\partial X_i} \right) \left(\frac{\partial f}{\partial X_j} \right) u(x_i, x_j) \quad (1.76)$$

dove $u(x_i, x_j)$ é la *covarianza stimata* associata a x_i e x_j .

Introducendo il *coefficiente di correlazione*

$$r(x_i, x_j) = \frac{u(x_i, x_j)}{u(x_i)u(x_j)} \quad (1.77)$$

si ha

$$u_c^2(y) = \sum_{i=1}^N c_i^2 u^2(x_i) + 2 \sum_{i=1}^{N-1} \sum_{j=i+1}^N c_i c_j u(x_i) u(x_j) r(x_i, x_j) \quad (1.78)$$

1.7.2 Incertezza estesa

L'incertezza estesa U é una quantità usata soprattutto in ambito normativo, ed é legata all'incertezza composta dalla relazione

$$U = K u_c(y) \quad (1.79)$$

dove K prende il nome di *fattore di copertura*.

L'incertezza estesa viene utilizzata per determinare un intervallo di osservazione in cui cada una percentuale desiderata delle osservazioni della grandezza Y da misurare

$$y - U \leq Y \leq y + U \quad (1.80)$$

L'osservazione generica della Y é compresa tra $y - U$ e $y + U$ con una probabilità P , che prende il nome di *livello di copertura* o *livello di fiducia*.

1.8 Correlazione tra le grandezze in ingresso

La correlazione tra le grandezze in ingresso puó avere diverse origini. Nel caso ad esempio che tra le grandezze di ingresso intervengano la temperatura e la frequenza di un oscillatore, quest'ultima é influenzata dalla temperatura. In questo caso la correlazione nasce da un preciso legame fisico tra le grandezze in gioco.

Le grandezze X_1, X_2, \dots, X_N possono a loro volta essere funzioni di altre grandezze Q_j . Consideriamo per semplicitá il caso di due sole grandezze, X_i e

X_j , dipendenti da L grandezze Q :

$$\begin{aligned} X_i &= F(Q_1, Q_2, \dots, Q_L) \\ X_j &= G(Q_1, Q_2, \dots, Q_L) \end{aligned} \quad (1.81)$$

Le grandezze Q_j sono assunte per ipotesi tra di loro scorrelate. La correlazione tra X_i e X_j dipende dal fatto che esse possono dipendere, rispettivamente tramite le funzioni F e G , da una stessa grandezza Q_j .

La covarianza di due variabili casuali é una espressione quantitativa della loro dipendenza mutua, e puó essere stimata disponendo di n coppie di osservazioni indipendenti simultanee

$$s(Y_i, Z_i) = \frac{1}{n-1} \sum_{j=1}^n (Y_j - \bar{Y}_i)(Z_j - \bar{Z}_i) \quad (1.82)$$

La covarianza stimata delle medie \bar{Y} e \bar{Z} risulta essere

$$s(\bar{Y}_i, \bar{Z}_i) = \frac{s(Y_i, Z_i)}{n} \quad (1.83)$$

Osservazione. Per definizione abbiamo

$$u(x_i, x_j) = s(\bar{X}_i, \bar{X}_j) \quad (1.84)$$

quindi

$$u(x_i, x_j) = \frac{1}{n(n-1)} \sum_{k=1}^n (X_{i,k} - \bar{X}_i)(X_{j,k} - \bar{X}_j) \quad (1.85)$$

Effettuando ora una approssimazione lineare della dipendenza di X_i e X_j dalle variabili Q_j si ottiene:

$$X_{i,k} - \bar{X}_i \simeq \sum_{l=1}^L \left(\frac{\partial F}{\partial Q_l} \right) (Q_{l,k} - \bar{Q}_l) \quad (1.86)$$

$$X_{j,k} - \bar{X}_j \simeq \sum_{l=1}^L \left(\frac{\partial G}{\partial Q_l} \right) (Q_{l,k} - \bar{Q}_l) \quad (1.87)$$

Sostituendo le (1.86) e (1.87) nella (1.85), poiché le variabili Q_j sono scorrelate, i termini misti si annullano (almeno in prima approssimazione). Rimane dunque:

$$\begin{aligned} u(x_i, x_j) &= \sum_{l=1}^L \left(\frac{\partial F}{\partial Q_l} \right) \left(\frac{\partial G}{\partial Q_l} \right) \left[\frac{1}{n(n-1)} \sum_{k=1}^n (Q_{l,k} - \bar{Q}_l)^2 \right] = \\ &= \sum_{l=1}^L \left(\frac{\partial F}{\partial Q_l} \right) \left(\frac{\partial G}{\partial Q_l} \right) u^2(q_l) \end{aligned} \quad (1.88)$$

Nel caso particolare in cui tutti i coefficienti di correlazione $r(x_i, x_j)$ siano pari a +1 si ha

$$u_c^2(y) = \left[\sum_{i=1}^N c_i u(x_i) \right]^2 \neq \sum_{i=1}^N c_i^2 u^2(x_i) \quad (1.89)$$

1.9 Esempi

In questo paragrafo si riportano alcuni esempi per meglio chiarire i concetti esposti in precedenza.

1.9.1 Realizzazione di una resistenza serie

Consideriamo una resistenza serie R_{serie} realizzata mediante dieci resistenze R_i :

$$Y = R_{serie} = f(R_1, R_2, \dots, R_{10}) = \sum_{i=1}^{10} R_i \quad (1.90)$$

Supponiamo che il valore delle R_i sia stato determinato attraverso il confronto con una resistenza campione (attraverso ad esempio il metodo a ponte) e che il metodo di misura usato non fornisca un contributo significativo di incertezza. Pertanto $u(r_i)$ dipenderá soltanto dall'incertezza della resistenza campione R_c impiegata nella misura di taratura, ossia $u(r_i) = u(r_c)$ e se abbiamo un errore su R_c , l'errore viene riportato con lo stesso segno in tutte le operazioni di taratura. Pertanto

$$u(y) = u(r_{serie}) = 10u(r_c) \quad (1.91)$$

1.9.2 Misura di impedenza

Una misura di impedenza può essere effettuata valutando il modulo della tensione V , il modulo della corrente I e lo sfasamento Φ . Come risultato della misurazione si ricavano la parte reale dell'impedenza, quella immaginaria e il modulo:

$$R = \frac{V}{I} \cos \Phi \quad (1.92)$$

$$X = \frac{V}{I} \sin \Phi \quad (1.93)$$

$$Z = \frac{V}{I} \quad (1.94)$$

Supponiamo di aver effettuato cinque misurazioni dei valori V , I e Φ . I risultati sono riportati nella tabella 1.1.

Gruppo	V (V)	I (mA)	Φ (rad)
1	5.007	19.663	1.0456
2	4.994	19.639	1.0438
3	5.005	19.640	1.0468
4	4.990	19.685	1.0428
5	4.999	19.678	1.0433
Media aritmetica	$\bar{V} = 4.9990$	$\bar{I} = 19.6610$	$\bar{\Phi} = 1.04446$
scarto tipo sperimentale della media	$s(\bar{V}) = 0.0032$	$s(\bar{I}) = 0.0095$	$s(\bar{\Phi}) = 0.00075$
Coefficienti di correlazione			
$r(\bar{V}, \bar{I}) = -0.36$			
$r(\bar{V}, \bar{\Phi}) = 0.86$			
$r(\bar{I}, \bar{\Phi}) = -0.65$			

Tabella 1.1: Valori delle grandezze di ingresso V , I e Φ ottenuti in 5 gruppi di osservazioni simultanee.

L'obiettivo è calcolare l'incertezza sui valori R , X e Z . A tal fine è possibile utilizzare due diversi metodi.

Metodo 1

Si effettua una valutazione attraverso i valori medi, ossia

$$\bar{Y} = f(\bar{X}_1, \bar{X}_2, \dots, \bar{X}_N) \quad (1.95)$$

I valori sono riportati nella tabella 1.2.

Misurando	Stima del misurando	Valore della stima	Incertezza tipo composta
1	$Y_1 = R = (\bar{V}/\bar{I}) \cos \bar{\Phi}$	127.732 Ω	0.071 Ω
2	$Y_2 = X = (\bar{V}/\bar{I}) \sin \bar{\Phi}$	219.847 Ω	0.295 Ω
3	$Y_3 = Z = \bar{V}/\bar{I}$	254.260 Ω	0.236 Ω
Coefficienti di correlazione			
$r(Y_1, Y_2) = -0.588$			
$r(Y_1, Y_3) = -0.485$			
$r(Y_2, Y_3) = 0.993$			

Tabella 1.2: Valori calcolati delle grandezze di uscita R , X e Z - metodo 1.

Metodo 2

In questo metodo osservazione per osservazione si valutano i valori della Y . Per ogni osservazione si determinano i valori R , X e Z e si calcolano i valori Y_1 , Y_2 , Y_3 come media aritmetica delle osservazioni. I risultati sono riportati nella tabella 1.3.

Risultati

Confrontando i risultati dei due metodi, si può osservare come questi siano sostanzialmente identici. Questo fatto si può spiegare osservando che esiste uno scostamento attorno ai valori medi abbastanza contenuto, per cui la funzione f , che in genere non è lineare, può essere linearizzata senza introdurre errori significativi. Nel caso in cui si avesse a che fare con incertezze più ampie i due metodi potrebbero invece fornire risultati differenti.

Gruppo	$R = (V/I) \cos \Phi \ (\Omega)$	$X = (V/I) \sin \Phi \ (\Omega)$	$Z = (V/I) \ (\Omega)$
1	127.67	220.32	254.64
2	127.89	219.79	254.29
3	127.51	220.64	254.84
4	127.71	218.97	254.49
5	127.88	219.51	254.04
Media aritmetica	$Y_1 = \bar{R} = 127.732$	$Y_2 = \bar{X} = 219.847$	$Y_3 = \bar{Z} = 254.260$
scarto tipo della media	$s(\bar{R}) = 0.071$	$s(\bar{X}) = 0.295$	$s(\bar{Z}) = 0.236$
Coefficienti di correlazione			
$r(Y_1, Y_2) = r(\bar{R}, \bar{X}) = -0.588$			
$r(Y_1, Y_3) = r(\bar{R}, \bar{Z}) = -0.485$			
$r(Y_2, Y_3) = r(\bar{X}, \bar{Z}) = -0.993$			

Tabella 1.3: Valori delle grandezze di uscita R , X e Z - metodo 2.

1.9.3 Misura di una barretta metallica per confronto

Detta L la lunghezza della barretta abbiamo

$$L = l(1 + \alpha\theta) \quad (1.96)$$

dove l rappresenta la lunghezza alla temperatura di riferimento $T = 20^\circ\text{C}$, α é il coefficiente di dilatazione termica e θ é la differenza di temperatura rispetto a 20°C .

Si effettua la misurazione confrontando la lunghezza della barretta con una barretta di riferimento lungo $l_s(1 + \alpha_s\theta_s)$. In altre parole la misura avviene valutando la differenza tra le due grandezze

$$d = l(1 + \alpha\theta) - l_s(1 + \alpha_s\theta_s) \quad (1.97)$$

dove d é la grandezza osservata.

Si vuole determinare la grandezza l , ottenibile dalla formula

$$l = \frac{l_s(1 + \alpha_s\theta_s) + d}{1 + \alpha\theta} \quad (1.98)$$

E' possibile mettere in pratica una serie di strategie per cercare di semplificare il problema in esame (formalmente espresso dalla (1.98)). Effettuiamo prima di tutto l'approssimazione $\frac{1}{1+x} \simeq 1 - x$, ottenendo:

$$\begin{aligned} l &\simeq [l_s(1 + \alpha_s\theta_s) + d](1 - \alpha\theta) = \\ &= l_s(1 + \alpha_s\theta_s) + d - l_s(1 + \alpha_s\theta_s)\alpha\theta - d\alpha\theta \simeq \\ &\simeq l_s + d + l_s(\alpha_s\theta_s - \alpha\theta) \end{aligned} \quad (1.99)$$

dove l'ultima approssimazione consiste nel trascurare i termini

$$l_s\alpha_s\theta_s\alpha\theta \quad (1.100)$$

e

$$d\alpha\theta \quad (1.101)$$

Effettuiamo ora un cambio di variabili ponendo

$$\alpha = \alpha_s + \delta\alpha \quad (1.102)$$

$$\theta_s = \theta - \delta\theta \quad (1.103)$$

In tal modo l'equazione (1.99) può essere riscritta come

$$l \simeq l_s + d + l_s(\alpha_s\theta - \alpha_s\delta\theta - \alpha_s\theta - \delta\alpha\theta) = l_s + d - l_s(\alpha_s\delta\theta + \delta\alpha\theta) \quad (1.104)$$

In questo modo si evita di dover considerare la correlazione che può esistere tra θ e θ_s in quanto θ_s non compare più. Analogo ragionamento può essere effettuato per α e α_s . Con gli accorgimenti visti l risulta essere funzione di $d, l_s, \alpha_s, \theta, \delta\theta$ e $\delta\alpha$, ossia

$$l \simeq f(d, l_s, \alpha_s, \theta, \delta\theta, \delta\alpha) \quad (1.105)$$

In questo modo sono state eliminate le grandezze α e θ_s introducendo le differenze $\delta\theta$ e $\delta\alpha$, con il vantaggio che $\delta\theta$ e θ possono essere considerate come scorrelate, così come $\delta\alpha$ e α_s .

1.10 Combinazione e confronto dei risultati di misure separate

Molto spesso si conoscono molte misurazioni differenti di una stessa grandezza fisica, e sorge il problema di come poterle combinare tra loro per ottenere una singola migliore stima. Consideriamo per semplicità due diverse misurazioni X_A e X_B di una certa grandezza X con i propri scarti tipo s_A e s_B . Come detto il problema è trovare il modo migliore di combinare X_A e X_B per ottenere una singola miglior stima di X .

Per risolvere il problema si ricorre ancora una volta all'ipotesi che la distribuzione di probabilità delle grandezze in gioco sia gaussiana e al criterio di massima verosimiglianza.

La probabilità di osservare X_A risulta essere

$$P(X_A) \propto \frac{1}{s_A} \exp \left[-\frac{(X_A - \bar{X})^2}{2s_A^2} \right] \quad (1.106)$$

mentre la probabilità di osservare X_B è

$$P(X_B) \propto \frac{1}{s_B} \exp \left[-\frac{(X_B - \bar{X})^2}{2s_B^2} \right] \quad (1.107)$$

Se le misure X_A e X_B sono indipendenti, la probabilità di osservare X_A e X_B risulta pari al prodotto delle due probabilità (1.106) e (1.107), ossia

$$P(X_A, X_B) \propto \frac{1}{s_A} \frac{1}{s_B} \exp \left[-\frac{(X_A - \bar{X})^2}{2s_A^2} \right] \exp \left[-\frac{(X_B - \bar{X})^2}{2s_B^2} \right] \quad (1.108)$$

In base al criterio di massima verosimiglianza si sceglierà il valore \bar{X} che rende massima la quantità $P(X_A, X_B)$. In altri termini, posto

$$\chi^2 = \frac{(X_A - \bar{X})^2}{2s_A^2} + \frac{(X_B - \bar{X})^2}{2s_B^2} \quad (1.109)$$

si tratta di risolvere il seguente problema di ottimizzazione:

$$\min_{\bar{X}} \chi^2 \quad (1.110)$$

Per risolvere il problema calcoliamo la derivata rispetto a \bar{X} dell'equazione (1.109) e uguagliamola a zero:

$$\frac{\partial \chi^2}{\partial \bar{X}} = 0 \Rightarrow \frac{(X_A - \bar{X})^2}{s_A^2} + \frac{(X_B - \bar{X})^2}{s_B^2} = 0 \quad (1.111)$$

Svolgendo i calcoli si ottiene:

$$\bar{X} = \frac{\frac{X_A}{s_A^2} + \frac{X_B}{s_B^2}}{\frac{1}{s_A^2} + \frac{1}{s_B^2}} = \frac{w_A X_A + w_B X_B}{w_A + w_B} \quad (1.112)$$

dove $w_A = \frac{1}{s_A^2}$ e $w_B = \frac{1}{s_B^2}$.

Analizzando il risultato ottenuto si nota come gli scarti tipo svolgano una funzione di peso. Il valor medio \bar{X} é una media pesata delle due misurazioni, con pesi inversamente proporzionali agli scarti tipo stessi. Pertanto si dá un peso maggiore alla misura che possiede uno scarto tipo (e quindi una incertezza) piú piccola.

Il caso generale di n misurazioni di una stessa grandezza X può essere facilmente ricavato estendendo i risultati precedenti. In particolare risulta:

$$\bar{X} = \frac{\sum_{i=1}^n w_i X_i}{\sum_{i=1}^n w_i} \quad (1.113)$$

1.10.1 Calcolo dell'incertezza da associare al valor medio

Per calcolare l'incertezza da associare al valor medio \bar{X} determinato precedentemente si utilizza la relazione dell'incertezza composta.

Espresso \bar{X} come funzione delle misurazioni effettuate

$$\bar{X} = f(X_A, X_B, \dots) \quad (1.114)$$

é possibile valutare lo scarto tipo composto nel seguente modo:

$$\begin{aligned}
 s_X^2 &= \left(\frac{\partial f}{\partial X_A} \right)^2 s_A^2 + \left(\frac{\partial f}{\partial X_B} \right)^2 s_B^2 + \dots = \\
 &= \left(\frac{w_A}{\sum_{i=1}^n w_i} \right)^2 s_A^2 + \left(\frac{w_B}{\sum_{i=1}^n w_i} \right)^2 s_B^2 + \dots = \\
 &= \frac{w_A}{(\sum_{i=1}^n w_i)^2} + \frac{w_B}{(\sum_{i=1}^n w_i)^2} + \dots = \\
 &= \frac{1}{\sum_{i=1}^n w_i}
 \end{aligned} \tag{1.115}$$

In altri termini

$$s_X^2 = \frac{1}{\frac{1}{s_A^2} + \frac{1}{s_B^2} + \dots} \tag{1.116}$$

1.11 Compatibilitá tra due misure

Consideriamo in un primo momento il caso di una grandezza il cui valore, indicato con X_{noto} , sia conoscibile con certezza. Supponiamo di aver effettuato una osservazione sperimentale di tale grandezza e di aver misurato li valore X_A , caratterizzato da uno scarto tipo s_A . Il problema che si vuole affrontare in questo paragrafo é il seguente: la misura X_A é accettabile oppure no?

Per rispondere al quesito introduciamo una particolare quantitá t che prende il nome di *discrepanza*:

$$t = \frac{|X_A - X_{\text{noto}}|}{s_A} \tag{1.117}$$

Analizzando la relazione (1.117), si nota come la differenza tra il valore misurato e X_{noto} venga rapportata allo scarto tipo s_A della misurazione. Grandi scostamenti dal valore noto possono essere ritenuti accettabili qualora s_A sia una quantitá elevata, e piccoli scostamenti dal valore noto non accettabili qualora la misura sia caratterizzata da una incertezza ridotta.

Supponendo una distribuzione di probabilitá gaussiana, é possibile calcolare la probabilitá che la differenza tra X_A e X_{noto} sia entro l'intervallo $t \cdot s_A$ nel

seguinte modo:

$$\begin{aligned} \Pr(\text{entro } t \cdot s_A) &= \frac{1}{s_A \sqrt{2\pi}} \int_{X_{\text{noto}} - ts_A}^{X_{\text{noto}} + ts_A} \exp \left[-\frac{(X_A - X_{\text{noto}})^2}{2s_A^2} \right] dX_A \\ &= \frac{1}{\sqrt{2\pi}} \int_{-t}^t \exp \left(-\frac{z^2}{2} \right) dz \end{aligned} \quad (1.118)$$

dove

$$z = \frac{X_A - X_{\text{noto}}}{s_A} \quad (1.119)$$

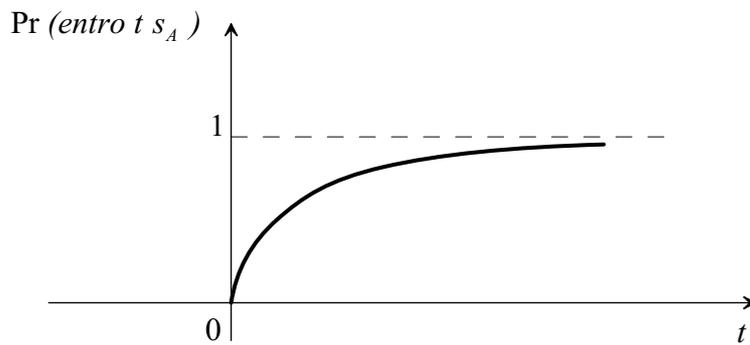


Figura 1.5: Rappresentazione grafica dell'equazione (1.118).

A partire da tale probabilità é possibile determinare la probabilità che la differenza tra X_A e X_{noto} sia entro l'intervallo $t \cdot s_A$ nel seguente modo:

$$\Pr(\text{fuori } t \cdot s_A) = 1 - \Pr(\text{entro } t \cdot s_A) \quad (1.120)$$

Si può dimostrare che

- per $t = 1$ si ha $\Pr(\text{fuori } t \cdot s_A) \approx 32\%$,
- per $t = 3$ si ha $\Pr(\text{fuori } t \cdot s_A) \approx 0.3\%$.

1.12 Accettabilità (confidenza) di due misure

Consideriamo una grandezza X di cui si siano effettuate due misurazioni X_A e X_B caratterizzate da scarti tipo rispettivamente pari a s_A e s_B . Queste misure sono accettabili?

Per rispondere alla domanda introduciamo la quantità

$$Y = X_A - X_B \quad (1.121)$$

Poiché X_A e X_B sono due misurazioni effettuate sulla stessa grandezza, esse fanno riferimento a una stessa distribuzione di probabilità, pertanto $\bar{Y} = 0$. Per quanto riguarda invece la varianza avremo $s_Y^2 = s_A^2 + s_B^2$. In analogia a quanto svolto nel paragrafo precedente, definiamo la discrepanza come

$$t = \frac{|X_A - X_B|}{s_Y} \quad (1.122)$$

Anche in questo caso é possibile determinare con quale probabilità si possono osservare valori della Y al di fuori dell'intervallo $t \cdot s_Y$ e valutare in tal modo l'accettabilità o meno delle misure considerate senza dover ricorrere a un valore di riferimento noto, come richiesto in precedenza.

1.12.1 Esempio

Si supponga di aver effettuato le due misurazioni seguenti della velocità di propagazione del suono in aria:

- $v_a = 334\text{m/s}$ con $s_a = 1\text{m/s}$,
- $v_b = 336\text{m/s}$ con $s_b = 2\text{m/s}$,

In questo caso si verifica facilmente che

$$t = \frac{|v_a - v_b|}{s_Y} = \frac{2}{\sqrt{1+4}} \simeq 0.89 \quad (1.123)$$

Possiamo affermare quindi che esiste una probabilità di circa il 30% (vedi figura 1.11) che le due misurazioni siano compatibili.

Se consideriamo invece questo altro gruppo di misurazioni

- $v_a = 334\text{m/s}$ con $s_a = 1\text{m/s}$,

- $v_b = 336\text{m/s}$ con $s_b = 5\text{m/s}$,

abbiamo

$$t = \frac{|v_a - v_b|}{s_Y} = \frac{2}{\sqrt{1 + 25}} \simeq 0.39 \quad (1.124)$$

Lo scarto sulla seconda misurazione é maggiore, la discrepanza é diminuita, quindi queste misure sono ancora piú accettabili delle precedenti.

1.12.2 Osservazione sulle grandezze in gioco

Fino a ora abbiamo trovato dei risultati nell'ipotesi che le distribuzioni delle grandezze in gioco siano gaussiane. Può sorgere spontanea la seguente domanda: é sempre possibile ipotizzare distribuzioni gaussiane? A tal proposito ci limitiamo a ricordare l'esistenza del *teorema del limite centrale*, secondo il quale in presenza di molti effetti tra loro non correlati, la distribuzione della misurazione assume proprio una forma gaussiana. Assumere una distribuzione gaussiana non significa ammettere l'esistenza di un fenomeno fisico strettamente gaussiano, ma significa far riferimento a una molteplicitá di fattori indipendenti che complessivamente producono una densitá di probabilitá gaussiana.

Capitolo 2

Distribuzioni di probabilità e modelli di rumore

In questo paragrafo vogliamo utilizzare alcune distribuzioni di probabilità per trattare due modelli di rumore, il *rumore termico* e il *rumore di granularità* (anche noto come *rumore shot*). A tal fine considereremo la distribuzione binomiale e in seguito le distribuzioni gaussiane e poissoniane.

2.1 Variabile di Bernoulli e distribuzione binomiale (*)

Consideriamo una sequenza di tentativi aleatori ripetuti. Questi tentativi prendono il nome di *eventi di Bernoulli* se:

- (1) a ogni tentativo può essere ottenuto uno di due risultati permessi. Esempi tipici sono il lancio di una moneta con due possibili risultati (testa o croce), o il lancio di un dado, dove definiamo, ad esempio, evento A un valore compreso tra uno e tre, ed evento B un valore compreso tra quattro e sei. In entrambi gli esempi i due eventi sono equiprobabili, ossia hanno probabilità pari a $\frac{1}{2}$;

(2) la probabilità di ognuno dei due risultati permessi non varia durante la serie di esperimenti.

La variabile aleatoria risultato dell'esperimento può assumere solamente due valori (diciamo A e B) e prende il nome di *variabile di Bernoulli*. Consideriamo n tentativi di Bernoulli, e chiediamoci quale sia la probabilità di ottenere ν risposte positive (ad esempio il verificarsi dell'evento A) in un ordine non specificato. Il numero di combinazioni possibili è uguale al numero di modi diversi in cui si possono scegliere le ν prove, tra le n totali, in cui abbia luogo l'evento A , ossia

$$\frac{n!}{\nu!(n-\nu)!} = \binom{n}{\nu} \quad (2.1)$$

Nel caso di eventi indipendenti, la probabilità di ogni combinazione è

$$p^\nu(1-p)^{n-\nu} \quad (2.2)$$

dove p è la probabilità elementare di una singola risposta positiva. Indicheremo nel seguito con $q = 1-p$ la probabilità di una singola risposta negativa. Pertanto la probabilità che n eventi di Bernoulli diano luogo a ν risposte positive è

$$b_{n,p}(\nu) = \binom{n}{\nu} p^\nu (1-p)^{n-\nu} \quad (2.3)$$

La relazione precedente prende il nome di *distribuzione binomiale*, ed è una distribuzione di probabilità discreta. Indicato con S_n il numero di successo in n prove, abbiamo

$$b_{n,p}(\nu) = Pr\{S_n = \nu\} \quad (2.4)$$

Dall'equazione (2.3) si ricava

$$\frac{b_{n,p}(\nu)}{b_{n,p}(\nu-1)} = \frac{(n-\nu+1)p}{\nu q} = 1 + \frac{(n+1)p-\nu}{\nu q} \quad (2.5)$$

Pertanto, il termine $b_{n,p}(\nu)$ è maggiore del precedente per $\nu < (n+1)p$ ed è più piccolo per $\nu > (n+1)p$. Se $(n+1)p = m$ è un numero intero, allora $b_{n,p}(m) = b_{n,p}(m-1)$. Perciò esiste un intero m tale che

$$(n+1)p - 1 < m \leq (n+1)p \quad (2.6)$$

e vale il seguente Teorema.

Teorema 2.1.1. *Per $0 \leq \nu \leq n$, il termine $b_{n,p}(\nu)$ prima cresce monotonicamente fino a raggiungere il valore massimo quando $\nu = m$ e poi decresce sempre in modo monotono. Quando $m = (n + 1)p$ si ha $b_{n,p}(m - 1) = b_{n,p}(m)$, ossia la distribuzione raggiunge il massimo per due distinti valori di μ .*

Il termine $b_{n,p}(m)$ prende il nome di *termine centrale*.

La distribuzione binomiale é *ben normalizzata*, ossia

$$\sum_{k=0}^n b_{n,p}(k) = 1 \quad (2.7)$$

Infatti la sommatoria $\sum_{k=0}^n b_{n,p}(k)$ é uguale alla somma dei termini dell'espansione di $[(1 - p) + p]^n$, che é proprio uguale a uno.

2.1.1 Caratteristiche della distribuzione binomiale (*)

Calcoliamo in primo luogo il valore atteso del numero di risposte positive in una serie di prove governate da una distribuzione binomiale, e la varianza associata. Per calcolare il valore atteso del numero di risposte positive dobbiamo sommare i possibili numeri di successi con il loro peso

$$E\{\nu\} = \bar{\nu} = \sum_{k=0}^n b_{n,p}(k)k \quad (2.8)$$

Quindi

$$\bar{\nu} = \sum_{k=0}^n \binom{n}{k} p^k (1 - p)^{n-k} k \quad (2.9)$$

Considerato che

$$\binom{n}{k} k = \frac{n!}{k!(n-k)!} k = n \frac{(n-1)!}{(k-1)!(n-k)!} = n \binom{n-1}{k-1} \quad (2.10)$$

si ottiene

$$\begin{aligned}
 \bar{\nu} &= \sum_{k=0}^n \binom{n}{k} p^k (1-p)^{n-k} k = \sum_{k=0}^n \binom{n-1}{k-1} p^{k-1} (1-p)^{n-k} np = \\
 &= \sum_{k-1=0}^{n-1} \binom{n-1}{k-1} p^{k-1} (1-p)^{n-1-(k-1)} np = \\
 &= \sum_{i=0}^m \binom{m}{i} p^i (1-p)^{m-i} np = \sum_{i=0}^m b_{m,p}(i) np
 \end{aligned} \tag{2.11}$$

Poiché la distribuzione é ben normalizzata

$$\sum_{i=0}^m b_{m,p}(i) = 1 \tag{2.12}$$

da cui si ricava

$$\bar{\nu} = np \tag{2.13}$$

Si può dimostrare che per una qualunque distribuzione vale la relazione

$$s_x^2 = \overline{x^2} - (\bar{x})^2 \tag{2.14}$$

Infatti

$$s_x^2 = \overline{(x - \bar{x})^2} = \overline{x^2 - 2x\bar{x} + (\bar{x})^2} = \overline{x^2} - 2\overline{x\bar{x}} + (\bar{x})^2 = \overline{x^2} - (\bar{x})^2 \tag{2.15}$$

Sfruttando la relazione (2.14), é possibile calcolare la varianza della distribuzione binomiale in questo modo:

$$s_\nu^2 = \sum_{k=0}^n k^2 b_{n,p}(k) - \bar{\nu}^2 = \left[\sum_{k=0}^n k^2 \binom{n}{k} p^k (1-p)^{n-k} \right] - (np)^2 \tag{2.16}$$

Ponendo $k^2 = k(k-1) + k$, si ottiene

$$\begin{aligned}
s_\nu^2 &= \left[\sum_{k=2}^n k(k-1) \binom{n}{k} p^k (1-p)^{n-k} + \sum_{k=1}^n k \binom{n}{k} p^k (1-p)^{n-k} \right] - (np)^2 = \\
&= \sum_{k=2}^n k(k-1) \binom{n}{k} p^k (1-p)^{n-k} + np - (np)^2 = \\
&= \sum_{k=2}^n k(k-1) \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} + np - (np)^2 = \\
&= \sum_{k=2}^{n-2} k(k-1) \frac{n(n-1)}{k(k-1)} p^2 \frac{(n-2)!}{(k-2)![n-2-(k-2)]!} p^{k-2} (1-p)^{[n-2-(k-2)]} + \\
&+ np - (np)^2 = n(n-1)p^2 \sum_{i=0}^m \frac{m!}{i!(m-i)!} p^i (1-p)^{m-i} + np - (np)^2
\end{aligned} \tag{2.17}$$

Poiché la distribuzione é ben normalizzata

$$\sum_{i=0}^m b_{m,p}(i) = 1 \tag{2.18}$$

si ricava

$$s_\nu^2 = n(n-1)p^2 + np - (np)^2 = -np^2 + np = np(1-p) \tag{2.19}$$

da cui

$$s_\nu^2 = np(1-p) \tag{2.20}$$

Se $p = \frac{1}{2}$, il numero medio di successi é proprio $\frac{n}{2}$, inoltre la distribuzione é simmetrica

$$b_{n,\frac{1}{2}}(\nu) = b_{n,\frac{1}{2}}(n-\nu) \tag{2.21}$$

2.2 Distribuzione di Gauss

2.2.1 La gaussiana come approssimazione della binomiale

Per valori di n sufficientemente grandi, la distribuzione binomiale puó essere approssimata da una distribuzione gaussiana avente stesso valor medio e stessa

varianza, come si può vedere dalla figura 2.1. Il teorema di De Moivre–Laplace (si veda al riguardo l'appendice D) permette di dimostrare questa affermazione.

Il fatto che la distribuzione binomiale possa essere approssimata dalla funzione di Gauss quando n è grande è molto utile in pratica. Infatti il calcolo della distribuzione binomiale con n maggiore di venti è molto pesante dal punto di vista computazionale a causa della presenza dei fattoriali (che potrebbero originare facilmente degli *overflow* durante il calcolo), mentre il calcolo della funzione di Gauss è molto semplice, qualunque siano il valor medio e la varianza.

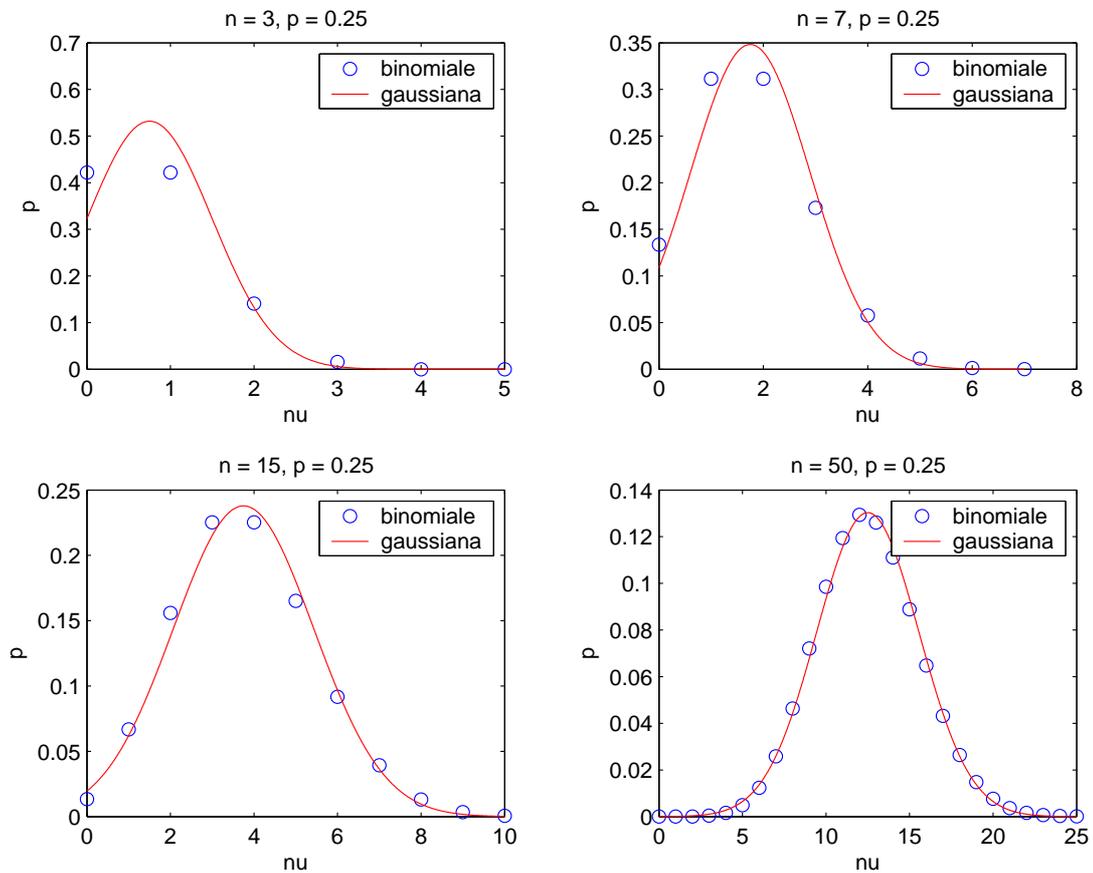


Figura 2.1: Andamento delle distribuzioni binomiale e gaussiana al variare di n .

Si noti come

- (1) all'aumentare di n e di p , la distribuzione si sposta a valori sempre più elevati;

- (2) all'aumentare di n la distribuzione diventa relativamente piú stretta (se la larghezza viene rapportata al valore centrale);
- (3) all'aumentare di n la distribuzione si infittisce e, per valori di p distanti da 0 e da 1, acquista una forma regolare "a campana".

In particolare quando $p = \frac{1}{2}$ la distribuzione binomiale é simmetrica.

2.2.2 Modello degli errori casuali

Supponiamo di voler misurare una grandezza X . Consideriamo un processo di misura in cui siano trascurabili eventuali errori sistematici, ma in cui siano presenti n sorgenti indipendenti di errori casuali. Supponiamo che queste sorgenti producano errori casuali della stessa entità ϵ , e segno diverso con eguale probabilità, $p = \frac{1}{2}$.

Il risultato della misura X_ν può essere espresso in funzione dei contributi delle sorgenti di errore positivo (ν) e dei restanti contributi con valore negativo (ν) in questo modo:

$$X_\nu = \bar{X} + \nu\epsilon - (n - \nu)\epsilon \quad (2.22)$$

La probabilità di avere ν errori positivi su n sorgenti di errore, avendo supposto rumori indipendenti é data dalla distribuzione binomiale

$$b_{n, \frac{1}{2}}(\nu) \quad (2.23)$$

All'aumentare di n si ottengono distribuzioni il cui andamento é simile a quello riportato nella figura 2.1. In questo caso le distribuzioni sono simmetriche perché $p = \frac{1}{2}$. Ripetendo quanto già accennato precedentemente, la binomiale si discosta notevolmente da una gaussiana quando n é piccolo, mentre quando il numero di sorgenti di errore n cresce, la distribuzione binomiale é ben approssimata da una gaussiana. Questo risultato conferma il *teorema del limite centrale*, secondo cui quando si ha a che fare con molte sorgenti di errore indipendenti, la misurazione assume una distribuzione di tipo gaussiano.

Nel caso in esame, poiché $p = \frac{1}{2}$, avremo

$$\begin{aligned}\bar{\nu} &= pn = \frac{n}{2} \\ s_{\nu}^2 &= p(1-p)n = \frac{n}{4}\end{aligned}\tag{2.24}$$

da cui

$$\begin{aligned}\bar{X} &= \bar{X}_{\nu} \\ s_x^2 &= 4\epsilon^2 s_{\nu}^2 = \epsilon^2 n\end{aligned}\tag{2.25}$$

2.3 Distribuzione di Poisson

La distribuzione di Poisson riguarda eventi casuali in un fenomeno in cui è definita una media temporale.

Esempi di fenomeni reali modellabili tramite una distribuzione di Poisson sono la corrente di particelle, i fenomeni di emissione radioattiva, il traffico dei veicoli e altri ancora.

Consideriamo come esempio il fenomeno del decadimento di un materiale radioattivo. Ogni nucleo radioattivo ha una probabilità definita di decadere in ogni intervallo di un minuto. Supponendo di avere a disposizione n nuclei di materiale radioattivo, e p rappresenta la probabilità che ogni nucleo decada, la probabilità di ν decadimenti in n prove è $b_{n,p}(\nu)$, ossia la distribuzione binomiale introdotta in precedenza. In realtà per il tipo di esperimento considerato, è possibile fare una notevole semplificazione. Il numero di prove (cioè nuclei) è di solito molto grande ($n \simeq 10^{20}$) e la probabilità di successo (cioè di decadimento) per ogni nucleo è molto piccola (spesso dell'ordine di $p \simeq 10^{-20}$). Sotto queste ipotesi si può dimostrare che la distribuzione binomiale è ben approssimata da una funzione più semplice, chiamata distribuzione di Poisson

$$p_{\mu}(\nu) = e^{-\mu} \frac{\mu^{\nu}}{\nu!}\tag{2.26}$$

2.3.1 Caratteristiche della distribuzione di Poisson

Calcoliamo dapprima il valor medio di questa distribuzione. Abbiamo

$$\bar{\nu} = \sum_{\nu=0}^{\infty} \nu p_{\mu}(\nu) = \sum_{\nu=0}^{\infty} e^{-\mu} \frac{\mu^{\nu}}{\nu!} \quad (2.27)$$

Il primo termine della sommatoria può essere eliminato (in quanto nullo). Inoltre sostituendo $\frac{\nu}{\nu!}$ con $\frac{1}{(\nu-1)!}$ e raccogliendo il fattore $\mu e^{-\mu}$ si ottiene

$$\bar{\nu} = \mu e^{-\mu} \sum_{\nu=1}^{\infty} \frac{\mu^{\nu-1}}{(\nu-1)!} = \mu e^{-\mu} \left[1 + \mu + \frac{\mu^2}{2!} + \frac{\mu^3}{3!} + \dots \right] = \mu e^{-\mu} e^{\mu} = \mu \quad (2.28)$$

Pertanto il parametro μ di una distribuzione di Poisson ne rappresenta il valor medio.

Per quanto riguarda la varianza, (sfruttando anche la relazione (2.14)) abbiamo

$$\begin{aligned} s_{\nu}^2 &= \sum_{\nu=0}^n (\nu - \bar{\nu})^2 p_{\mu}(\nu) = \sum_{\nu=0}^n \nu^2 p_{\mu}(\nu) - \bar{\nu}^2 = \sum_{\nu=1}^n \nu^2 e^{-\mu} \frac{\mu^{\nu}}{\nu!} - \mu^2 = \\ &= \mu e^{-\mu} \left[\sum_{\nu=1}^n \left[\nu \frac{\mu^{\nu-1}}{(\nu-1)!} - \frac{\mu^{\nu-1}}{(\nu-1)!} + \frac{\mu^{\nu-1}}{(\nu-1)!} \right] \right] - \mu^2 = \\ &= \mu e^{-\mu} \left[\sum_{\nu=1}^n \frac{\mu^{\nu-1}}{(\nu-1)!} + \sum_{\nu=1}^n (\nu-1) \frac{\mu^{\nu-1}}{(\nu-1)!} \right] - \mu^2 = \\ &= \mu e^{-\mu} \left[e^{\mu} + \mu \sum_{\nu=2}^n \frac{\mu^{\nu-2}}{(\nu-2)!} \right] - \mu^2 = \\ &= \mu e^{-\mu} [e^{\mu} + \mu e^{\mu}] - \mu^2 = \mu + \mu^2 - \mu^2 = \mu \end{aligned} \quad (2.29)$$

In sintesi sia il valor medio che la varianza della distribuzione di Poisson valgono μ .

2.3.2 La poissoniana come approssimazione della binomiale

Consideriamo la distribuzione binomiale $b_{n,p}(\nu)$, e supponiamo di avere n abbastanza grande e p abbastanza piccolo. In particolare scegliamo n e p in

modo che la media temporale

$$\bar{\nu} = pn = \mu \quad (2.30)$$

sia un quantità finita. Vogliamo dimostrare come, sotto queste ipotesi, la distribuzione di Poisson possa essere utilizzata per approssimare la distribuzione binomiale.

Partiamo dalla distribuzione binomiale

$$b_{n,p}(\nu) = \binom{n}{\nu} p^\nu (1-p)^{n-\nu} = \frac{n(n-1)\cdots(n-\nu+1)}{\nu!} p^\nu (1-p)^{n-\nu} \quad (2.31)$$

Per n elevato possiamo approssimare ogni fattore proprio con n , in quanto i termini $-1, -2, \dots, -\nu + 1$ sono trascurabili rispetto a n :

$$\begin{aligned} n-1 &\simeq n \\ &\vdots \\ n-\nu+1 &\simeq n \end{aligned} \quad (2.32)$$

Poiché sono presenti ν fattori, la quantità al numeratore $n(n-1)\cdots(n-\nu+1)$ è approssimabile con n^ν . Otteniamo quindi

$$b_{n,p}(\nu) = \binom{n}{\nu} p^\nu (1-p)^{n-\nu} \simeq \frac{(pn)^\nu}{\nu!} (1-p)^{n-\nu} \quad (2.33)$$

Consideriamo ora il termine $(1-p)^{n-\nu}$. Per n elevato abbiamo

$$(1-p)^{n-\nu} \simeq (1-p)^n \quad (2.34)$$

Utilizzando l'espansione di Taylor, si ricava

$$(1-p)^n = 1 - np + \frac{1}{2}n(n-1)p^2 + \cdots \simeq 1 - pn + \frac{1}{2}(pn)^2 + \cdots = e^{-pn} \quad (2.35)$$

Sostituite le approssimazioni svolte nella (2.31) e posto $\mu = pn$, si ottiene

$$b_{n,p}(\nu) \simeq \frac{\mu^\nu e^{-\mu}}{\nu!} \quad (2.36)$$

ossia la distribuzione binomiale è approssimata da una distribuzione di Poisson.

La relazione (2.36) potrebbe non essere adeguata per valori di ν prossimi a n . Infatti se $\nu \simeq n$ non è più lecita l'approssimazione $n - \nu + 1 \simeq n$. Tuttavia il caso in cui $\nu \simeq n$ ha una probabilità bassa, e pertanto è possibile accettare una approssimazione più grossolana.

Come una distribuzione binomiale al crescere di n può essere approssimata con una distribuzione gaussiana, anche la distribuzione di Poisson al crescere di μ si avvicina a una gaussiana, come mostrato nella figura 2.2.

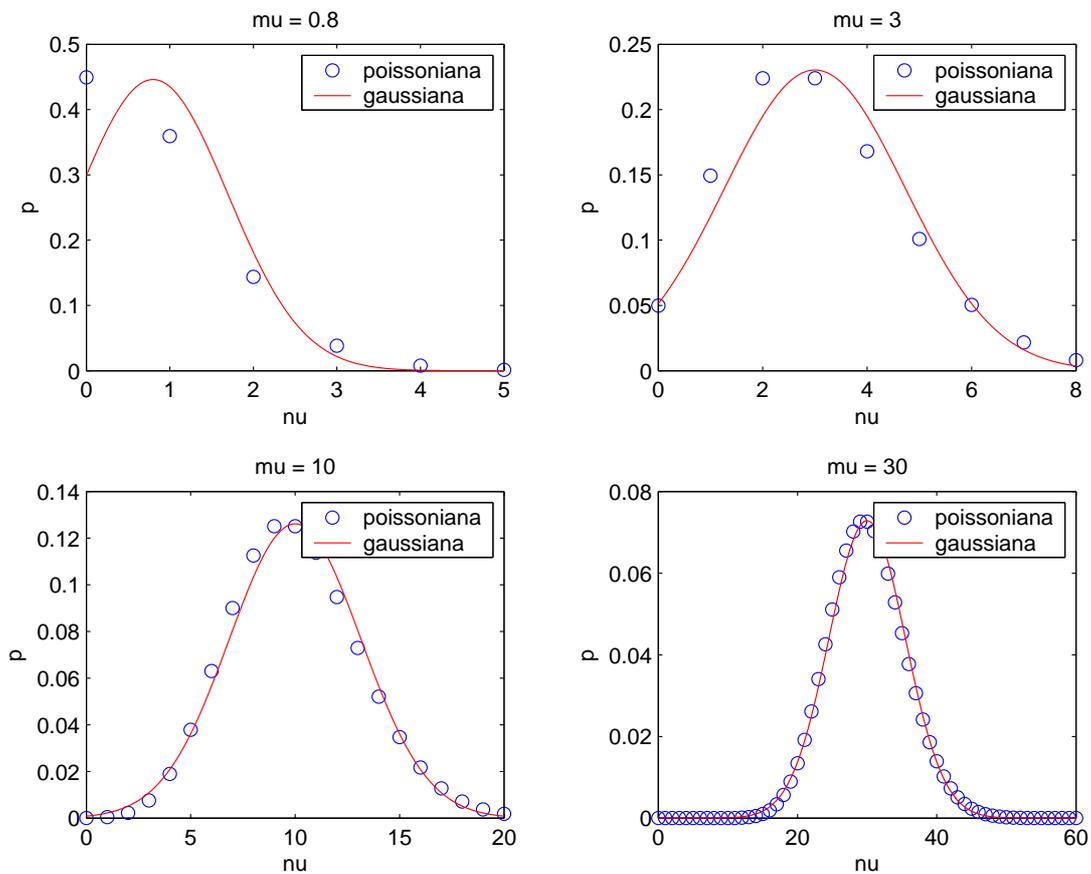


Figura 2.2: Andamento delle distribuzioni poissoniana e gaussiana al variare di μ .

La poissoniana differisce dalla gaussiana per valori bassi di μ . Per valori di μ crescenti le distribuzioni sono quasi coincidenti. La distribuzione di Gauss è continua, mentre la distribuzione di Poisson è una distribuzione discreta in

quanto ν può assumere solo valori interi. Inoltre come mostrato nelle relazioni (2.28) e (2.29) il parametro μ di una distribuzione di Poisson fissa sia il valor medio che la varianza della distribuzione, mentre in una gaussiana tali valori sono indipendenti.

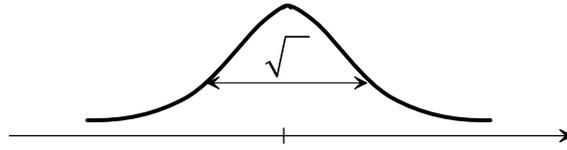


Figura 2.3: La larghezza di una distribuzione poissoniana é legata al valor medio μ .

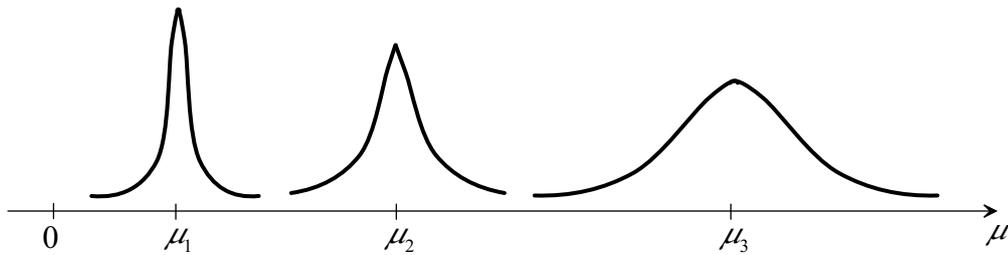


Figura 2.4: Forma delle distribuzioni di Poisson all'aumentare del parametro μ . Le distribuzioni sono tanto piú larghe quanto piú é grande μ .

La figura 2.4 mostra come al crescere del valore di μ la distribuzione poissoniana tenda ad allargarsi (per valori di μ molto piccoli la distribuzione non é neppure simmetrica).

Esempio. Consideriamo come semplice esempio l'osservazione del traffico stradale, ipotizzando che l'arrivo di veicoli a un semaforo sia di tipo casuale e abbia un tasso pari a m . In altre parole, m rappresenta il numero medio dei veicoli che arrivano al semaforo nell'unitá di tempo. Fissiamo un intervallo ΔT di osservazione del flusso di veicoli, risulterà

$$\mu = m\Delta T \quad (2.37)$$

Commenti. Nel caso si osservi il fenomeno per una finestra temporale ridotta, ΔT sarà piccolo, e di conseguenza anche μ sarà piccolo. La distribuzione che si ottiene in questo caso é una distribuzione asimmetrica. Aumentando l'intervallo di osservazione μ aumenta e la distribuzione dei veicoli assomiglia sempre di piú a una gaussiana. Osservare per un intervallo di tempo maggiore significa avere una incertezza $\sqrt{\mu}$ piú grande. Tuttavia l'incertezza relativa

$$\frac{s_\nu}{\bar{\nu}} = \frac{\sqrt{\mu}}{\mu} = \frac{1}{\sqrt{\mu}} \quad (2.38)$$

diminuisce al crescere dell'intervallo di osservazione.

2.4 Il rumore nel dominio delle frequenze e del tempo

Il rumore é la piú importante sorgente di incertezza nella strumentazione elettronica. Si definisce rumore (*noise*) qualsiasi segnale estraneo a quello di interesse per la misura e interferente con questo. In generale un rumore $n(t)$ si presenta come un processo aleatorio, e quindi, osservato a un certo istante t_1 fissato, costituisce una variabile aleatoria. Trattandosi di una variabile aleatoria possiamo definirne la statistica, in particolare valore medio, valore quadratico medio, varianza e funzione di autocorrelazione.

Supponendo di aver effettuato N osservazioni $n_i(t_1)$ del rumore all'istante t_1 , il *valore medio* é dato da

$$\bar{n}(t_1) = \lim_{N \rightarrow +\infty} \frac{1}{N} \sum_{i=1}^N n_i(t_1) \quad (2.39)$$

Nel caso in cui

$$\bar{n}(t_1) = \bar{n}(t_2) \quad (2.40)$$

allora il rumore si dice *stazionario*, e la media é costante nel tempo.

Definiamo il *valore quadratico medio* come

$$\overline{n^2}(t_1) = \lim_{N \rightarrow +\infty} \frac{1}{N} \sum_{i=1}^N n_i^2(t_1) \quad (2.41)$$

e la *varianza* come

$$\sigma^2(t_1) = \lim_{N \rightarrow +\infty} \frac{1}{N} \sum_{i=1}^N [n_i(t_1) - \overline{n}(t_1)]^2 \quad (2.42)$$

Se il rumore é stazionario, anche valore quadratico medio e varianza sono costanti nel tempo.

Si definisce *funzione di autocorrelazione* la quantità

$$R_n(t_1, t_2) = E\{n(t_1)n(t_2)\} \quad (2.43)$$

Se il processo aleatorio é stazionario la funzione di autocorrelazione dipende solo dalla differenza $t_2 - t_1 = \tau$, ossia

$$R_n(t_1, t_2) = R_n(t_2 - t_1) = R_n(\tau) \quad (2.44)$$

Un rumore stazionario si dice *ergodico nella media*, quando il valore medio, valutato a partire da un istante t_1 , eguaglia la media temporale, qualunque sia l'istante t_1 , ossia

$$\overline{n} = \overline{n}_t \quad (2.45)$$

dove \overline{n} é il valor medio

$$\overline{n} = \lim_{N \rightarrow +\infty} \frac{1}{N} \sum_{i=1}^N n_i(t_1) \quad (2.46)$$

e \overline{n}_t é la media temporale

$$\overline{n}_t = \lim_{T \rightarrow +\infty} \frac{1}{T} \int_{-\frac{T}{2}}^{+\frac{T}{2}} n(t) dt \quad (2.47)$$

Un processo stazionario si definisce *ergodico nel valore quadratico medio* quando

$$\overline{n^2} = \overline{n}_t^2 \quad (2.48)$$

dove

$$\overline{n_t^2} = \lim_{T \rightarrow +\infty} \frac{1}{T} \int_{-\frac{T}{2}}^{+\frac{T}{2}} n^2(t) dt \quad (2.49)$$

rappresenta il valore quadratico medio temporale. Analogamente un processo stazionario si definisce *ergodico nella varianza* quando

$$\sigma^2 = \sigma_t^2 \quad (2.50)$$

dove

$$\sigma_t^2 = \lim_{T \rightarrow +\infty} \frac{1}{T} \int_{-\frac{T}{2}}^{+\frac{T}{2}} [n(t) - \bar{n}]^2 dt \quad (2.51)$$

rappresenta la varianza temporale.

In modo del tutto analogo ai casi precedenti, definiamo *ergodico nell'autocorrelazione*, un processo stazionario per cui vale la relazione

$$R_n(\tau) = \lim_{T \rightarrow +\infty} \frac{1}{T} \int_{-\frac{T}{2}}^{+\frac{T}{2}} n(t)n(t+\tau) dt \quad (2.52)$$

dove il limite rappresenta l'autocorrelazione temporale di una data realizzazione del processo aleatorio, e $R_n(\tau)$ é l'autocorrelazione statistica definita come

$$R_n(\tau) = E\{n(t)n(t+\tau)\} \quad (2.53)$$

La trattazione dei segnali nasce prima dello studio del rumore elettrico, scoperto in seguito al trattamento e amplificazione dei segnali nelle reti elettriche. Il fatto che il rumore elettrico sia stato osservato in questo contesto ha portato come conseguenza naturale il tentativo di inquadrare anche questi processi nello schema matematico del trattamento dei segnali, per i quali l'approccio nel dominio delle frequenze (serie e trasformate di Fourier) si era dimostrato particolarmente utile.

Consideriamo dunque la densità spettrale di potenza, che fornisce la distribuzione del valore quadratico medio del rumore in funzione della frequenza. Il teorema di *Wiener-Khintchine* permette di affermare che lo spettro di densità di

potenza di un processo aleatorio stazionario é dato dalla trasformata di Fourier della funzione di autocorrelazione

$$G_n(f) = \mathcal{F}\{R_n(\tau)\} \quad (2.54)$$

Lo spettro di densità di potenza permette di conoscere la distribuzione media della potenza del processo sull'asse delle frequenze. La potenza delle componenti frequenziali del processo aleatorio comprese tra f_1 ed f_2 può essere calcolata come

$$\int_{-f_2}^{-f_1} G_n(f)df + \int_{f_1}^{f_2} G_n(f)df = 2 \int_{f_1}^{f_2} G_n(f)df \quad (2.55)$$

in quanto la densità spettrale di potenza é una funzione a simmetria pari.

2.5 Tipologie di rumore

Nei circuiti elettronici esistono cinque sorgenti fondamentali di rumore:

- (1) *shot noise*: viene generato quando le cariche elettriche attraversano le barriere di potenziale presenti nei diodi e nei transistori. Tale attraversamento é un evento puramente casuale, dunque la corrente che viene osservata può essere considerata come l'effetto macroscopico di impulsi casuali di corrente;
- (2) *thermal noise*: é proprio dei dispositivi attivi a semiconduttore; viene stimato come la media geometrica dalla tensione termica e della tensione sviluppata nel FET dalla carica di un singolo elettrone;
- (3) *flicker noise*: é dovuto nei dispositivi attivi a trappole che catturano e rilasciano in modo casuale portatori di carica, causando fluttuazioni nel flusso di corrente di natura casuale;
- (4) *burst noise*: é detto anche *popcorn*; sembra essere dovuto alla contaminazione del semiconduttore da parte dei metalli pesanti e alle condizioni non perfette sulla superficie dei semiconduttori durante i processi tecnologici sul wafer;

- (5) *avalanche noise* (effetto valanga): si presenta in giunzioni pn che operano nella zona di breakdown; la generazione per effetto valanga di coppie elettroni-lacune é casuale e si manifesta con impulsi di corrente attraverso la giunzione inversa.

Nel seguito ci occuperemo esclusivamente del rumore shot e del rumore termico.

2.5.1 Rumore shot

Lo shot noise é una corrente di rumore sempre associata con il passaggio della corrente in transistor, diodi o tubi a vuoto. Esso é prodotto ogni qualvolta delle cariche elettriche passano attraverso una barriera di potenziale, come in una giunzione pn . Non esiste shot noise nei conduttori semplici.

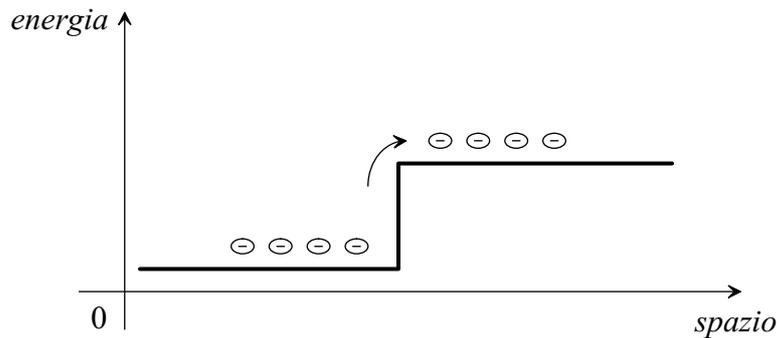


Figura 2.5: Rappresentazione schematica della barriera di energia che gli elettroni devono oltrepassare.

Attraversare una barriera di potenziale (vedi figura 2.5) é un evento puramente casuale: si può modellare l'emissione di cariche elettriche con un processo di Poisson.

Ipotizziamo che gli eventi di attraversamento della barriera di potenziale siano mutuamente indipendenti, dato che non c'è nessuna ragione per la quale dovrebbe essere vero il contrario, e ipotizziamo che le variabili aleatorie che costituiscono questo processo siano identicamente distribuite in modo esponenziale, con

parametro μ . Definiamo dunque il processo stocastico del rumore shot come

$$s(t) = q \sum_i h(t - t_i) \quad (2.56)$$

dove q é la carica del singolo portatore, le quantità t_i sono variabili aleatorie estratte da un processo poissoniano di parametro μ , e $h(t)$ é una funzione reale che descrive l'andamento temporale della corrente in corrispondenza del passaggio di una carica attraverso la barriera.

Nel caso in cui consideriamo un passaggio di cariche istantaneo, la funzione $h(t)$ corrisponde a un impulso. Se consideriamo una situazione piú realistica di tempo non istantaneo, possiamo modellare la funzione $h(t)$ nel seguente modo

$$h(t) = \begin{cases} \alpha t & 0 \leq t \leq T \\ 0 & \text{altrove} \end{cases} \quad (2.57)$$

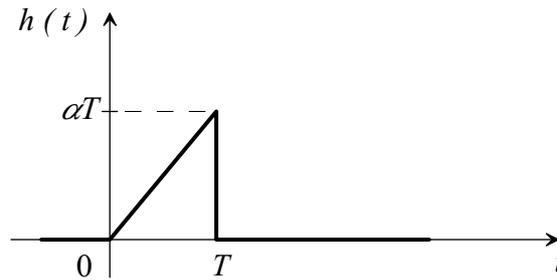


Figura 2.6: Funzione $h(t)$ nel caso in cui il passaggio di elettroni attraverso la barriera non sia istantaneo, ma di durata T (tempo di transito).

Tuttavia é possibile semplificare il problema trattando come infinitesimo il tempo di transito, e quindi rappresentare $h(t)$ come un treno di impulsi centrato sugli istanti casuali t_i definiti prima, ossia

$$s(t) = q \sum_{i=1}^m \delta(t - t_i) \quad (2.58)$$

Per un processo ergodico nell'autocorrelazione, si ha

$$R(\tau) = \lim_{T \rightarrow +\infty} q^2 \frac{1}{T} \int_{-\infty}^{+\infty} \sum_{i=1}^m \delta(t - t_i) \sum_{j=1}^m \delta(t - t_j + \tau) dt \quad (2.59)$$

Poiché

$$\sum_{i=1}^m \delta(t-t_i) \sum_{j=1}^m \delta(t-t_j+\tau) = \sum_{i=1}^m \delta(t-t_i)\delta(t-t_i+\tau) + \sum_{i \neq j} \delta(t-t_i)\delta(t-t_j+\tau) \quad (2.60)$$

si ricava

$$R(\tau) = \lim_{T \rightarrow +\infty} q^2 \frac{1}{T} \int_{-\infty}^{+\infty} \left[\sum_{i=1}^m \delta(t-t_i)\delta(t-t_i+\tau) + \sum_{i \neq j} \delta(t-t_i)\delta(t-t_j+\tau) \right] dt \quad (2.61)$$

Trascurando il valor medio

$$\sum_{i \neq j} \delta(t-t_i)\delta(t-t_j+\tau) \quad (2.62)$$

la precedente relazione diventa

$$R(\tau) = \lim_{T \rightarrow +\infty} q^2 \frac{1}{T} \int_{-\infty}^{+\infty} \sum_{i=1}^m \delta(t-t_i)\delta(t-t_i+\tau) dt \quad (2.63)$$

Invertendo il segno di sommatoria con quello di integrale, si ottiene

$$R(\tau) = \lim_{T \rightarrow +\infty} q^2 \frac{1}{T} \sum_{i=1}^m \int_{-\infty}^{+\infty} \delta(t-t_i)\delta(t-t_i+\tau) dt \quad (2.64)$$

Sfruttando la nota proprietà dell'impulso di Dirac

$$\int_{-\infty}^{+\infty} \delta(\theta)\delta(\theta+\tau) d\theta = \delta(\tau) \quad (2.65)$$

si ricava

$$R(\tau) = \lim_{T \rightarrow +\infty} q^2 \frac{1}{T} \sum_{i=1}^m \int_{-\infty}^{+\infty} \delta(t-t_i)\delta(t-t_i+\tau) dt = q^2 \frac{1}{T} \sum_{i=1}^m \delta(\tau) = q^2 \frac{m}{T} \delta(\tau) \quad (2.66)$$

dove $\frac{m}{T} = \mu$ rappresenta la frequenza di arrivo degli impulsi. Sostituendo nella precedente relazione si ottiene

$$R(\tau) = q^2 \mu \delta(\tau) \quad (2.67)$$

Il passaggio degli elettroni non ha un carattere uniforme, ma presenta una variabile descritta dal parametro della distribuzione di Poisson. Si ricava dunque lo spettro di densità di potenza

$$G(f) = \mathcal{F}\{R(\tau)\} = qI \quad (2.68)$$

dove $I = \mu q$.

Come si evince facilmente dalla relazione precedente, se ipotizziamo un passaggio istantaneo delle cariche, la corrente avrà un andamento nel tempo di tipo impulsivo, e sarà caratterizzata lo spettro di densità di potenza costante al variare della frequenza.

2.5.2 Rumore termico

Il rumore termico (thermal noise), è prodotto dall'agitazione termica dei portatori di carica (elettroni o lacune) in un conduttore. Si tratta di un rumore presente in tutti gli elementi resistivi passivi a una temperatura maggiore dello zero assoluto. A differenza dello shot noise è indipendente dal passaggio di corrente. Il rumore termico è stato modellato da Nyquist nel 1928, in relazione agli esperimenti di Johnson dello stesso anno, che ne evidenziò le caratteristiche. Il ragionamento di Nyquist è basato su una considerazione di tipo termodinamico e di meccanica statistica. Consideriamo un circuito costituito da due resistori R_I e R_{II} dello stesso valore di resistenza ($R_I = R_2 = R$). Nel circuito non sono presenti generatori indipendenti, e la temperatura dei due resistori è la stessa.

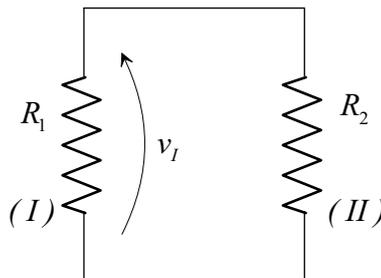


Figura 2.7: Circuito costituito da due resistenze uguali per la spiegazione del rumore termico.

Per effetto dell'agitazione termica ai capi dei due resistori si forma una differenza di potenziale, che dá luogo al passaggio di una corrente I e a un trasferimento di potenza da un resistore all'altro. Il trasferimento di potenza dal

resistore R_I al resistore R_{II} é espresso dalla relazione

$$\left[\frac{v_I(t)}{2R} \right]^2 R_{II} \quad (2.69)$$

dove $v_I(t)$ rappresenta la tensione presente ai capi del resistore R_I dovuta all'agitazione termica. In modo del tutto analogo possiamo definire il trasferimento di potenza dal resistore R_{II} al resistore R_I

$$\left[\frac{v_{II}(t)}{2R} \right]^2 R_I \quad (2.70)$$

Per il secondo principio della termodinamica non c'è scambio netto di energia per corpi che hanno la stessa temperatura. Quindi esiste un equilibrio tra le potenze scambiate tra i due resistori, che vale ad ogni frequenza. Se così non fosse, sarebbe possibile violare il principio della termodinamica inserendo un opportuno circuito risonante (puramente reattivo) tra le due resistenze (vedi figura 2.8): tale circuito reattivo annullando lo scambio termico alla frequenza di risonanza $\omega = \frac{1}{\sqrt{LC}}$ (supposto squilibrato), renderebbe squilibrato lo scambio termico complessivo.

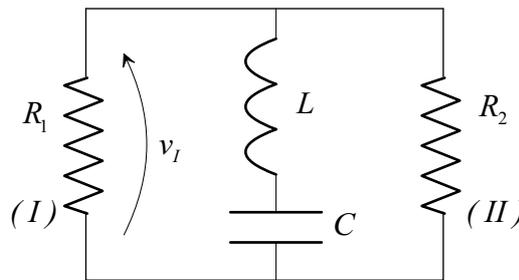


Figura 2.8: Inserimento di un circuito risonante nel circuito costituito da due resistenze uguali.

L'equilibrio vale dunque frequenza per frequenza. Fatte queste considerazioni preliminari, vediamo ora di ricavare la densità spettrale di potenza, esprimendola in termini di resistenza, temperatura e frequenza.

Spettro di densità di potenza del rumore termico secondo Nyquist

Vediamo ora di ricavare l'espressione dello spettro di densità di potenza del rumore termico. Per fare questo colleghiamo le due resistenze R_I e R_{II} , entrambe alla stessa temperatura T , con una linea di trasmissione priva di perdite, anch'essa a temperatura T , lunga l , avente una impedenza caratteristica pari a R . Essendo ideale, la linea di trasmissione non modifica gli scambi energetici tra i due resistori.

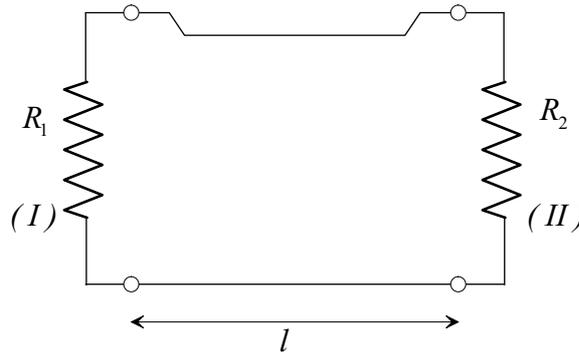


Figura 2.9: Inserimento di una linea di trasmissione ideale tra le due resistenze.

Chiamiamo u la velocità con la quale le onde elettromagnetiche si propagano nella linea. Poiché abbiamo supposto che la linea non sia dispersiva, la velocità di propagazione è la stessa per tutte le frequenze. Valutiamo la potenza trasmessa, supponendo che l'energia immagazzinata nella linea non vari se si cortocircuitano gli estremi, studiando i modi di oscillazione della linea cortocircuitata (vedi figura 2.10).

I modi di oscillazione avranno frequenze multiple del modo dominante

$$f_0 = \frac{u}{2l} \quad (2.71)$$

I modi di ordine superiore saranno

$$\begin{aligned} f_1 &= 2\frac{u}{2l} \\ f_2 &= 3\frac{u}{2l} \end{aligned} \quad (2.72)$$

e così via al crescere della frequenza.

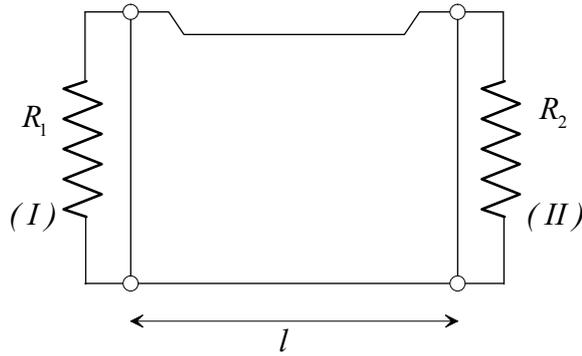


Figura 2.10: Linea di trasmissione ideale tra le due resistenze cortocircuitata agli estremi.

Consideriamo ora un intervallo Δf di frequenza. Il numero di modi di oscillazione presenti all'interno del suddetto intervallo é pari a

$$n_m = \frac{\Delta f}{\frac{u}{2l}} = \frac{2l}{u} \Delta f \quad (2.73)$$

Per il *principio di equipartizione dell'energia*, possiamo affermare che per ogni modo di propagazione presente nella linea, esiste una quantità KT di energia accumulata nella linea stessa, dove T é la temperatura assoluta della linea, e $K = 1.38066 \cdot 10^{-23} \text{J/K}$ é la costante di Boltzmann. Possiamo dunque esprimere l'energia immagazzinata come il prodotto dell'energia del singolo modo di oscillazione e del numero di modi, ossia

$$E = n_m \cdot KT \quad (2.74)$$

Si ricava quindi la potenza trasferita per effetto termico fra i due resistori

$$P = \frac{\frac{1}{2}E}{t_t} = \frac{1}{2} \frac{2l}{u} \frac{KT \Delta f}{\frac{l}{u}} = KT \Delta f \quad (2.75)$$

dove $t_t = \frac{l}{u}$ é il tempo che le onde elettromagnetiche impiegano per percorrere la linea di trasmissione (tempo di transito). Il fattore $\frac{1}{2}$ si puó spiegare affermando che l'energia che si trasferisce da R_I a R_{II} é la metà dell'energia totale immagazzinata all'interno della linea cortocircuitata (l'altra metà corrisponde infatti

alla porzione di energia che da R_{II} si trasferisce verso R_I). Giustamente l e u (parametri arbitrari della linea) non intervengono nel risultato finale.

Esprimiamo ora l'energia accumulata nella linea a partire dallo spettro di densità di potenza $G(f)$. Due resistenze in equilibrio termico collegate in parallelo trasferiscono l'una verso l'altra la medesima potenza. Nel dominio del tempo, la potenza scambiata risulta essere

$$p(t) = R [I(t)]^2 = R \left[\frac{v(t)}{2R} \right]^2 = \frac{[v(t)]^2}{4R} \quad (2.76)$$

Passando nel dominio delle frequenze, si introduce lo spettro di densità di potenza, ottenendo

$$P = \frac{G(f)}{4R} \Delta f \quad (2.77)$$

dove Δf è ancora l'intervallo di frequenze considerato.

Uguagliando le relazioni (2.75) e (2.77) si ottiene

$$KT \Delta f = \frac{G(f)}{4R} \Delta f t_t \quad (2.78)$$

Questo ci permette di ricavare la densità spettrale di potenza

$$G(f) = 4RKT \quad (2.79)$$

Il risultato, dovuto a Nyquist, si basa sul principio di equipartizione dell'energia, e risulta del tutto indipendente dalla frequenza. Secondo questo ragionamento lo spettro di densità di potenza della tensione $v_I(t)$ è costante al variare della frequenza (si veda al riguardo la figura 2.11).

Una trattazione più adeguata deve tenere tenere in considerazione anche la quantizzazione dei livelli energetici, ignorata dalla statistica di Boltzmann. Tenendo conto di tale quantizzazione, la statistica di Boltzmann viene modificata dal modello di Planck, secondo cui l'energia E_f di un modo di oscillazione a frequenza f risulta essere pari a

$$E_f = \frac{hf}{e^{\frac{hf}{kT}} - 1} \quad (2.80)$$

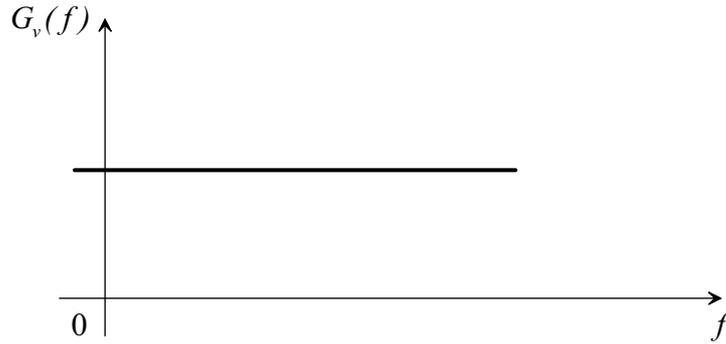


Figura 2.11: Densità spettrale della tensione secondo la statistica di Boltzmann.

dove $h = 6.62 \cdot 10^{-34} \text{ J s}$, é la costante di Planck.

Lo spettro di densità di potenza risulta essere pertanto

$$G(f) = 4R \frac{hf}{e^{\frac{hf}{kT}} - 1} \quad (2.81)$$

La differenza tra la (2.79) e la (2.81) si apprezza soltanto per frequenze tali per cui $hf \simeq kT$ o superiori. Nella pratica la densità spettrale del rumore termico si può trattare come costante per a frequenze inferiori a circa 10^{12} Hz (infatti per una temperatura $T = 290 \text{ K}$, l'uguaglianza $hf = kT$ si ottiene per valori di f circa uguali a 10^{12} Hz). Per frequenze maggiori si ha un andamento di tipo decrescente, come riportato in figura 2.12.

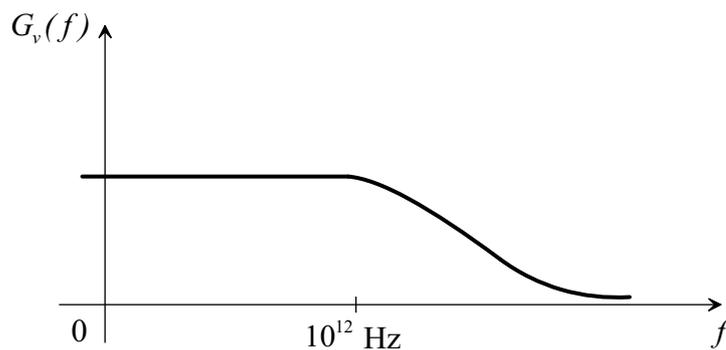


Figura 2.12: Densità spettrale della tensione secondo il modello di Planck.

Capitolo 3

Il problema della stima

3.1 Stima di parametri

Accade sovente che le caratteristiche del nostro sistema non siano ricavabili da misure dirette: in questo caso bisogna ricorrere a procedimenti di stima indiretta che permettono di risalire a una valutazione delle quantità incognite.

Per stima intendiamo un processo di selezione di un punto (*stima ottima*) in uno spazio continuo. La stima si differenzia dalla *decisione*, che consiste nella scelta di un punto tra un numero finito di alternative possibili (spazio discreto).

Alcuni domini applicativi del problema della stima riguardano ad esempio il tracking, il controllo, le telecomunicazioni e l'elaborazione dei segnali (signal processing).

Le quantità da stimare possono assumere varie forme. Distinguiamo due diverse categorie di problemi:

- *problema di stima statico*, in cui vengono stimati dei parametri, ossia quantità tempo invarianti;
- *problema di stima dinamico*, in cui viene stimato lo stato di un sistema, che può cambiare col trascorrere del tempo.

In questo capitolo ci occuperemo esclusivamente della stima di parametri.

3.1.1 Principi fondamentali della teoria della stima

Supponiamo di avere n osservazioni di un segnale x

$$x_i \equiv x[(i-1)T] \quad i = 1, 2, \dots, n \quad (3.1)$$

Si vuole stimare una quantità α . D'ora in avanti indicheremo con $\hat{\alpha} = g(x_1, x_2, \dots, x_n)$ la stima di α . Sia le misure sia la quantità da stimare sono variabili aleatorie. Possiamo quindi associare ad α una densità di probabilità. La densità di probabilità associata alla stima $\hat{\alpha}$, dato che esiste una quantità α è indicata con

$$p(\hat{\alpha}|\alpha) \quad (3.2)$$

In genere ci aspettiamo che il valor medio di questa distribuzione sia centrato sul valore che vogliamo stimare. Se questo non accade abbiamo a che fare con un cattivo stimatore.

3.1.2 Criteri di bontà della stima

In generale si prendono in considerazione due indicatori rappresentativi della bontà della stima:

1. polarizzazione,
2. stabilità.

Uno stimatore viene detto *buono* quando la sua polarizzazione

$$b \stackrel{\text{def}}{=} \alpha - E\{\hat{\alpha}|\alpha\} = 0 \quad (3.3)$$

dove

$$E\{\hat{\alpha}|\alpha\} = \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} \hat{\alpha}(x_1, x_2, \dots, x_n) p(x_1, x_2, \dots, x_n|\alpha) dx_1 \dots dx_n \quad (3.4)$$

è il valor medio condizionale della stima.

Una stima caratterizzata da $b = 0$ prende il nome di *stima non polarizzata*.

La varianza fornisce invece una idea dell'incertezza. Abbiamo

$$\sigma_{\hat{\alpha}}^2 = E\{(\hat{\alpha} - E\{\hat{\alpha}|\alpha\})^2|\alpha\} = E\{\hat{\alpha}^2|\alpha\} - E\{\hat{\alpha}|\alpha\}^2 \quad (3.5)$$

Effettuando un gran numero di misure é auspicabile che la varianza tenda a zero, ossia $\sigma_{\hat{\alpha}}^2 \rightarrow 0$.

Lo stimatore viene detto *consistente* se valgono entrambe le condizioni $b \rightarrow 0$ e $\sigma_{\hat{\alpha}}^2 \rightarrow 0$. In questo caso si dice che la stima é asintoticamente stabile.

Si può dimostrare che date n osservazioni x_1, x_2, \dots, x_n , esiste un limite inferiore per $\sigma_{\hat{\alpha}}^2$ che prende il nome di limite di *Cramer-Rao*:

$$\sigma_{\hat{\alpha}}^2 \geq \frac{1}{\int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} \left[\frac{d}{d\alpha} \ln p(x_1, x_2, \dots, x_n|\alpha) \right]^2 p(x_1, x_2, \dots, x_n|\alpha) dx_1 \dots dx_n} \quad (3.6)$$

(si veda l'Appendice C per una dimostrazione di tale disuguaglianza).

L'uguaglianza nell'equazione (3.6) si ottiene per

$$\frac{d}{d\alpha} \ln p(x_1, x_2, \dots, x_n|\alpha) = \lambda(\alpha)(\hat{\alpha} - \alpha) \quad (3.7)$$

dove $\lambda(\alpha)$ é una funzione arbitraria diversa da zero che dipende solo da α e non dalle osservazioni x_1, x_2, \dots, x_n . Se questo accade lo stimatore raggiunge il limite e la stima é detta *efficiente* e viene indicata con $\hat{\alpha}_{EFF}$. Questa stima si ricava dalla (3.7) ponendo $\hat{\alpha} = \hat{\alpha}_{EFF}$. In particolare se

$$\hat{\alpha}_{EFF} = \alpha \quad (3.8)$$

abbiamo

$$\frac{d}{d\alpha} \ln p(x_1, x_2, \dots, x_n|\alpha) = 0 \quad (3.9)$$

3.2 Stima secondo il criterio del minimo errore quadratico medio

Si vuole trovare la stima $\hat{\alpha}$ che minimizza l'errore quadratico medio

$$P = E\{(\alpha - \hat{\alpha})^2\} \quad (3.10)$$

Bisogna quindi risolvere il problema

$$\hat{\alpha} = \arg \min_{\hat{\alpha}} \{E\{(\alpha - \hat{\alpha})^2\}\} \quad (3.11)$$

calcolando la derivata rispetto ad $\hat{\alpha}$ dell'errore quadratico medio e uguagliandola a zero:

$$\frac{d}{d\hat{\alpha}} E\{(\alpha - \hat{\alpha})^2\} = 0 \quad (3.12)$$

3.2.1 Stima lineare (caso monodimensionale)

Approssimazione della stima con una costante

Prima di descrivere il metodo generale, consideriamo il semplice caso dell'approssimazione di α con un parametro costante C

$$\hat{\alpha} = C = \text{costante} \quad (3.13)$$

A tal fine calcoliamo

$$\frac{\partial}{\partial C} E\{(\alpha - C)^2\} = 0 \quad (3.14)$$

ossia

$$-E\{2(\alpha - C)\} = 0 \quad \Rightarrow \quad E\{\alpha\} - E\{C\} = 0 \quad \Rightarrow \quad C = E\{\alpha\} = \bar{\alpha} \quad (3.15)$$

Approssimazione della stima con una funzione lineare

Consideriamo ora il caso dell'approssimazione di α con una funzione lineare, ossia

$$\hat{\alpha} = ax + b \quad (3.16)$$

Dovremo determinare i parametri a e b che minimizzino l'errore quadratico medio, per cui occorre risolvere il seguente problema di ottimizzazione:

$$\arg \min_{a,b} E\{(\alpha - ax - b)^2\} \quad (3.17)$$

Se si considera come un'unica variabile aleatoria $\alpha - ax$ si ottiene un problema che può essere ricondotto al caso precedente. Poniamo dunque

$$z = \alpha - ax \quad (3.18)$$

dove z è una nuova variabile aleatoria.

Applicando quanto ricavato per il caso precedente, si ottiene

$$\frac{\partial}{\partial z} E\{(z - b)^2\} = E\{2(z - b)\} = 0 \quad (3.19)$$

da cui

$$E\{z\} - E\{b\} = 0 \quad \Rightarrow \quad b = E\{z\} = \bar{z} \quad \Rightarrow \quad b = \bar{\alpha} - a\bar{x} \quad (3.20)$$

Sostituendo il valore di b appena ricavato nell'equazione (3.17), si ottiene

$$\begin{aligned} \arg \min_{a,b} E\{(\alpha - ax - b)^2\} &= \arg \min_{a,b} E\{(\alpha - ax - \bar{\alpha} + a\bar{x})^2\} = \\ &= \arg \min_{a,b} E\{[(\alpha - \bar{\alpha}) - a(x - \bar{x})]^2\} = \\ &= \arg \min_{a,b} [a^2 E\{(x - \bar{x})^2\} - 2aE\{(x - \bar{x})(\alpha - \bar{\alpha})\} + E\{(\alpha - \bar{\alpha})^2\}] \end{aligned} \quad (3.21)$$

Derivando rispetto ad a e uguagliando il tutto a zero si ottiene:

$$\frac{\partial}{\partial a} [a^2 E\{(x - \bar{x})^2\} - 2aE\{(x - \bar{x})(\alpha - \bar{\alpha})\} + E\{(\alpha - \bar{\alpha})^2\}] = 0 \quad (3.22)$$

ossia

$$aE\{(x - \bar{x})^2\} - E\{(x - \bar{x})(\alpha - \bar{\alpha})\} = 0 \quad (3.23)$$

da cui

$$a\sigma_x^2 - \mu_{x\alpha} = 0 \quad (3.24)$$

dove $\mu_{x\alpha}$ indica la covarianza tra x ed α .

E' possibile quindi ricavare il valore dei coefficienti a e b come

$$a = \frac{\mu_{x\alpha}}{\sigma_x^2} \quad (3.25)$$

$$b = \bar{\alpha} - \frac{\mu_{x\alpha}}{\sigma_x^2} \bar{x} \quad (3.26)$$

La stima lineare secondo il criterio del minimo errore quadratico medio quindi risulta

$$\hat{\alpha} = \frac{\mu_{x\alpha}}{\sigma_x^2} x + \bar{\alpha} - \frac{\mu_{x\alpha}}{\sigma_x^2} \bar{x} \quad (3.27)$$

Introducendo il *coefficiente di correlazione* tra la variabile α e la variabile x

$$r = \frac{\mu_{x\alpha}}{\sigma_x \sigma_\alpha} \quad (3.28)$$

possiamo riscrivere $\hat{\alpha}$ nel seguente modo:

$$\hat{\alpha} = \bar{\alpha} + \frac{\sigma_\alpha}{\sigma_x} r (x - \bar{x}) \quad (3.29)$$

La stima é quindi legata ad $\bar{\alpha}$ ma anche a un termine che riguarda la correlazione tra le due variabili, e a $(x - \bar{x})$, che misura quanto x si discosta dal valore medio.

Osservazione. Per semplicitá consideriamo il caso

$$\bar{x} = 0 \quad e \quad \bar{\alpha} = 0 \quad (3.30)$$

da cui

$$\hat{\alpha} = \frac{\sigma_\alpha}{\sigma_x} \cdot r \cdot x \quad (3.31)$$

Osserviamo che

- se r é piccolo, $\hat{\alpha}$ tende ad $\bar{\alpha}$,
- se $r \simeq 1$ le misure sono molto correlate con la quantità che si vuole stimare e $\hat{\alpha} \simeq \frac{\sigma_\alpha}{\sigma_x} x$.

Pertanto se x assume valori prossimi alla sua dispersione σ_x , allora anche la stima tenderá ad assumere valori vicini alla sua dispersione

$$\hat{\alpha} \simeq \frac{\sigma_\alpha}{\sigma_x} x \simeq \frac{\sigma_\alpha}{\sigma_x} \sigma_x = \sigma_\alpha \quad (3.32)$$

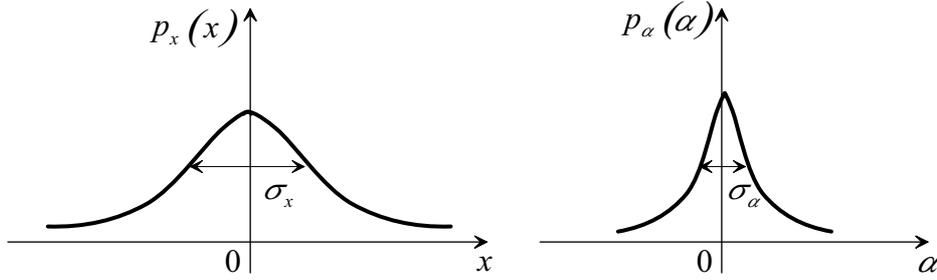


Figura 3.1: Rappresentazione grafica delle dispersioni di x e di α .

Questa stima é buona perché rispetta il principio della non polarizzazione. Inoltre nel caso di coefficiente di correlazione prossimo a uno avremo anche una riduzione dell'incertezza della stima rispetto all'incertezza di x .

3.2.2 Stima non lineare (caso monodimensionale)

Si vuole ora approssimare α con una stima non lineare, ossia $\hat{\alpha}$ é espressa come funzione generica g della misurazione x a disposizione:

$$\hat{\alpha} = g(x) \quad (3.33)$$

Consideriamo l'errore quadratico medio

$$P = E\{(\hat{\alpha} - \alpha)^2\} = E\{(g(x) - \alpha)^2\} = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} [g(x) - \alpha]^2 p_{x\alpha}(x, \alpha) d\alpha dx \quad (3.34)$$

dove la funzione

$$[g(x) - \alpha]^2 \quad (3.35)$$

é una quantità che dipende solo da x e α . Pertanto il valor medio che consideriamo deve dipendere dalla distribuzione congiunta $p_{x\alpha}$. Esprimiamo questa

distribuzione congiunta come prodotto di due densità

$$p_{x\alpha} = p_x(x)p_{\alpha|x}(\alpha|x) \quad (3.36)$$

Quindi l'integrale della (3.34) diventa

$$P = \int_{-\infty}^{+\infty} p_x(x) \int_{-\infty}^{+\infty} [g(x) - \alpha]^2 p_{\alpha|x}(\alpha|x) d\alpha dx \quad (3.37)$$

Le funzioni integrande sono tutte maggiori o uguali a zero, pertanto cercare il valore che rende minima la (3.34) equivale a cercare il valore che rende minimo il solo integrale piú interno

$$\int_{-\infty}^{+\infty} [g(x) - \alpha]^2 p_{\alpha|x}(\alpha|x) d\alpha \quad (3.38)$$

Per ricavare la funzione $g(x)$ che rende minimo l'integrale (3.38) consideriamone la derivata e uguagliamola a zero:

$$\frac{\partial}{\partial g(x)} \left[\int_{-\infty}^{+\infty} [g(x) - \alpha]^2 p_{\alpha|x}(\alpha|x) d\alpha \right] = 0 \quad (3.39)$$

da cui

$$2 \int_{-\infty}^{+\infty} (g(x) - \alpha) p_{\alpha|x}(\alpha|x) d\alpha = 0 \quad (3.40)$$

quindi

$$\int_{-\infty}^{+\infty} g(x) p_{\alpha|x}(\alpha|x) d\alpha = \int_{-\infty}^{+\infty} \alpha p_{\alpha|x}(\alpha|x) d\alpha \quad (3.41)$$

poiché

$$\int_{-\infty}^{+\infty} g(x) p_{\alpha|x}(\alpha|x) d\alpha = g(x) \underbrace{\int_{-\infty}^{+\infty} p_{\alpha|x}(\alpha|x) d\alpha}_1 \quad (3.42)$$

si ricava allora che la $g(x)$ che permette di ottenere la stima ottima voluta é il valor medio condizionato della variabile α che si vuole stimare, ossia la *curva di regressione* del parametro α

$$\hat{\alpha} = g(x) = \int_{-\infty}^{+\infty} \alpha p_{\alpha|x}(\alpha|x) d\alpha = E\{\alpha|x\} \quad (3.43)$$

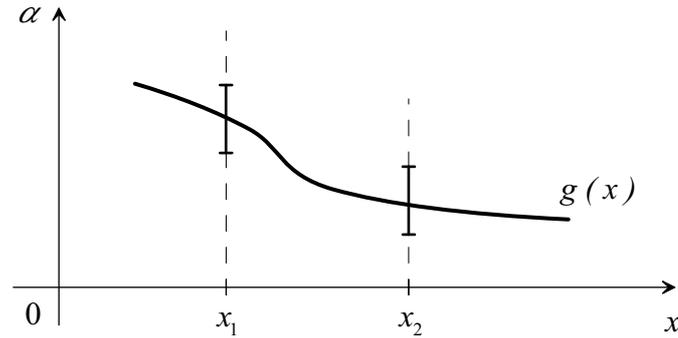


Figura 3.2: Rappresentazione grafica della funzione di regressione $g(x)$.

3.2.3 Principio di ortogonalit 

Il principio di ortogonalit  afferma che l'errore della stima secondo il criterio MEQM (minimo errore quadratico medio) risulta ortogonale ai dati. L'ortogonalit  qui   intesa in senso statistico come

$$E\{\text{errore} \cdot \text{dati}\} = 0 \quad (3.44)$$

Consideriamo l'esempio del caso lineare monodimensionale, dove la funzione stima  

$$\hat{\alpha} = ax + b \quad (3.45)$$

Si devono ricavare i parametri di a e b che minimizzano l'errore quadratico medio

$$P = E\{(\alpha - ax - b)^2\} \quad (3.46)$$

pertanto dobbiamo risolvere il seguente problema di ottimizzazione:

$$\arg \min_{a,b} E\{(\alpha - ax - b)^2\} \quad (3.47)$$

Derivando l'equazione precedente rispetto ad a e uguagliando il tutto a zero si ottiene

$$\frac{\partial}{\partial a} E\{(\alpha - ax - b)^2\} = 2E\{(\alpha - ax - b)(-x)\} = 0 \quad (3.48)$$

dove l'errore   rappresentato dalla quantit  $(\alpha - ax - b)$ mentre i dati (le misure) sono rappresentate dalla quantit  $(-x)$ (il segno meno non conta nulla in quanto

abbiamo imposto l'uguaglianza a zero). Riarrangiando i termini dell'equazione (3.48) si ottiene

$$E\{(\alpha - ax - b)(x)\} = 0 \quad (3.49)$$

che é una forma del principio di ortogonalit  (3.44).

E' possibile dare una interpretazione geometrica del principio di ortogonalit  ricorrendo per semplicit  a uno stimatore lineare omogeneo (caratterizzato cio  da $b = 0$).

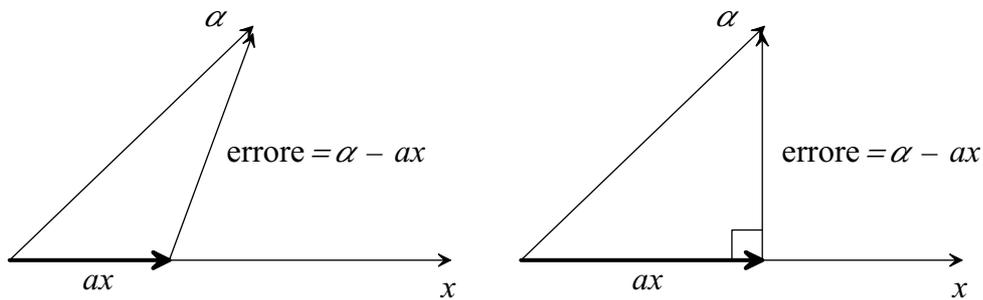


Figura 3.3: Rappresentazione grafica del principio di ortogonalit .

Il vettore x costituisce l'insieme di dati, il vettore α é la quantit  da stimare. Si ha minimo errore quando ax é la proiezione ortogonale di α sui dati.

Il principio di ortogonalit  assicura quindi che si ha stima quando l'errore é ortogonale ai dati.

Osservazione

Il quadrato della lunghezza di $(\alpha - ax)$ uguaglia il prodotto scalare di α con l'errore $(\alpha - ax)$.

Dimostrazione. Definiamo l'errore commesso nel seguente modo:

$$\begin{aligned} \text{err} &= E\{(\alpha - ax)^2\} = E\{(\alpha - ax)(\alpha - ax)\} = \\ &= E\{(\alpha - ax)\alpha\} - E\{(\alpha - ax)ax\} = \\ &= E\{\alpha^2\} - E\{a\alpha x\} - E\{a\alpha x\} + E\{(ax)^2\} \end{aligned} \quad (3.50)$$

Posto

$$E\{(ax)^2\} = -E\{(ax)^2\} + 2E\{(ax)^2\} \quad (3.51)$$

e sostituendo nell'equazione (3.50) si ottiene

$$\text{err} = E\{\alpha^2\} - E\{(ax)^2\} - 2aE\{(\alpha - ax)x\} \quad (3.52)$$

Per il principio di ortogonalit a l'ultimo termine dell'equazione precedente   uguale a zero. Quindi rimane

$$\text{err} = E\{\alpha^2\} - E\{(ax)^2\} = E\{(\alpha - ax)\alpha\} \quad (3.53)$$

3.2.4 Stima lineare (caso a pi u dimensioni)

Date n osservazioni x_1, x_2, \dots, x_n , nel caso di stima lineare dovremo determinare la funzione

$$\hat{\alpha} = a_1x_1 + a_2x_2 + \dots + a_nx_n \quad (3.54)$$

che minimizza l'errore quadratico medio

$$P = E\{[\bar{\alpha}(a_1x_1 + a_2x_2 + \dots + a_nx_n)]^2\} \quad (3.55)$$

ossia dovremo risolvere il problema di ottimizzazione

$$\arg \min_A E\{[\bar{\alpha}(a_1x_1 + a_2x_2 + \dots + a_nx_n)]^2\} \quad (3.56)$$

dove $A \stackrel{\text{def}}{=} [a_1 \ a_2 \ \dots \ a_n]$.

Si pu o applicare il principio di ortogonalit a, che vale anche nel caso a pi u dimensioni, e fare in modo che ogni errore sia ortogonale ai dati. In altri termini deve essere verificato

$$E\{[\bar{\alpha}(a_1x_1 + a_2x_2 + \dots + a_nx_n)]x_i\} = 0 \quad \forall i = 1, 2, \dots, n \quad (3.57)$$

Si potevano ricavare le stesse condizioni imponendo

$$\frac{\partial}{\partial a_i} E\{[\bar{\alpha}(a_1x_1 + a_2x_2 + \dots + a_nx_n)]^2\} = 0 \quad \forall i = 1, \dots, n \quad (3.58)$$

covarianza infatti é una matrice di elementi

$$C_{ij} = E\{(x_i - \bar{x}_i)(x_j - \bar{x}_j)\} \quad (3.64)$$

Osservazione. L'errore quadratico medio P puó essere scritto come

$$P = R_{00} - (a_1 R_{01} + a_2 R_{02} + \dots + a_n R_{0n})^2 \quad (3.65)$$

dove $R_{00} = E\{\alpha^2\}$.

I risultati trovati per lo stimatore lineare omogeneo, si possono estendere al caso di stimatore non omogeneo (in cui $b \neq 0$). Avremo in questo caso

$$b = a_0, \quad \hat{\alpha} = a_0 + a_1 x_1 + a_2 x_2 + \dots + a_n x_n \quad (3.66)$$

Posto $a_0 = a_0 x_0$ con $x_0 = 1$ si ritorna formalmente al caso precedente. Otterremo un sistema di equazioni del tipo

$$\left\{ \begin{array}{l} a_0 + \bar{x}_1 a_1 + \dots + \bar{x}_n a_n = \bar{\alpha} \\ \bar{x}_1 a_0 + R_{11} a_1 + R_{12} a_2 + \dots + R_{1n} a_n = R_{01} \\ \bar{x}_2 a_0 + R_{21} a_1 + R_{22} a_2 + \dots + R_{2n} a_n = R_{02} \\ \vdots \\ \bar{x}_n a_0 + R_{n1} a_1 + R_{n2} a_2 + \dots + R_{nn} a_n = R_{0n} \end{array} \right. \quad (3.67)$$

con $n+1$ incognite e $n+1$ equazioni. Si noti come gli elementi di R in cui compare x_0 abbiano la seguente forma

$$E\{x_0 x_j\} = \begin{cases} 1 & \text{se } j = 0 \\ \bar{x}_j & \text{se } j \neq 0 \end{cases} \quad (3.68)$$

Dalla soluzione del sistema si ottiene il vettore A e quindi i coefficienti della stima.

3.2.5 Stima non lineare (caso a piú dimensioni)

In questo caso la stima $\hat{\alpha}$ della grandezza α é una funzione generica g delle n osservazioni x_1, x_2, \dots, x_n

$$\hat{\alpha} = g(x_1, x_2, \dots, x_n) \quad (3.69)$$

La stima $\hat{\alpha}$ deve essere tale per cui l'errore quadratico medio

$$P = E\{[\alpha - g(x_1, x_2, \dots, x_n)]^2 | x_1, x_2, \dots, x_n\} \quad (3.70)$$

sia minimo. Generalizzando i concetti visti in una dimensione si ha

$$\hat{\alpha} = g(x_1, x_2, \dots, x_n) = \int_{-\infty}^{+\infty} \alpha p(\alpha | x_1, x_2, \dots, x_n) d\alpha \quad (3.71)$$

Nel caso in cui le variabili aleatorie abbiano distribuzione gaussiana e valor medio nullo (una distribuzione del genere é indicata spesso con la notazione $N(0, \sigma^2)$), la stima lineare e quella non lineare coincidono.

Osservazione. Nel caso della stima lineare si é soliti usare una notazione particolare per indicare la stima stessa

$$\hat{\alpha} = \hat{E}\{\alpha | x_1, x_2, \dots, x_n\} \quad (3.72)$$

dove l'operatore $\hat{E}\{\cdot\}$ indica la stima nel senso del minimo errore quadratico medio. Si faccia attenzione a non confondere questo operatore con l'operatore "expectation" $E\{\cdot\}$. Solamente nell'ipotesi di densitá di probabilitá normali $N(0, \sigma^2)$ abbiamo

$$\hat{E}\{\alpha | x_1, x_2, \dots, x_n\} \equiv E\{\alpha | x_1, x_2, \dots, x_n\} \quad (3.73)$$

Sotto l'ipotesi di gaussianitá e valor medio nullo é anche possibile scrivere l'espressione della densitá di probabilitá a posteriori $p(\alpha | x_1, x_2, \dots, x_n)$. Essendo le variabili in gioco gaussiane, anche tale densitá di probabilitá sará gaussiana, con valor medio pari alla stima $\hat{\alpha}$ e varianza pari all'errore quadratico medio P . Infatti poiché $\bar{\alpha} = \hat{\alpha}$, la varianza risulta essere

$$\sigma^2 = E\{(\alpha - \bar{\alpha})^2 | x_1, x_2, \dots, x_n\} = E\{(\alpha - \hat{\alpha})^2 | x_1, x_2, \dots, x_n\} = P. \quad (3.74)$$

3.2.6 Determinazione di densità di probabilità condizionali

Abbiamo visto come

$$p(\alpha|x_1, x_2, \dots, x_n) = \frac{1}{\sqrt{2\pi P}} \exp \left[-\frac{[\alpha - (a_1x_1 + a_2x_2 + \dots + a_nx_n)]^2}{2P} \right] \quad (3.75)$$

Supponiamo di avere x_1 e x_2 congiuntamente normali a valor medio nullo e di voler calcolare $p(x_2|x_1)$. Utilizzando i risultati esposti in precedenza abbiamo

$$\hat{E}\{x_2|x_1\} = ax_1 \quad (3.76)$$

dove $a = \frac{R_{12}}{R_{11}}$.

Per quanto riguarda la varianza, introduciamo dapprima una espressione più operativa per P :

$$P = E\{(\alpha - \hat{\alpha})^2\} = E\{(\alpha - \hat{\alpha})(\alpha - \hat{\alpha})\} = E\{\alpha(\alpha - \hat{\alpha})\} - E\{\hat{\alpha}(\alpha - \hat{\alpha})\} \quad (3.77)$$

Per il principio di ortogonalità l'errore $\alpha - \hat{\alpha}$ è ortogonale ai dati, quindi il secondo termine dell'equazione (3.77) si annulla. Resta quindi

$$P = E\{\alpha(\alpha - \hat{\alpha})\} \quad (3.78)$$

Pertanto la varianza risulta essere pari a

$$\sigma_{x_2|x_1}^2 = P = E\{x_2(x_2 - ax_1)\} = R_{22} - aR_{21} = R_{22} - aR_{12} \quad (3.79)$$

Sostituendo i risultati trovati nella (3.75) si ha

$$p(x_2|x_1) = \frac{1}{\sqrt{2\pi P}} \exp \left[-\frac{[x_2 - \frac{R_{12}}{R_{11}}x_1]^2}{2P} \right]. \quad (3.80)$$

3.3 Stima Bayesiana

Anche in questo caso abbiamo una serie di misurazioni x_1, x_2, \dots, x_n e una quantità α (parametro) da stimare. E' possibile scegliere di calcolare la stima $\hat{\alpha}$

di α in vari modi. Ad esempio abbiamo visto che si può ricorrere al valor medio condizionato

$$\hat{\alpha} = \int_{-\infty}^{+\infty} \alpha p(\alpha|x_1, x_2, \dots, x_n) d\alpha \quad (3.81)$$

In alternativa é possibile scegliere $\hat{\alpha}$ in corrispondenza del valore di picco della densitá di probabilitá $p(\alpha|x_1, x_2, \dots, x_n)$. Nel caso in cui la distribuzione sia simmetrica le due scelte coincidono, mentre in casi piú generali di assenza di simmetria potrebbero esserci differenze anche notevoli nella scelta dello stimatore.

Ci proponiamo di trovare un metodo che permetta di stabilire di volta in volta quale definizione di stimatore risulta piú opportuna per i nostri scopi. La scelta di quale sia lo stimatore migliore dipende dalla quantificazione dell'errore che si vuole dare. In altri termini, ci possono essere alcuni metodi di stima che privilegiano una certa tipologia di errore e altri che invece tendono a ridurre altri tipi di errore. Si considera allora una sorta di misura dell'errore introducendo una *funzione di costo* o *funzione di perdita* (*loss function*) $L(\hat{\alpha}, \alpha)$. Lo scopo di questa funzione é associare a ciascun errore un peso che dica quanto importante é l'errore che si sta commettendo: errori particolarmente dannosi avranno peso elevato, mentre errori trascurabili avranno un peso in genere ridotto. Solitamente la funzione $L(\hat{\alpha}, \alpha)$ é non negativa e convessa.

3.3.1 Funzioni di perdita tipiche

Vediamo ora due esempi classici funzione di peso $L(\hat{\alpha}, \alpha)$.

Funzione A. La funzione $L(\hat{\alpha}, \alpha)$ (vedi figura 3.4) é costante per tutti i valori di $\hat{\alpha}$ esterni a un intervallo di ampiezza $\epsilon \geq 0$ centrato attorno al valore che si vuole stimare. Gli errori tali per cui $|\hat{\alpha} - \alpha| < \epsilon$ sono tollerati perché hanno un peso nullo, mentre gli errori tali per cui $|\hat{\alpha} - \alpha| \geq \epsilon$ sono tutti pesati allo stesso modo, ossia non ci sono errori privilegiati.

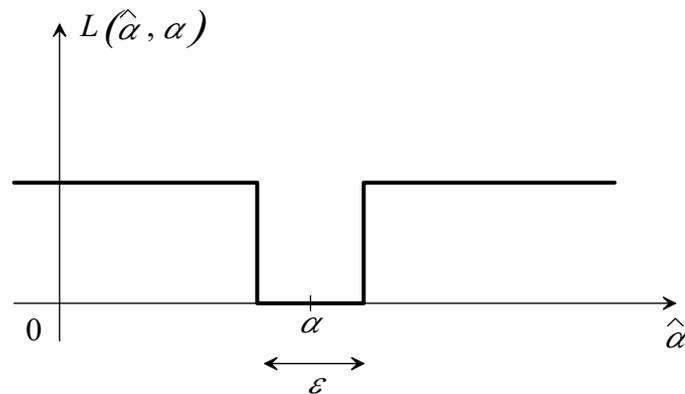


Figura 3.4: Funzione di peso di tipo A.

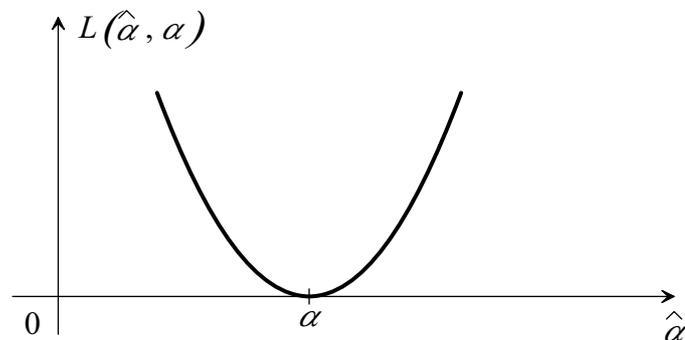


Figura 3.5: Funzione di peso di tipo B.

Funzione B. La funzione $L(\hat{\alpha}, \alpha)$ (vedi figura 3.5) é una funzione quadratica centrata sul valore da stimare in cui gli errori vengono pesati diversamente: errori piccoli (quindi piú accettabili) hanno un peso piccolo ma comunque diverso da zero, mentre al crescere dell'errore cresce anche il peso a esso attribuito.

3.3.2 Lo stimatore ottimo bayesiano

Lo stimatore bayesiano é uno stimatore che sceglie la stima $\hat{\alpha}$ in modo da minimizzare la media d'insieme \bar{L} della funzione di perdita

$$\bar{L} = \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} L(\hat{\alpha}, \alpha) p(x_1, x_2, \dots, x_n, \alpha) dx_1 dx_2 \dots dx_n d\alpha \quad (3.82)$$

Applicando Il teorema di Bayes alla distribuzione di probabilitá congiunta

$p(x_1, x_2, \dots, x_n, \alpha)$ si ricava

$$p(x_1, x_2, \dots, x_n, \alpha) = p(\alpha|x_1, x_2, \dots, x_n)p_x(x_1, x_2, \dots, x_n) \quad (3.83)$$

Sostituendo nella relazione (3.82) abbiamo

$$\bar{L} = \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} p_x(x_1, x_2, \dots, x_n) \int_{-\infty}^{+\infty} L(\hat{\alpha}, \alpha) p(\alpha|x_1, x_2, \dots, x_n) d\alpha dx_1 \dots dx_n \quad (3.84)$$

Poiché sicuramente $p_x(x_1, x_2, \dots, x_n) \geq 0$, la ricerca del valore $\hat{\alpha}$ che rende minimo \bar{L} si riduce all'ricerca del valore che rende minimo l'integrale piú interno

$$\int_{-\infty}^{+\infty} L(\hat{\alpha}, \alpha) p(\alpha|x_1, x_2, \dots, x_n) d\alpha \quad (3.85)$$

La funzione integranda $L(\hat{\alpha}, \alpha) p(\alpha|x_1, x_2, \dots, x_n)$ prende il nome di *funzione di rischio*.

Nel caso della funzione di peso di tipo A (vedi figura 3.4) possiamo scrivere

$$\bar{L} = 1 - \int_{\hat{\alpha}-\frac{\epsilon}{2}}^{\hat{\alpha}+\frac{\epsilon}{2}} p(\alpha|x_1, x_2, \dots, x_n) d\alpha \quad (3.86)$$

che nel caso di ϵ infinitesimo puó essere approssimata con

$$\bar{L} = 1 - \epsilon p(\hat{\alpha}|x_1, x_2, \dots, x_n) \quad (3.87)$$

Poiché $\epsilon \geq 0$ e $p(\hat{\alpha}|x_1, x_2, \dots, x_n) \geq 0$, avremo

$$\arg \min_{\hat{\alpha}} \bar{L} = \arg \max_{\hat{\alpha}} p(\hat{\alpha}|x_1, x_2, \dots, x_n) \quad (3.88)$$

La densità di probabilità $p(\hat{\alpha}|x_1, x_2, \dots, x_n)$ prende il nome di densità di probabilità a *posteriori*. Il criterio descritto per la determinazione di $\hat{\alpha}$ prende il nome di *criterio MAP*, o *criterio della massima probabilità a posteriori*.

Ci si attende che la densità di probabilità a posteriori sia una funzione centrata sul valore da stimare α . Tale densità é genere piú "stretta" rispetto alla densità a priori, che inoltre potrebbe anche non essere centrata sul valore α del parametro.

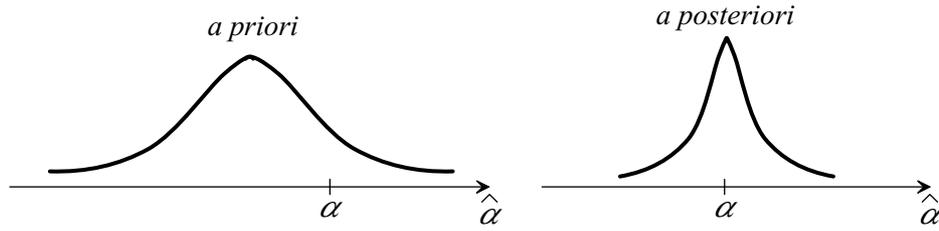


Figura 3.6: Esempio di densità di probabilità a priori (a sinistra) e di densità di probabilità a posteriori (a destra).

Nel caso della funzione di peso di tipo B (vedi figura 3.5) possiamo scrivere

$$\bar{L} = \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} (\alpha - \hat{\alpha})^2 p_x(x_1, x_2, \dots, x_n) p(\alpha | x_1, x_2, \dots, x_n) d\alpha dx_1 \dots dx_n \quad (3.89)$$

Come nel caso precedente, cercare il valore $\hat{\alpha}$ che rende minima questa quantità equivale a cercare il valore che rende minimo l'errore quadratico medio

$$P = \int_{-\infty}^{+\infty} (\alpha - \hat{\alpha})^2 p(\alpha | x_1, x_2, \dots, x_n) d\alpha \quad (3.90)$$

Calcoliamo allora la derivata e uguagliamola a zero:

$$\frac{d}{d\hat{\alpha}} \left(\int_{-\infty}^{+\infty} (\alpha - \hat{\alpha})^2 p(\alpha | x_1, x_2, \dots, x_n) d\alpha \right) = 0 \quad (3.91)$$

Svolgendo i calcoli necessari si ricava

$$\hat{\alpha} = \int_{-\infty}^{+\infty} \alpha p(\alpha | x_1, x_2, \dots, x_n) d\alpha \quad (3.92)$$

Notiamo che nel caso di una funzione di perdita di tipo B la stima bayesiana coincide con la stima al minimo errore quadratico medio.

Osservazione. Lo stesso risultato si ottiene anche scegliendo come funzione di perdita una qualunque funzione L purché:

1. $L(\hat{\alpha}, \alpha) = L(\hat{\alpha} - \alpha)$ ossia L dipenda solamente dalla differenza $\hat{\alpha} - \alpha$;
2. $L(\xi) = L(-\xi)$ ossia L sia una funzione simmetrica;

3. L sia una funzione convessa;
4. $p(\alpha|x_1, x_2, \dots, x_n)$ sia una funzione simmetrica attorno alla stima $\hat{\alpha}$ ottenuta con la stima al minimo errore quadratico medio.

3.4 Stima a massima verosimiglianza

Introduciamo ora la stima ML (o stima *maximum likelihood* o della *massima verosimiglianza*) effettuando un confronto con la stima MAP introdotta nel paragrafo precedente.

Il difetto maggiore della stima MAP é la necessità di dover conoscere la densità di probabilità a posteriori $p(\alpha|x_1, x_2, \dots, x_n)$, che di solito non é nota. Più spesso invece é nota la densità $p(x_1, x_2, \dots, x_n|\alpha)$. Per dedurre $p(\alpha|x_1, x_2, \dots, x_n)$ da $p(x_1, x_2, \dots, x_n|\alpha)$ si utilizza il teorema di Bayes:

$$p(\alpha|x_1, x_2, \dots, x_n) = \frac{1}{p_x(x_1, x_2, \dots, x_n)} p_\alpha(\alpha) p(x_1, x_2, \dots, x_n|\alpha) \quad (3.93)$$

Posto $1/p_x(x_1, x_2, \dots, x_n) = K \geq 0$ indipendentemente dalla quantità da stimare α , possiamo ottenere la stima calcolando

$$\frac{d}{d\alpha} p(\alpha|x_1, x_2, \dots, x_n) = 0 \quad (3.94)$$

ossia

$$\frac{d}{d\alpha} p_\alpha(\alpha) p(x_1, x_2, \dots, x_n|\alpha) + p_\alpha(\alpha) \frac{d}{d\alpha} p(x_1, x_2, \dots, x_n|\alpha) = 0 \quad (3.95)$$

Dividendo per $p_\alpha(\alpha) p(x_1, x_2, \dots, x_n|\alpha)$ si ottiene

$$\frac{1}{p_\alpha(\alpha)} \frac{d}{d\alpha} p_\alpha(\alpha) + \frac{1}{p(x_1, x_2, \dots, x_n|\alpha)} \frac{d}{d\alpha} p(x_1, x_2, \dots, x_n|\alpha) = 0 \quad (3.96)$$

Poiché $\frac{1}{p} \frac{dp}{d\alpha} = \frac{d}{d\alpha} \ln p$, la (3.96) diventa

$$\frac{d}{d\alpha} \ln p_\alpha(\alpha) + \frac{d}{d\alpha} \ln p(x_1, x_2, \dots, x_n|\alpha) = 0 \quad (3.97)$$

Le relazioni (3.96) e (3.97) forniscono la stima ML. In entrambe le equazioni occorre conoscere sia $p_\alpha(\alpha)$ sia $p(x_1, x_2, \dots, x_n|\alpha)$. Nel caso, del resto abbastanza frequente in pratica, in cui $p_\alpha(\alpha)$ non sia nota, si assume $p_\alpha(\alpha) = \text{costante}$ in quanto non si possiede nessuna informazione su α . In questo caso

$$\frac{d}{d\alpha} \ln p_\alpha(\alpha) = 0 \quad (3.98)$$

e le (3.96) e (3.97) si riducono a

$$\frac{d}{d\alpha} p(x_1, x_2, \dots, x_n|\alpha) = 0 \quad (3.99)$$

$$\frac{d}{d\alpha} \ln p(x_1, x_2, \dots, x_n|\alpha) = 0. \quad (3.100)$$

3.5 Stima di parametri con rumore additivo

Consideriamo la situazione in cui una informazione α da stimare venga impressa in un segnale di forma nota, attribuendo l'informazione nel valore dei suoi parametri caratteristici (quali ad esempio ampiezza, fase etc), al fine di facilitarne il trasferimento o la memorizzazione. Il caso piú comune di questa operazione risiede in una modulazione d'ampiezza, in cui un segnale da trasmettere viene moltiplicato per un segnale di forma nota, alterandone l'ampiezza. Il segnale disponibile per la misura (segnale osservabile) sará la sovrapposizione del segnale originario e di un rumore additivo. Ad esempio, in seguito a interferenze sul canale, viene sommato al risultato della modulazione un rumore che distorce il segnale originariamente trasmesso.

La situazione puó essere rappresentata con lo schema a blocchi riportato in figura 3.7, dove

- s_i é un segnale noto con certezza per quanto riguarda la forma (ad esempio sinusoidale, esponenziale etc.);

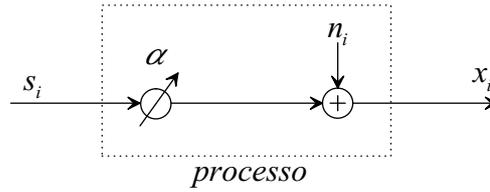


Figura 3.7: Rappresentazione schematica del processo di somma di un rumore su un segnale s_i di forma nota in cui l'informazione risiede nel parametro α .

- α é una variabile aleatoria che amplifica ad esempio l'ampiezza del segnale: si tratta del parametro da stimare;
- n_i é un rumore additivo;
- x_i é il segnale che misuriamo alla fine del processo.

La quantità misurata in uscita si può esprimere nel seguente modo:

$$x_i = \alpha s_i + n_i \quad i = 1, 2, \dots, n \quad (3.101)$$

dove sono state effettuate n misurazioni.

Solitamente si suppone che il rumore additivo sia un rumore *bianco*, ossia che il suo spettro in frequenza sia costante. Inoltre si assume che esso sia caratterizzato da una distribuzione di probabilità gaussiana a valor medio nullo e varianza σ^2 . In sintesi

$$p_n(n) = N(0, \sigma^2) \quad (3.102)$$

Usiamo il criterio MAP per stimare α con una quantità $\hat{\alpha}$. Consideriamo anzitutto la seguente probabilità condizionale:

$$p(\alpha|x_i) = \frac{p(x_i, \alpha)}{p_x(x_i)} \quad (3.103)$$

Per semplicità consideriamo una sola misura, ossia $x_i = x$. Dobbiamo quindi calcolare $p(x, \alpha)$, con $x = \alpha s + n$.

Osserviamo che tutta l'aleatorietà di x é rappresentata dall'aleatorietà del rumore considerando α come “dato”. E' possibile esprimere il rumore come

$$n = g(x) = x - \alpha s \quad (3.104)$$

Le grandezze n e x sono legate attraverso una funzione $g(x)$, pertanto é possibile esprimere la densità di probabilità del rumore in funzione della densità di probabilità del segnale misurato e della funzione g sfruttando la teoria delle funzioni di variabile aleatoria. In particolare abbiamo

$$p_n(n) = \frac{p_x(x)}{\left| \frac{dg}{dx} \right|} \Rightarrow p_x(x) = p_n(n) \left| \frac{dg}{dx} \right| \quad (3.105)$$

E' possibile usare questa espressione per calcolare $p(x, \alpha)$, ottenendo

$$p(x, \alpha) = p(n, \alpha) \left| \frac{dg}{dx} \right| \quad (3.106)$$

Poiché in questo caso $\frac{dg}{dx} = 1$, abbiamo $p(x, \alpha) = p(n, \alpha)$. Ipotizzando che il rumore e la quantità da stimare α siano tra loro statisticamente indipendenti, (questa ipotesi é del tutto ragionevole in quanto non c'è nessun motivo per cui dovrebbero essere legate statisticamente tra loro) si può scrivere

$$p(n, \alpha) = p_n(n)p_\alpha(\alpha) \quad (3.107)$$

Sostituendo $n = x - \alpha s$ nella relazione precedente si ottiene

$$p(n, \alpha) = p_n(n)p_\alpha(\alpha) = p_n(x - \alpha s)p_\alpha(\alpha) \quad (3.108)$$

Per il principio della massima probabilità a posteriori, la stima $\hat{\alpha}$ ottima é il valore di α che massimizza la densità di probabilità $p(\alpha|x)$, ossia occorre trovare il massimo di $p_n(x - \alpha s)p_\alpha(\alpha)$. Una rappresentazione tridimensionale del problema di massimizzazione é riportata in figura 3.8.

La $p_n(n)p_\alpha(\alpha)$ é una gaussiana bidimensionale. Proiettando la figura 3.8 sul piano $(x, x - \alpha s)$ si ottiene la rappresentazione bidimensionale attraverso curve di livello schematizzata in figura 3.9.

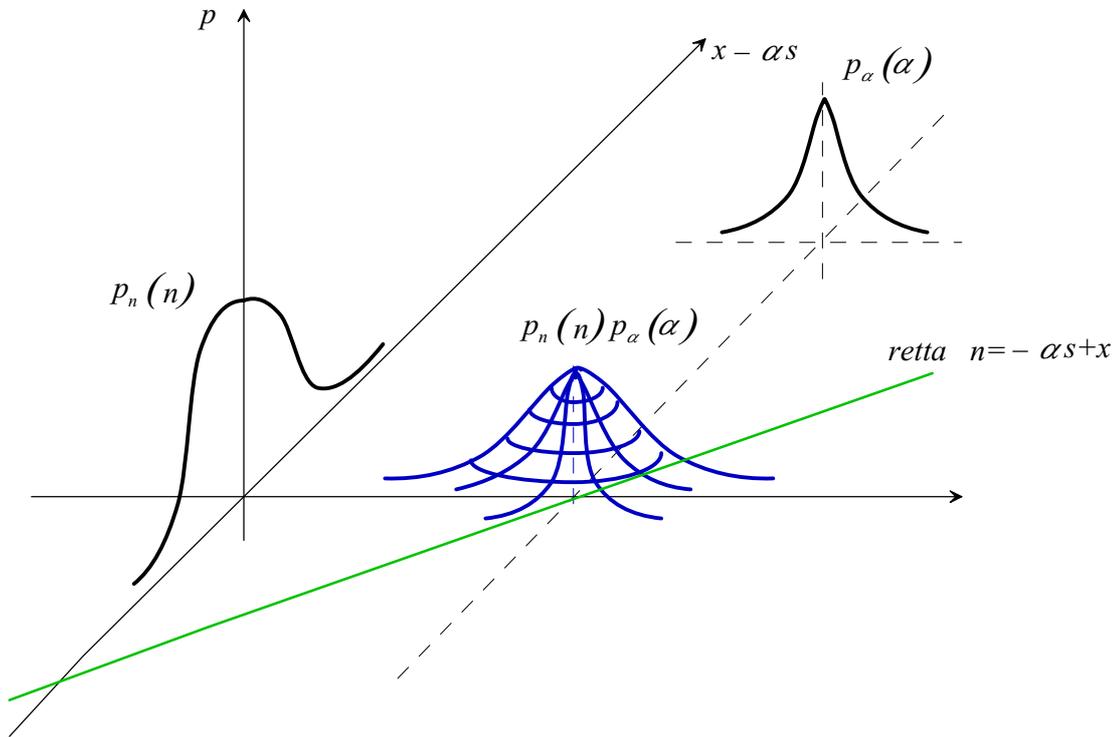


Figura 3.8: Rappresentazione tridimensionale della costruzione della densità di probabilità $p_\alpha(\alpha) = p_n(x - \alpha s)$ e sua massimizzazione.

La retta $n = -\alpha s + x$ permette di prendere in considerazione le misure svolte (l'unica misura svolta in questo caso). La distribuzione di probabilità congiunta $p(n, \alpha)$ viene tagliata dalla retta originando una nuova curva che ha solitamente varianza inferiore rispetto alla probabilità a priori $p_\alpha(\alpha)$. Per risolvere il problema occorre calcolare il massimo di questa densità sul vincolo della retta. Questo passaggio da probabilità a priori a probabilità a posteriori prende il nome di *apprendimento bayesiano*.

Nel caso in cui non si conoscesse $p_\alpha(\alpha)$, occorrerebbe utilizzare il criterio ML, in cui si assume $p_\alpha(\alpha)$ uniforme in un certo range di valori. In questo caso il prodotto tra la distribuzione gaussiana del rumore e la distribuzione costante di α , darà luogo a una sorta di “cilindro” con sezione gaussiana. Calcolando l'intersezione con la retta delle misurazioni si ricava la funzione da minimizza-

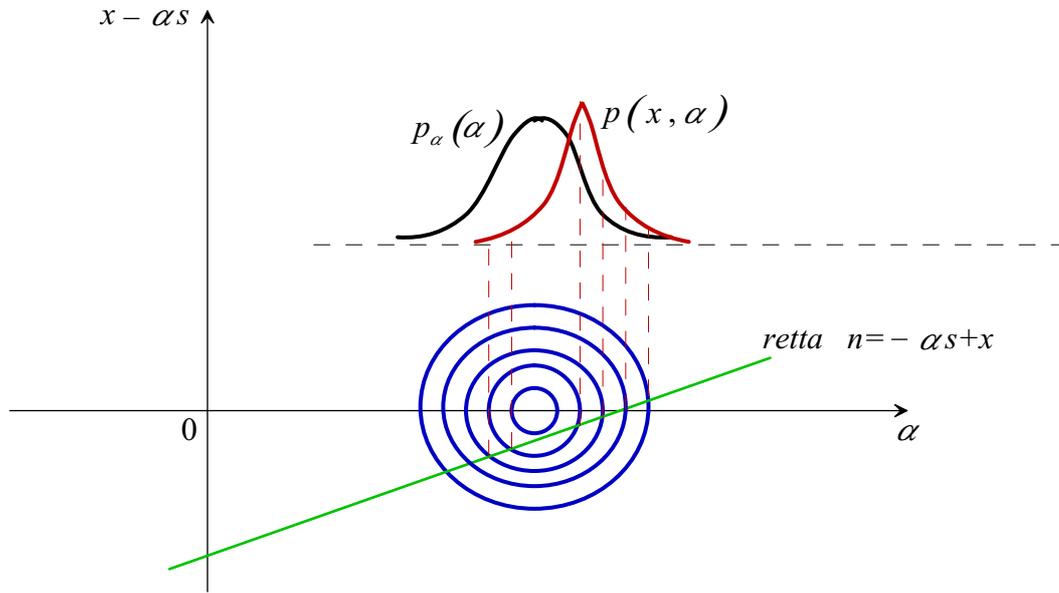


Figura 3.9: Proiezione sul piano $x, x - \alpha s$ della figura 3.8 e intersezione con il vincolo (retta $n = -\alpha s + x$).

re. Pertanto, anche utilizzando il criterio della massima verosimiglianza bisogna adottare lo stesso procedimento del criterio MAP: l'unica cosa che cambia é la funzione da massimizzare.

Il caso monodimensionale analizzato puó essere facilmente esteso al caso in cui si posseggano n misure x_i .

Osservazione sull'ipotesi di rumore gaussiano bianco

Fino ad ora abbiamo ipotizzato l'esistenza di un rumore caratterizzato da una densità di probabilità di tipo gaussiano a valor medio nullo (caratterizzato cioè da una densità di probabilità $p_n(n) = N(0, \sigma^2)$) e spettro costante su tutte le frequenze (*rumore bianco*). Questa ipotesi puó essere parzialmente modificata ipotizzando che il rumore sia *quasi bianco*, ossia abbia uno spettro costante solo nella zona di frequenza in cui il segnale possiede la maggior parte della sua energia. Ad esempio il segnale s avrà uno spettro in frequenza tale per cui una percentuale

fissata (poniamo ad esempio il 95%) dell'energia totale sarà contenuta in frequenze inferiori a un certo limite B (vedi figura 3.10).

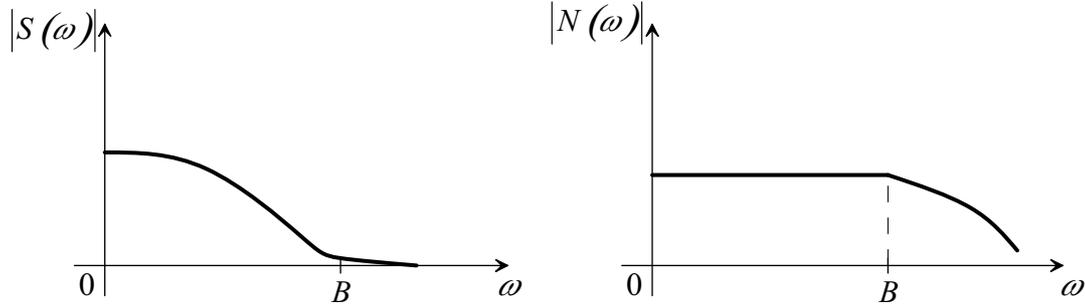


Figura 3.10: Spettro in frequenza tipico di un segnale $s(t)$.

3.6 Quadro complessivo delle varie tipologie di stima

I diversi criteri di stima visti nei paragrafi precedenti hanno origine dalla ricerca di un valore $\hat{\alpha}$ che risolva il seguente problema di ottimizzazione:

$$\min_{\hat{\alpha}} \left[\int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} L(\hat{\alpha}, \alpha) p(x_1, x_2, \dots, x_n, \alpha) dx_1 \dots dx_n d\alpha \right] \quad (3.109)$$

dove $L(\hat{\alpha}, \alpha)$ prende il nome come visto di funzione di perdita, e ha il compito di associare un costo agli errori che si commettono durante il processo di stima. La soluzione del problema espresso dalla (3.109) si può ricondurre alla soluzione del problema

$$\min_{\hat{\alpha}} \left[\left(\int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} p_x(x_1, \dots, x_n) dx_1 \dots dx_n \right) \left(\int_{-\infty}^{+\infty} L(\hat{\alpha}, \alpha) p(\alpha | x_1 \dots x_n) d\alpha \right) \right] \quad (3.110)$$

Poiché il primo termine del prodotto precedente è sempre una quantità maggiore o uguale a zero, l'unico termine che conta nel processo di minimizzazione è il secondo. Quindi il problema di ottimizzazione si riduce a

$$\min_{\hat{\alpha}} \left\{ \int_{-\infty}^{+\infty} L(\hat{\alpha}, \alpha) p(\alpha | x_1, x_2, \dots, x_n) d\alpha \right\} \quad (3.111)$$

A seconda delle informazioni che si hanno a disposizione, si possono ricavare stime diverse, che sono ottime rispetto alle informazioni che si posseggono. Ovviamente minori sono i dati a disposizione, peggiore é la stima.

In generale si possono individuare tre quantità importanti per la stima, di cui é auspicabile avere conoscenza:

- (1) la funzione di perdita L ;
- (2) la densità di probabilità a priori di α , ossia $p_\alpha(\alpha)$;
- (3) la densità di probabilità delle osservazioni, ossia $p(x_1, x_2, \dots, x_n | \alpha)$ (infatti $p(\alpha | x_1, x_2, \dots, x_n)$ si può ricavare partendo da $p_\alpha(\alpha)$ e $p(x_1, x_2, \dots, x_n | \alpha)$).

A seconda che si conoscano o meno le quantità (1), (2) e (3) si ottengono le diverse stime analizzate precedentemente. Vedremo in particolare come tutte le stime prese in considerazione possano essere considerate come casi particolari della stima bayesiana.

3.6.1 Stima bayesiana

In questo caso si suppongono note tutte le informazioni (1), (2) e (3). Scegliendo una funzione di perdita L quadratica, ossia

$$L(\hat{\alpha}_{BAYES}, \alpha) = (\alpha - \hat{\alpha})^2 \quad (3.112)$$

la stima bayesiana si può esprimere come

$$\hat{\alpha} = \int_{-\infty}^{+\infty} \alpha p(\alpha | x_1, x_2, \dots, x_n) dx \quad (3.113)$$

Questa stima viene detta stima bayesiana al minimo errore quadratico medio (anche nota come stima *least mean square*) ed é indicata con $\hat{\alpha}_{LMS}$. Ovviamente a seconda della funzione di perdita L scelta si troveranno stime diverse.

3.6.2 Stima MAP

In questo caso si perde l'informazione sulla funzione L . Quando non si possiede nessuna informazione su quali errori privilegiare, tutti gli errori sono da considerarsi ugualmente dannosi, ossia si deve ipotizzare una funzione peso costante in cui vengono penalizzati allo stesso modo tutti gli errori. Questa funzione é un caso limite della funzione di peso di tipo A introdotta nel paragrafo 3.3.1, in cui ϵ viene fatto tendere a zero. L'unico punto in cui la funzione é nulla é quello per cui $\hat{\alpha} = \alpha$, ossia il punto a errore nullo (vedi figura 3.11).

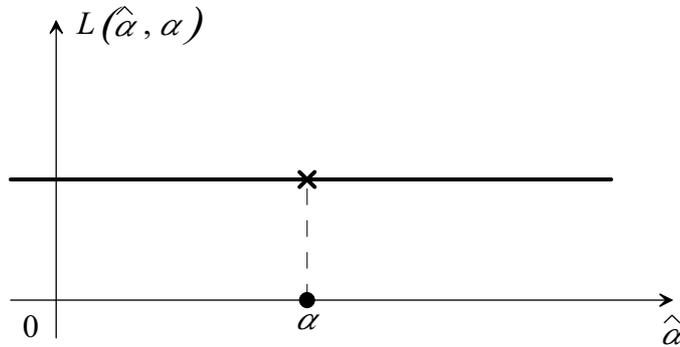


Figura 3.11: Funzione di perdita di tipo A nel caso limite in cui $\epsilon \rightarrow 0$.

Come accennato, nella stima MAP (massima probabilità a posteriori) si hanno a disposizione solo le informazioni (2) e (3). La stima può essere espressa come

$$\hat{\alpha}_{MAP} = \arg \max_{\alpha} \{p(\alpha|x_1, x_2, \dots, x_n)\} \quad (3.114)$$

3.6.3 Stima ML

Nella stima ML (massima verosimiglianza) si possiede solo l'informazione (3) in quanto anche la distribuzione a priori $p_{\alpha}(\alpha)$ non é nota. E' nota solo la densità di probabilità $p(x_1, x_2, \dots, x_n|\alpha)$ che si può ricavare a partire dalle osservazioni effettuate. La stima si può esprimere nel seguente modo:

$$\hat{\alpha}_{ML} = \arg \max_{\alpha} \{p(x_1, x_2, \dots, x_n|\alpha)\} \quad (3.115)$$

3.6.4 Stima lineare

Anche la stima lineare può essere vista come un caso particolare della stima bayesiana. Nella stima lineare MEMQ (al minimo errore quadratico medio) nessuna delle quantità (1), (2) o (3) è nota. In questo caso occorre riferirsi a informazioni statistiche di ordine inferiore, quali ad esempio la *funzione di autocorrelazione*, in quanto non è possibile risalire alla densità di probabilità. La stima lineare può essere espressa come

$$\hat{\alpha}_{MEMQ} = \sum_{i=1}^n a_i x_i \quad (3.116)$$

dove i coefficienti a_i si ricavano minimizzando l'errore quadratico medio oppure ricorrendo all'inversione della matrice di autocorrelazione R .

3.7 Applicazioni dei concetti esposti in precedenza

Consideriamo un segnale

$$x(t) = s(t, \alpha) + n(t) \quad (3.117)$$

dove $s(t, \alpha)$ è segnale di forma nota e $n(t)$ è un rumore additivo. Il parametro da stimare è ancora α . Dire che $s(t, \alpha)$ è un segnale di forma nota non significa che ne è nota con esattezza l'espressione analitica con tutti i parametri, ma solo la tipologia, essendo uno o più parametri fissati dalla quantità aleatoria costante nel tempo α . Ad esempio $s(t, \alpha)$ potrebbe essere pensato come un segnale sinusoidale (la forma del segnale è nota: una sinusoide appunto) di cui non è nota l'ampiezza (oppure la fase) perché fissata dal parametro α .

Supponiamo di avere n campioni del segnale $x(t)$ in un intervallo di durata T . I vari campioni saranno legati dalla relazione

$$x_i = s_i(\alpha) + n_i \quad i = 1, 2, \dots, n \quad (3.118)$$

Calcoliamo la stima $\hat{\alpha}$ di α nel seguente modo:

$$\hat{\alpha} = \arg \max_{\alpha} \{p(\alpha|x_1, x_2, \dots, x_n)\} = \arg \max_{\alpha} \{p_{\alpha}(\alpha)p(x_1, x_2, \dots, x_n|\alpha)\} \quad (3.119)$$

dove é stato utilizzato il teorema di Bayes.

L'espressione precedente contiene la quantità $p(x_1, x_2, \dots, x_n|\alpha)$, ossia una densità di probabilità condizionata al parametro α da stimare. Il fatto che α compaia come fattore condizionante permette di affermare che, almeno limitatamente al termine $p(x_1, x_2, \dots, x_n|\alpha)$, l'unica aleatorietà su x_i é data dal rumore n_i .

Supponendo ancora un rumore gaussiano a media nulla e varianza σ^2 , ossia supponendo

$$p_n(n) = N(0, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma}} \exp -\frac{n^2}{2\sigma^2} \quad (3.120)$$

e che tutte le osservazioni siano indipendenti tra loro, abbiamo

$$p(x_1, x_2, \dots, x_n|\alpha) = p(x_1|\alpha)p(x_2|\alpha) \cdots p(x_n|\alpha) \quad (3.121)$$

Poiché

$$p(x_i|\alpha) = p_n(x_i - s_i(\alpha)) \quad (3.122)$$

si ottiene

$$\hat{\alpha} = \arg \max_{\alpha} \left\{ p_{\alpha}(\alpha) \prod_{i=1}^n p_n(x_i - s_i(\alpha)) \right\} \quad (3.123)$$

Poiché si assume spesso (come in questo caso) che $p_n(n)$ sia una gaussiana, conviene massimizzare il logaritmo della equazione precedente. In particolare si ottiene

$$\hat{\alpha} = \arg \max_{\alpha} \left\{ \ln p_{\alpha}(\alpha) + \ln \left(\frac{1}{\sqrt{2\pi\sigma}} \right)^n - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - s_i(\alpha))^2 \right\} \quad (3.124)$$

Il termine $\ln \left(\frac{1}{\sqrt{2\pi\sigma}} \right)^n$ può essere trascurato durante il processo di minimizzazione in quanto indipendente da α . Svolgendo il quadrato nell'espressione precedente si ottiene

$$\hat{\alpha} = \arg \max_{\alpha} \left\{ \ln p_{\alpha}(\alpha) - \frac{1}{2\sigma^2} \left[\sum_{i=1}^n x_i^2 - 2 \sum_{i=1}^n x_i s_i(\alpha) + \sum_{i=1}^n s_i^2(\alpha) \right] \right\} \quad (3.125)$$

Trascurando il termine $\sum_{i=1}^n x_i^2$ che non dipende da α si ottiene

$$\hat{\alpha} = \arg \max_{\alpha} \left\{ 2\sigma^2 \ln p_{\alpha}(\alpha) + 2 \sum_{i=1}^n x_i s_i(\alpha) - \sum_{i=1}^n s_i^2(\alpha) \right\} \quad (3.126)$$

Osservazione. Fino ad ora abbiamo ragionato a tempo discreto considerando n campioni di $x(t)$, $s(t, \alpha)$ e $n(t)$. Ritornando al tempo continuo, la sommatoria $\sum_{i=1}^n x_i s_i(\alpha)$ equivale a un integrale esteso nel dominio in cui si posseggono i campioni, ossia equivale all'integrale $\int_0^T x(t) s(t, \alpha) dt$. Analogamente $\sum_{i=1}^n s_i^2(\alpha)$ corrisponde a $\int_0^T s^2(t, \alpha) dt$.

Vediamo ora alcuni esempi particolari di forme tipiche di $s(t, \alpha)$.

3.7.1 Esempio A

Consideriamo un segnale $s(t, \alpha)$ del tipo

$$s(t, \alpha) = \alpha \cos(\omega_0 t) \quad (3.127)$$

in cui l'informazione α è impressa nell'ampiezza di un segnale sinusoidale. Per ipotesi si suppone che l'intervallo di osservazione del segnale sia molto maggiore del periodo del coseno, ossia

$$T \gg T_0 = \frac{1}{f_0} = \frac{2\pi}{\omega_0} \quad (3.128)$$

Per semplicità si suppone ancora

$$p_{\alpha}(\alpha) = \text{costante} \quad (3.129)$$

Questo equivale a non avere nessuna informazione su $p_{\alpha}(\alpha)$, per cui verrà utilizzata la stima ML. Questa ipotesi permette di trascurare il termine $2\sigma^2 \ln p_{\alpha}(\alpha)$ nell'equazione (3.126) in quanto costante e indipendente da α . Resta quindi da calcolare

$$\hat{\alpha}_{ML} = \arg \max_{\alpha} \left\{ 2 \int_0^T x(t) \alpha \cos(\omega_0 t) dt - \int_0^T \alpha^2 \cos^2(\omega_0 t) dt \right\} \quad (3.130)$$

Calcoliamo la derivata rispetto a α della funzione tra parentesi graffe nell'espressione precedente e uguagliamola a zero al fine di trovare $\hat{\alpha}$:

$$\frac{d}{d\alpha} \left[2 \int_0^T x(t) \alpha \cos(\omega_0 t) dt - \int_0^T \alpha^2 \cos^2(\omega_0 t) dt \right] = 0 \quad (3.131)$$

Svolgendo i calcoli si ottiene

$$2 \int_0^T x(t) \cos(\omega_0 t) dt - 2\alpha \int_0^T \cos^2(\omega_0 t) dt = 0 \quad (3.132)$$

da cui si ricava

$$\hat{\alpha}_{ML} = \frac{\int_0^T x(t) \cos(\omega_0 t) dt}{\int_0^T \cos^2(\omega_0 t) dt} = \frac{\int_0^T x(t) \cos(\omega_0 t) dt}{\frac{T}{2} - \frac{1}{4\omega_0} \sin(2\omega_0 T)} \simeq \frac{\int_0^T x(t) \cos(\omega_0 t) dt}{\frac{T}{2}} \quad (3.133)$$

Abbiamo ricavato quindi un modo operativo per estrarre l'informazione α dal segnale. Lo schema a blocchi per stimare α è mostrato in figura 3.12.

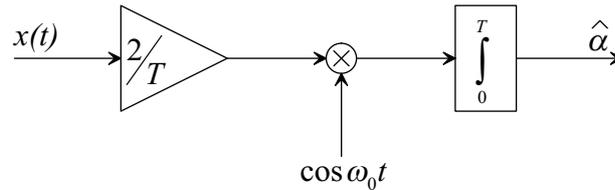


Figura 3.12: Schema a blocchi per il calcolo di $\hat{\alpha}$ nell'esempio A.

3.7.2 Esempio B

Consideriamo un segnale $s(t, \alpha)$ del tipo

$$s(t, \alpha) = \cos(\omega_0 t + \alpha) \quad (3.134)$$

in cui l'informazione α è inserita nella fase.

Supponiamo ancora $T \gg T_0 = \frac{1}{f_0}$ e $p_\alpha(\alpha) = \text{costante}$, ossia consideriamo ancora la stima alla massima verosimiglianza. In questo caso la stima $\hat{\alpha}$ si ricava risolvendo il problema

$$\hat{\alpha}_{ML} = \arg \max_{\alpha} \left\{ 2 \int_0^T x(t) \cos(\omega_0 t + \alpha) dt - \int_0^T \cos^2(\omega_0 t + \alpha) dt \right\} \quad (3.135)$$

Seguendo lo stesso procedimento adottato per il caso precedente calcoliamo la derivata della quantità tra parentesi graffe e uguagliamola a zero:

$$\frac{d}{d\alpha} \left[2 \int_0^T x(t) \cos(\omega_0 t + \alpha) dt - \int_0^T \cos^2(\omega_0 t + \alpha) dt \right] = 0 \quad (3.136)$$

Poiché abbiamo ipotizzato $T \gg T_0 = \frac{1}{f_0}$, l'ultimo integrale nell'equazione (3.136) può essere considerato indipendente dal valore di α . Infatti tale termine è l'integrale di un coseno avente un numero di oscillazioni molto elevato all'interno dell'intervallo di integrazione $[0, T]$ (vedi figura 3.13). Per questo motivo un piccolo spostamento dovuto al contributo di α nella fase non provoca un cambiamento significativo del valore dell'integrale.

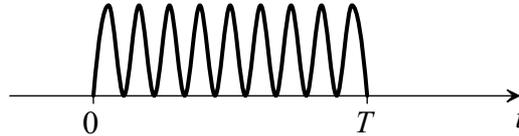


Figura 3.13: Oscillazioni continue del segnale s in un periodo T

Inoltre, considerando che

$$\cos(\omega_0 t + \alpha) = \cos(\omega_0 t) \cos(\alpha) - \sin(\omega_0 t) \sin(\alpha) \quad (3.137)$$

e svolgendo la derivata nella (3.136) si ha

$$\begin{aligned} \frac{d}{d\alpha} \left[2 \int_0^T x(t) \cos(\omega_0 t + \alpha) dt - \int_0^T \cos^2(\omega_0 t + \alpha) dt \right] &\simeq \\ &\simeq -2 \sin(\alpha) \int_0^T x(t) \cos(\omega_0 t) dt - 2 \cos(\alpha) \int_0^T x(t) \sin(\omega_0 t) dt = 0 \end{aligned} \quad (3.138)$$

Pertanto, si ricava:

$$\hat{\alpha}_{ML} = -\arctan \frac{\int_0^T x(t) \sin(\omega_0 t) dt}{\int_0^T x(t) \cos(\omega_0 t) dt} \pm \pi \quad (3.139)$$

Il termine finale $\pm\pi$ è necessario in quanto l'arcotangente fornisce l'informazione di fase a meno di un periodo.

Anche in questo caso è possibile schematizzare il procedimento di estrazione dell'informazione da $x(t)$ con un opportuno schema a blocchi, riportato in figura 3.14.

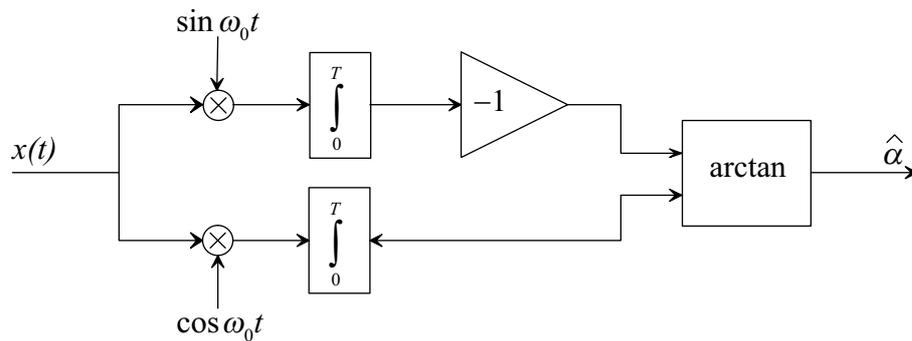


Figura 3.14: Schema a blocchi per il calcolo di $\hat{\alpha}$ nell'esempio B.

3.7.3 Esempio C

Consideriamo un segnale $s(t, \alpha)$ del tipo

$$s(t, \alpha) = s_0(t - \alpha) \quad (3.140)$$

in cui l'informazione che si vuole conoscere é un ritardo.

Questo problema si ritrova ad esempio nelle applicazioni radar, in cui viene trasmesso un segnale noto verso un corpo posto a una distanza incognita. Misurando il tempo che trascorre dall'invio dell'impulso all'istante in cui l'impulso stesso riflesso dal corpo viene ricevuto dal radar, é possibile ricavare l'informazione della distanza del corpo dalla sorgente.

Per ipotesi assumiamo un impulso di durata $T_0 \ll T$, dove T rappresenta l'intervallo di osservazione del fenomeno (si vedano le figure 3.15 e 3.16).

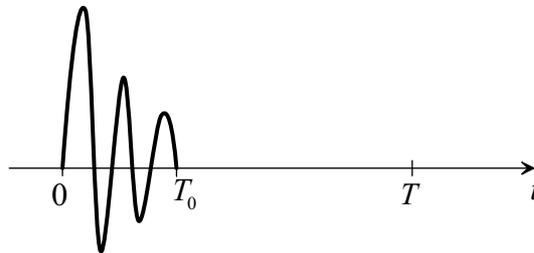


Figura 3.15: Impulso di durata $T_0 \ll T$.

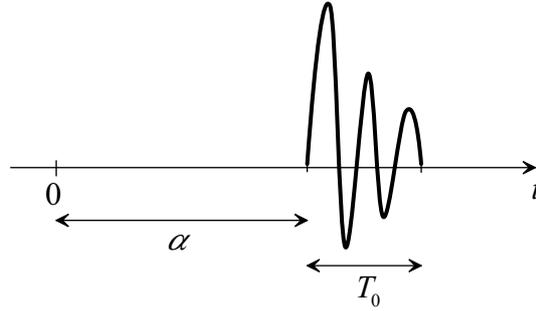


Figura 3.16: Impulso di durata $T_0 \ll T$ ritardato di una quantità α .

Anche in questo caso consideriamo $p_\alpha(\alpha) = \text{costante}$ (stima ML) e cerchiamo

$$\hat{\alpha}_{ML} = \arg \max_{\alpha} \left\{ 2 \int_0^T x(t) s_0(t - \alpha) dt - \int_0^T s_0^2(t - \alpha) dt \right\} \quad (3.141)$$

Conviene osservare che poiché $T_0 \ll T$, il termine s_0^2 non cambia significativamente se si sposta leggermente l'impulso di una quantità α . Pertanto il termine $\int_0^T s_0^2(t - \alpha) dt$ può essere considerato indipendente da α ed essere quindi trascurato nel processo di minimizzazione. Si tratta allora di risolvere il problema seguente:

$$\hat{\alpha}_{ML} = \arg \max_{\alpha} \left\{ \int_0^T x(t) s_0(t - \alpha) dt \right\} \quad (3.142)$$

Cerchiamo di ricavare $\hat{\alpha}_{ML}$ attraverso una *operazione di filtraggio*. In altre parole cerchiamo il filtro che fornisce $\hat{\alpha}_{ML}$.

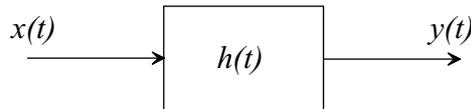


Figura 3.17: Operazione di filtraggio per la ricerca della stima ottima $\hat{\alpha}_{ML}$.

Tale filtro sarà caratterizzato da una risposta all'impulso $h(t)$ del tipo

$$h(t) = s_0(T_0 - t) \quad (3.143)$$

L'uscita $y(t)$ del filtro è la convoluzione dell'ingresso con la risposta all'impulso

del filtro, ossia

$$y(t) = \int_0^T x(\xi)h(t - \xi)d\xi = \int_0^T x(\xi)s_0(T_0 - t + \xi)d\xi \quad (3.144)$$

Questa quantità é molto simile alla quantità da minimizzare nell'operazione (3.142). Infatti l'integrale

$$\int_0^T x(\xi)s_0[\xi - (t - T_0)]d\xi \quad (3.145)$$

é formalmente identico a

$$\int_0^T x(t)s_0(t - \alpha)dt \quad (3.146)$$

a patto di far corrispondere $\xi \rightarrow t$ e $(t - T_0) \rightarrow \alpha$.

A partire dall'equazione (3.144) é possibile ricavare il valore t_{MAX} tale per cui $y(t)$ é massima. Trovato questo valore é possibile ricavare la stima cercata ponendo

$$\hat{\alpha}_{ML} = t_{MAX} - T_0 \quad (3.147)$$

Il filtro utilizzato in questo caso prende il nome di *filtro adattato*.

3.8 Rappresentazioni distribuite di una misura

Il concetto di stima puó essere utilizzato per generalizzare il concetto di misura. Solitamente siamo abituati a pensare a sistema di misura come uno o piú trasduttori che operano una trasformazione dalla grandezza oggetto della misura a un'altra che puó essere facilmente maneggiata. Ad esempio un amperometro effettua una trasformazione dalla grandezza corrente allo spostamento di un indice su una scala graduata. Il trasduttore traduce ogni valore di corrente in una posizione diversa dell'indice sulla scala.

Detta α la grandezza incognita da misurare, il trasduttore effettuerá la trasformazione in un'altra grandezza nota utilizzando una certa funzione di trasduzione (vedi figura 3.18).

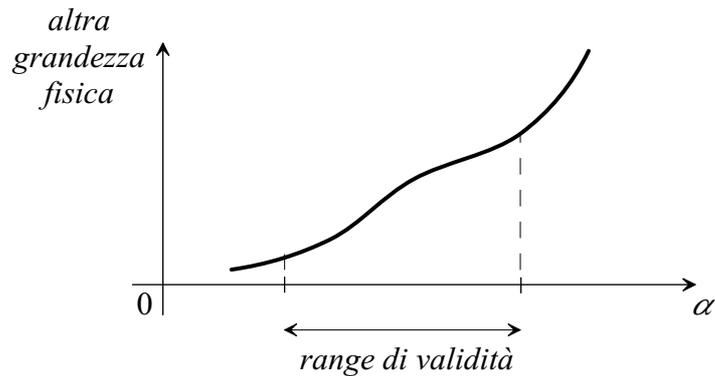


Figura 3.18: Il trasduttore fa corrispondere mediante una opportuna funzione di trasformazione ogni valore di α in un diverso valore di un'altra grandezza fisica.

In alcuni casi però potrebbe non esistere in natura un trasduttore capace di fornire la trasformazione desiderata. Inoltre la relazione di trasformazione potrebbe non essere monotona, quindi ci sarebbero casi in cui uno stesso valore della grandezza fisica in ordinata corrisponde a diversi valori α_1 e α_2 della quantità da misurare, rendendo impossibile il processo di misurazione (vedi figura 3.19).

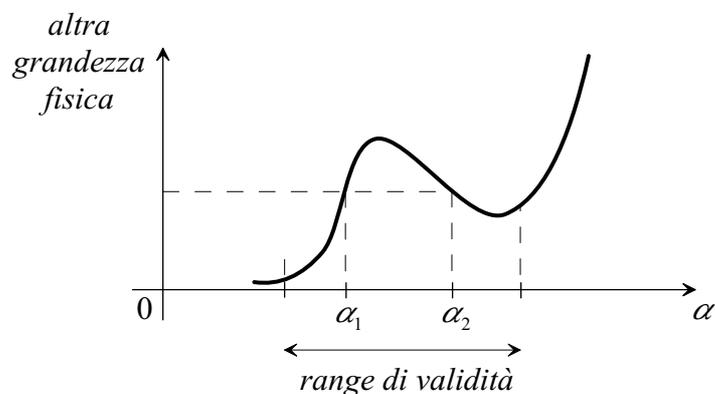


Figura 3.19: Funzione di trasduzione non monotona. In questo caso valori di α diversi forniscono lo stesso risultato della misurazione.

E' evidente come in casi del genere non sia possibile costruire un sistema di misura con dei semplici trasduttori. In molti casi si utilizzano dei *rivelatori* al posto dei trasduttori.

Un semplice schema di misura consiste nel prevedere una serie di rivelatori ognuno con una propria regione di sensibilità (*cella*). Quando la grandezza α da misurare si trova all'interno del range di valori corrispondente alla zona di sensibilità di un particolare rivelatore, il rivelatore interessato risponde diventando attivo, mentre tutti gli altri si mettono in una condizione di inattività. Andando a vedere quale rivelatore ha risposto é possibile risalire al valore assunto dalla grandezza oggetto della misurazione. Ogni rivelatore si occupa esclusivamente della propria cella, senza influenzare i rivelatori vicini.

Realizzare un sistema di misura in questo modo comporta una serie di problemi, primo fra tutti quello della *risoluzione*: Per avere una elevata precisione nella misura occorre avere un numero di rivelatori molto alto. Strettamente collegato a questo problema é anche quello del *costo*: come detto per avere una buona risoluzione occorre un gran numero di sensori, di cui solamente uno per ogni misurazione é attivo. Si ha quindi una grande quantità di risorse che per la maggior parte del tempo non sono utilizzate e occupano spazio.

In alternativa, si potrebbe usare una rappresentazione distribuita a “cellule”, in cui i rivelatori possiedono una certa *curva o funzione di sensibilità* che si estende anche a zone di competenza di altri rivelatori. In altre parole esiste una sovrapposizione tra le varie funzioni di sensibilità che fa in modo che, con diverso peso, tutti i rivelatori diano il loro contributo nel processo di misura.

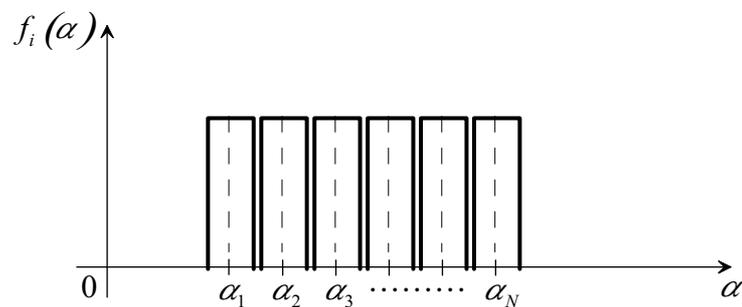


Figura 3.20: Curve di sensibilità di una serie di rivelatori. In questo caso ogni rivelatore risponde solo nella sua zona di competenza.

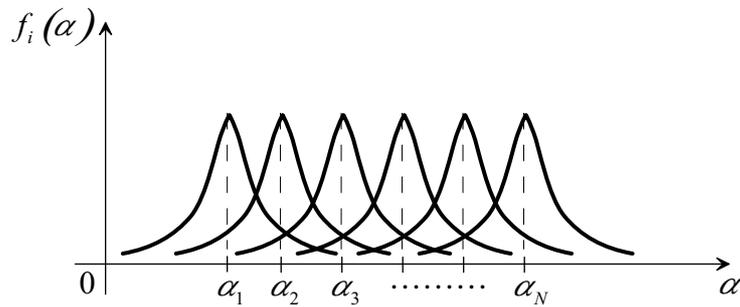


Figura 3.21: Curve di sensibilità di una serie di rivelatori. Si noti la sovrapposizione tra le varie curve.

Grazie alla forte sovrapposizione delle curve di sensibilità, la finestra di osservazione (regione di sensibilità) di ogni rivelatore è potenzialmente infinita. Questa soluzione prende il nome di *codifica di popolazione* in quanto per ogni misura risponde l'intera popolazione di rivelatori e non uno solo, come accadeva nel caso di curve di sensibilità non sovrapposte in cui ogni rivelatore era responsabile solo di una determinata cella.

I vantaggi di questa soluzione sono soprattutto la risoluzione elevata, in quanto è possibile combinare con la teoria della stima le risposte di tutti i rivelatori. Inoltre grazie alla sovrapposizione è possibile utilizzare un numero inferiore di rivelatori. Un ulteriore vantaggio risiede nella robustezza. Infatti grazie alla sovrapposizione delle curve di sensibilità, nel caso di guasto di un rivelatore non si avrebbe una zona oscura in cui il sistema di misura non risponderebbe (come nel caso di rivelatori dedicati a una sola cella) in quanto si avrebbe comunque il contributo degli altri rivelatori.

3.8.1 Utilizzo di rivelatori per la misura di un parametro

Scendiamo ora leggermente più nel dettaglio di come usare questa tecnica per stimare il parametro α .

Consideriamo N rivelatori accordati (sintonizzati) su N valori diversi del

parametro α ($\alpha_1, \dots, \alpha_N$). Ogni rivelatore sarà caratterizzato da una risposta

$$x_i = f_i(\alpha) + n_i \quad (3.148)$$

dove n_i è il rumore additivo di canale che si aggiunge inevitabilmente in seguito alla misurazione.

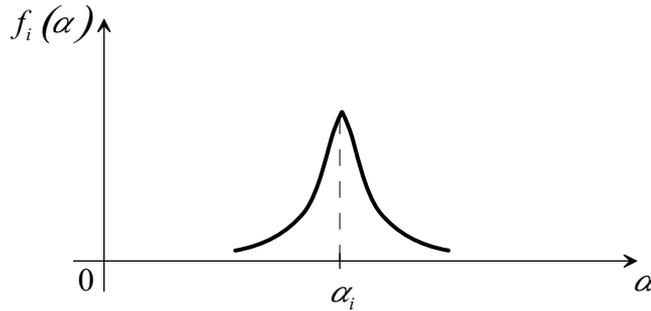


Figura 3.22: Curva di sensibilità di un rivelatore centrato sul valore α_i .

Tipicamente $f_i(\alpha)$ è una gaussiana con valor medio α_i e varianza σ_i^2

$$f_i(\alpha) = x_{MAX} \exp \left[-\frac{(\alpha - \alpha_i)^2}{2\sigma_i^2} \right] \quad (3.149)$$

dove x_{MAX} è il valore massimo di x .

Osservazione: Nel caso si abbia una densità elevata di campane ugualmente spaziate,

$$\sum_{i=1}^n f_i(\alpha) \simeq \text{costante} \quad (3.150)$$

La sommatoria diventa quindi indipendente da α .

Ricaviamo ora la stima $\hat{\alpha}_{ML}$ usando il criterio della massima verosimiglianza, che in questo caso coincide con il criterio della massima probabilità a posteriori. Per ipotesi, dato che trattiamo una stima ML, supponiamo $p_\alpha(\alpha) = \text{costante}$ e ricaviamo $p(x_i|\alpha)$. Con ragionamenti analoghi a quanto fatto in precedenza, la probabilità condizionata con α che compare come termine condizionante permette di considerare il rumore l'unica fonte di aleatorietà di x_i .

In questo caso conviene usare una distribuzione poissoniana del rumore invece che la solita distribuzione gaussiana. Avremo allora

$$p(x_i|\alpha) = \frac{1}{x_i!} \exp[-f_i(\alpha)] f_i(\alpha)^{x_i} \quad (3.151)$$

dove si ipotizza che x_i assuma valori interi. Questo é ragionevole nel caso in cui si quantizzi la variabile x_i .

Se tutti i rivelatori sono indipendenti tra loro abbiamo

$$p(x_1, x_2, \dots, x_N|\alpha) = \prod_{i=1}^N \frac{f_i(\alpha)^{x_i}}{x_i!} \exp[-f_i(\alpha)] \quad (3.152)$$

Dobbiamo trovare il valore $\hat{\alpha}_{ML}$ tale che

$$\hat{\alpha}_{ML} = \arg \max_{\alpha} \{p(\mathbf{x}|\alpha)\} \quad (3.153)$$

Tale valore é ricavabile a partire da ragionamenti del tutto analoghi a quelli fatti nei paragrafi precedenti. Poiché si ha a che fare anche in questo caso con degli esponenziali, si procede con la minimizzazione del logaritmo della quantità tra parentesi graffe nell'equazione precedente. In particolare si ottiene

$$\ln p(x_1, x_2, \dots, x_N|\alpha) = \sum_{i=1}^N x_i \ln f_i(\alpha) - \sum_{i=1}^N \ln(x_i!) - \sum_{i=1}^N f_i(\alpha) \quad (3.154)$$

dove $\sum_{i=1}^N \ln(x_i!)$ non dipende da α , e $\sum_{i=1}^N f_i(\alpha)$ é la somma delle varie funzioni di sensibilità, e quindi può essere considerato approssimativamente costante. Questi termini quindi non hanno alcuna influenza nel processo di minimizzazione, e possono quindi essere trascurati.

Calcolando la derivata della funzione da minimizzare e uguagliandola a zero si ottiene:

$$\frac{d}{d\alpha} \ln p(x_1, x_2, \dots, x_N|\alpha) = \sum_{i=1}^N x_i \frac{f_i'(\alpha)}{f_i(\alpha)} = \sum_{i=1}^N x_i \frac{\alpha_i - \alpha}{\sigma_i^2} = 0 \quad (3.155)$$

da cui si ricava

$$\hat{\alpha}_{ML} = \frac{\sum_{i=1}^N x_i \frac{\alpha_i}{\sigma_i^2}}{\sum_{i=1}^N \frac{x_i}{\sigma_i^2}} \quad (3.156)$$

Nell'ipotesi in cui i vari rivelatori abbiano funzioni di sensibilità caratterizzate tutte dalla stessa varianza σ_i^2 l'equazione precedente si riduce a

$$\hat{\alpha}_{ML} = \frac{\sum_{i=1}^N x_i \alpha_i}{\sum_{i=1}^N x_i} \quad (3.157)$$

Questa espressione fornisce un risultato abbastanza intuitivo: $\hat{\alpha}_{ML}$ è il centro di massa delle risposte x_i dell'intera popolazione di rivelatori. Ogni rivelatore contribuisce alla determinazione del valore $\hat{\alpha}_{ML}$ con un peso diverso a seconda del valore su cui è centrata la propria funzione di sensibilità.

Capitolo 4

Stima dinamica

4.1 Stima dello stato

Nel Capitolo 3 é stato discusso il problema di stima statico, che riguardava la stima di parametri. In questo capitolo affronteremo il problema di stima dinamico, che tratta la stima dello stato di un sistema, una quantità variabile nel tempo. A differenza dei parametri, che possono essere grandezze scalari o vettoriali, lo stato di un sistema é generalmente un vettore.

4.1.1 Principi fondamentali della teoria della stima dinamica

Il problema della stima dinamica può essere visto come un problema di stima di un processo stocastico. Si possono impostare due tipi di formulazione:

1. *Qualificazione statistica del processo stocastico*, il cui scopo é conoscere la statistica del processo stocastico. In questo caso ci si accontenta ad esempio della conoscenza della funzione di autocorrelazione o dello spettro di densità di potenza, senza voler conoscere il valore del segnale a ogni istante di tempo;

2. *Stima del processo ad ogni istante t* , per cui si vuole conoscere il valore del segnale a ogni istante, e non solo alcune sue caratteristiche statistiche. In questo caso possono presentarsi due situazioni, e quindi due tipi di problemi, a seconda che il processo sia stazionario o meno:

- *Problema di Wiener-Hopf*, in cui il processo é stazionario e, per ipotesi, si considera nota la funzione di autocorrelazione;
- *Problema di Kalman-Bucy*, in cui il processo puó non essere stazionario e non é possibile calcolarne la funzione di autocorrelazione.

4.2 Qualificazione statistica di un processo stocastico

Come detto, questo problema di stima dinamica consiste nella stima della statistica di un processo stocastico. Si vuole calcolare la funzione di autocorrelazione e lo spettro di densitá di potenza. Per ipotesi il processo stocastico é stazionario, quindi la funzione di autocorrelazione non dipende dai tempi assoluti t_1 e t_2 delle realizzazioni ma solo dalla loro differenza $\tau = t_1 - t_2$. Si suppone inoltre che il processo sia ergodico. Se si conoscesse totalmente il processo, ovvero si conoscesse la sua densitá di probabilitá, si potrebbero calcolare la funzione di autocorrelazione $R_x(\tau)$ e lo spettro di densitá di potenza $G_x(\omega)$ direttamente a partire dalle loro definizioni. Poiché nella realtá non si conosce tutto il processo, é possibile dare una stima della funzione di autocorrelazione e dello spettro di densitá di potenza, a partire dall'osservazione di una realizzazione di durata T .

4.2.1 Criteri di bontá della stima

Si puó valutare la bontá della stima considerando la polarizzazione (valore medio) e la stabilitá (varianza), in analogia a quanto giá visto per la stima di

parametri (vedi paragrafo 3.1.2). La stima si dice *non polarizzata* se si verifica

$$E \left\{ \hat{R}_x(\tau) \right\} = R_x(\tau) \quad (4.1)$$

Per valutare la stabilità si deve osservare quanto è dispersa la stima attorno al valore medio. La stima si dice *stabile* se si verifica

$$\lim_{T \rightarrow \infty} \sigma_{\hat{R}_x(\tau)}^2 = \lim_{T \rightarrow \infty} E \left\{ \left[\hat{R}_x(\tau) - E \left\{ \hat{R}_x(\tau) \right\} \right]^2 \right\} = 0 \quad (4.2)$$

dove T è la durata dell'osservazione del processo.

4.2.2 Stima della funzione di autocorrelazione

Si vuole ricavare la funzione di autocorrelazione a partire dall'osservazione del processo stocastico. Per un processo ergodico, data una finestra di osservazione di durata $2T$ (vedi figura 4.1), vale la seguente relazione

$$R_x(\tau) = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^{+T} x(t)x(t+\tau)dt \quad (4.3)$$

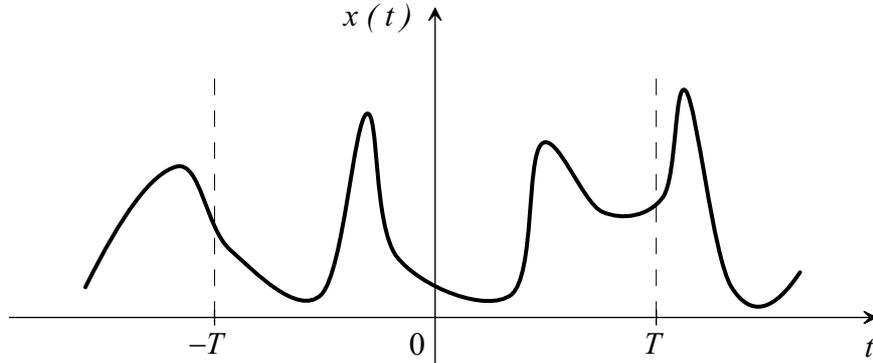


Figura 4.1: Realizzazione del processo stocastico $x(t)$ osservata in una finestra temporale di durata $2T$.

Una possibile stima $\hat{R}_x(\tau)$ di $R_x(\tau)$, può dunque essere

$$\hat{R}_x(\tau) = \frac{1}{2T} \int_{-T}^{+T} x(t)x(t+\tau)dt \quad (4.4)$$

Verifichiamo ora la polarizzazione di questa stima. Consideriamo il valore atteso della stima

$$E \left\{ \hat{R}_x(\tau) \right\} = E \left\{ \frac{1}{2T} \int_{-T}^{+T} x(t)x(t+\tau)dt \right\} \quad (4.5)$$

Poiché l'expectation é un'operatore lineare, si puó invertire il segno di integrazione con quello di expectation, ottenendo

$$E \left\{ \hat{R}_x(\tau) \right\} = \frac{1}{2T} \int_{-T}^{+T} E \{x(t)x(t+\tau)\} dt \quad (4.6)$$

La definizione di autocorrelazione di un segnale stocastico stazionario é

$$R(\tau) = E \{x(t)x(t+\tau)\} \quad (4.7)$$

Sostituendo nella relazione precedente si ottiene

$$\begin{aligned} E \left\{ \hat{R}_x(\tau) \right\} &= \frac{1}{2T} \int_{-T}^{+T} E \{x(t)x(t+\tau)\} dt = \frac{1}{2T} \int_{-T}^{+T} R(\tau)dt = \\ &= \frac{1}{2T} R(\tau) \int_{-T}^{+T} dt = \frac{2T}{2T} R(\tau) = R(\tau) \end{aligned} \quad (4.8)$$

Si conclude quindi che la polarizzazione di questa stima é nulla.

Occorre però osservare che il segnale é noto solo nell'intervallo $2T$. Al di fuori di questo intervallo non abbiamo alcuna informazione, quindi possiamo considerare nullo sia il segnale che la stima della funzione di autocorrelazione (vedi figura 4.2).

Alla luce di quanto detto, il prodotto $x(t)x(t+\tau)$ é nullo per $|\tau| > 2T$.

Pertanto si puó ottenere un'altra espressione per la stima della funzione di autocorrelazione cambiando gli estremi di integrazione

$$\hat{R}_x(\tau) = \frac{1}{2T} \int_{-(T-|\tau|)}^{+T} x(t)x(t+\tau)dt \quad (4.9)$$

Verifichiamo anche in questo caso la polarizzazione di questa stima. Consideriamo il valore atteso

$$E \left\{ \hat{R}_x(\tau) \right\} = E \left\{ \frac{1}{2T} \int_{-(T-|\tau|)}^{+T} x(t)x(t+\tau)dt \right\} \quad (4.10)$$

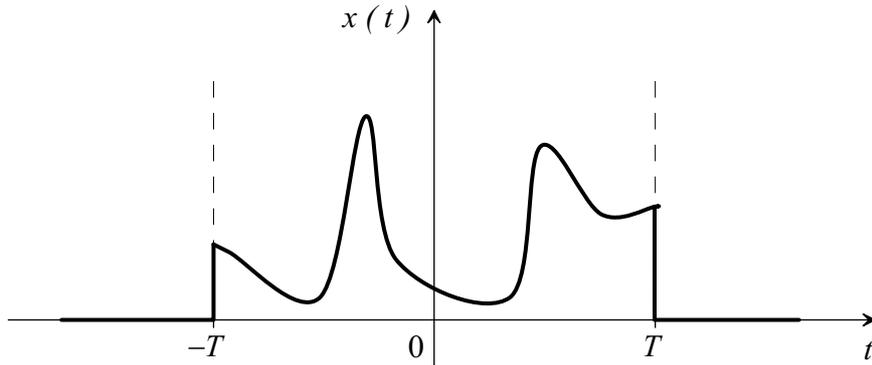


Figura 4.2: Realizzazione del processo stocastico $x(t)$ osservata in una finestra temporale di durata $2T$. All'esterno della finestra il segnale non è noto, quindi si suppone che esso assuma il valore zero.

Poiché l'expectation è un'operatore lineare, si può invertire il segno di integrazione con quello di expectation, ottenendo

$$\frac{1}{2T} \int_{-(T-|\tau|)}^{+T} E \{x(t)x(t+\tau)\} dt \quad (4.11)$$

Sfruttando la definizione di autocorrelazione di un segnale stocastico stazionario, e sostituendo nella relazione precedente si ottiene

$$\begin{aligned} E \{ \hat{R}_x(\tau) \} &= \frac{1}{2T} \int_{-(T+|\tau|)}^{+T} E \{x(t)x(t+\tau)\} dt = \frac{1}{2T} \int_{-(T+|\tau|)}^{+T} R(\tau) dt = \\ &= \frac{1}{2T} R(\tau) \int_{-(T+|\tau|)}^{+T} dt = \frac{2T-|\tau|}{2T} R(\tau) \end{aligned} \quad (4.12)$$

L'espressione è valida per $|\tau| \leq 2T$. Sinteticamente possiamo scrivere

$$E \{ \hat{R}_x(\tau) \} = \begin{cases} \frac{2T-|\tau|}{2T} R(\tau) & \text{se } |\tau| \leq 2T \\ 0 & \text{se } |\tau| > 2T \end{cases} \quad (4.13)$$

Si vede come ci sia una polarizzazione dovuta alla presenza del termine $\frac{2T-|\tau|}{2T}$, che funge da fattore di correzione, in quanto $E \{ \hat{R}_x(\tau) \} \neq R_x(\tau)$.

Poiché il fattore di correzione è un termine puramente proporzionale, è possibile rendere la stima non polarizzata riducendo l'intervallo di osservazione del segnale, ossia considerando solamente l'intervallo dove si possiede effettivamente

il segnale invece che tutto l'intervallo di durata $2T$. Con queste considerazioni si ottiene

$$\hat{R}_x(\tau) = \frac{1}{2T - |\tau|} \int_{-(T-|\tau|)}^{+T} x(t)x(t+\tau)dt \quad (4.14)$$

In questo modo la stima, almeno all'interno dell'intervallo $|\tau| \leq 2T$, é buona perché non polarizzata.

Osservazione

Poiché $R_x(\tau)$ é una funzione pari (simmetrica rispetto all'asse delle ordinate), non ha senso usare dei punti per stimare una parte che si può ricavare per simmetria.

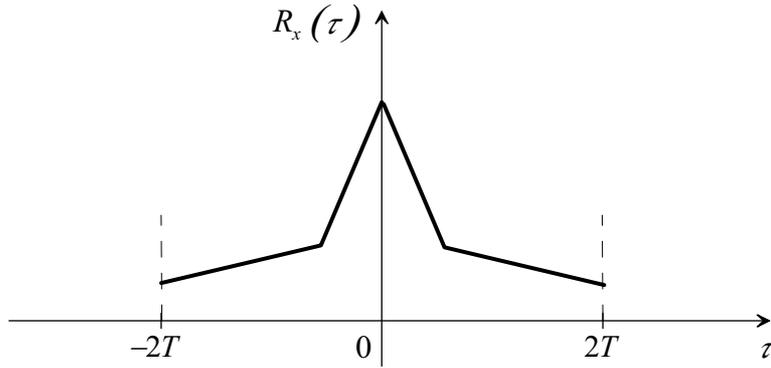


Figura 4.3: La funzione di autocorrelazione di un processo stocastico é una funzione pari.

Convien quindi porre l'origine del sistema di riferimento in un estremo dell'intervallo di osservazione, come mostrato in figura 4.4.

In questa situazione si ricava la seguente espressione:

$$\hat{R}_x(\tau) = \frac{1}{T - |\tau|} \int_{|\tau|}^T x(t)x(t+\tau)dt \quad (4.15)$$

La polarizzazione della stima é nulla, infatti

$$\begin{aligned} E \left\{ \hat{R}_x(\tau) \right\} &= \frac{1}{T - |\tau|} \int_{|\tau|}^T E \{ x(t)x(t+\tau) \} dt = \frac{1}{T - |\tau|} \int_{|\tau|}^T R(\tau) dt = \\ &= \frac{1}{T - |\tau|} R(\tau) \int_{|\tau|}^T dt = \frac{T - |\tau|}{T - |\tau|} R(\tau) = R(\tau) \end{aligned} \quad (4.16)$$

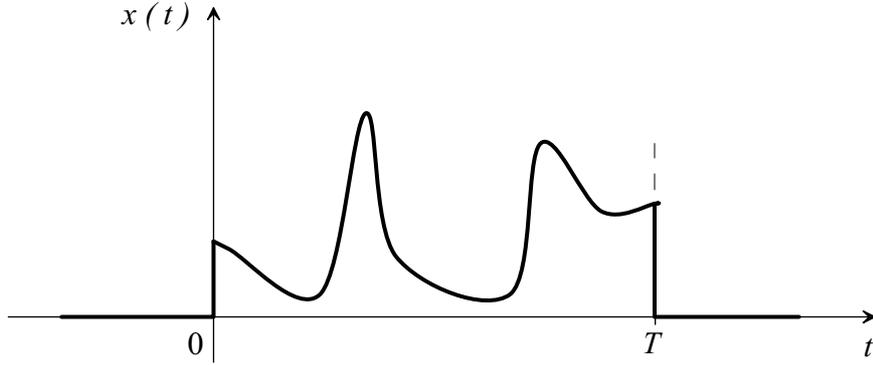


Figura 4.4: Processo stocastico osservato nell'intervallo $[0, T]$.

Valutiamo ora la stabilit  delle stime considerate fino a questo momento. In generale ricavare informazioni analitiche della varianza della stima della funzione di autocorrelazione   molto complesso, anche nel caso di ipotesi semplificative.

Si pu  ricavare la seguente espressione approssimata per la stima non polarizzata (4.4):

$$\sigma_{\hat{R}_x(\tau), NP}^2 \simeq \frac{2T}{(2T - |\tau|)^2} \int_{-\infty}^{+\infty} [R_x^2(t) + R_x(t + \tau)R_x(t - \tau)]^2 dt \quad (4.17)$$

L'approssimazione vale solamente per $T \gg |\tau|$.

Si pu  notare che in queste condizioni la stima non polarizzata   una stima asintoticamente stabile

$$\lim_{T \rightarrow \infty} \sigma_{\hat{R}_x(\tau)}^2 = 0 \quad (4.18)$$

Per la stima polarizzata (4.9), si ricava

$$\begin{aligned} \sigma_{\hat{R}_x(\tau), P}^2 &= \frac{(2T - |\tau|)^2}{(2T)^2} \sigma_{\hat{R}_x(\tau), NP}^2 \simeq \\ &\simeq \frac{(2T - |\tau|)^2}{(2T)^2} \frac{2T}{(2T - |\tau|)^2} \int_{-\infty}^{+\infty} [R_x^2(t) + R_x(t + \tau)R_x(t - \tau)]^2 dt \end{aligned} \quad (4.19)$$

Quindi

$$\sigma_{\hat{R}_x(\tau), P}^2 \simeq \frac{1}{2T} \int_{-\infty}^{+\infty} [R_x^2(t) + R_x(t + \tau)R_x(t - \tau)]^2 dt \quad (4.20)$$

Come in precedenza l'approssimazione vale per $T \gg |\tau|$.

Si può notare che

$$\lim_{T \rightarrow \infty} \sigma_{\hat{R}_x(\tau), P}^2 = 0 \quad (4.21)$$

Pertanto anche la stima polarizzata sia una stima asintoticamente stabile.

Nel caso in cui non sia più valida l'ipotesi $T \gg |\tau|$, la stima non polarizzata tende però ad avere problemi di stabilità. Infatti la varianza della stima della funzione di autocorrelazione presenta il termine $\frac{1}{2T-|\tau|}$ che dipende da τ : se $T \simeq \tau$ la varianza tende a crescere, e quindi la stima peggiora. La stima polarizzata, invece, non risente di questo problema in quanto presenta un fattore $\frac{1}{2T}$ indipendente da τ . Si può concludere che la stima non polarizzata è una buona stima per quanto riguarda la polarizzazione, mentre è una cattiva stima se si considera la stabilità. Ragionamenti inversi devono essere fatti per la stima polarizzata.

Un criterio spesso utilizzato per semplificare il confronto tra due stimatori tiene conto contemporaneamente della polarizzazione e della stabilità e si basa sul calcolo dell'errore quadratico medio:

$$E \left\{ \left[\hat{R}_x(\tau) - R_x(\tau) \right]^2 \right\} = \sigma_{\hat{R}_x(\tau)}^2 + b^2 \quad (4.22)$$

dove

$$b^2 = [E \{R_x(\tau)\}]^2 \quad (4.23)$$

Nell'espressione precedente è presente un contributo legato alla varianza della stima ($\sigma_{\hat{R}_x(\tau)}^2$) e uno legato alla polarizzazione (b).

4.2.3 Stima della densità spettrale di potenza

Spesso è utile avere informazioni riguardo la densità spettrale, in quanto essa ad esempio permette di stimare l'occupazione di banda del segnale, e quindi ricavare l'intervallo di campionamento opportuno al fine di rispettare il teorema del campionamento. Tra vari metodi per ricavare una stima $\hat{G}_x(\omega)$ di $G_x(\omega)$ considereremo:

1. *stima indiretta*, a partire dalla funzione di autocorrelazione;

2. *stima diretta*, a partire dalle osservazioni del segnale stocastico.

Stima indiretta

La densità spettrale di potenza é, per definizione, la trasformata di Fourier della funzione di autocorrelazione, pertanto la sua stima puó essere pensata come la trasformata di Fourier della stima della funzione di autocorrelazione, ossia

$$\hat{G}_x(\omega) = \mathcal{F} \left\{ \hat{R}_x(\tau) \right\} = \int_{-\infty}^{+\infty} \hat{R}_x(\tau) e^{-j\omega\tau} d\tau \quad (4.24)$$

Poiché però $\hat{R}_x(\tau)$ é definita nell'intervallo $[-T, T]$, é necessario ridurre l'intervallo di integrazione, ponendo

$$\hat{G}_x(\omega) = \int_{-T}^T \hat{R}_x(\tau) e^{-j\omega\tau} d\tau \quad (4.25)$$

Verifichiamo ora la polarizzazione della stima. Consideriamo il valore atteso

$$E \left\{ \hat{G}_x(\omega) \right\} = E \left\{ \int_{-T}^T \hat{R}_x(\tau) e^{-j\omega\tau} d\tau \right\} = \int_{-T}^T E \left\{ \hat{R}_x(\tau) \right\} e^{-j\omega\tau} d\tau \quad (4.26)$$

Se $\hat{R}_x(\tau)$ é una stima non polarizzata di $R_x(\tau)$, allora si ha

$$E \left\{ \hat{R}_x(\tau) \right\} = R_x(\tau) \quad (4.27)$$

Quindi, sostituendo nella relazione precedente, si ottiene

$$E \left\{ \hat{G}_x(\omega) \right\} = \int_{-T}^T R_x(\tau) e^{-j\omega\tau} d\tau \quad (4.28)$$

Si vede come sia presente una sostanziale differenza dalla definizione di densità spettrale di potenza

$$G_x(\omega) = \mathcal{F} \left\{ R_x(\tau) \right\} = \int_{-\infty}^{+\infty} R_x(\tau) e^{-j\omega\tau} d\tau \quad (4.29)$$

Questa differenza si traduce in una polarizzazione, esprimibile attraverso una convoluzione. Vediamo di ricavare questa relazione.

Possiamo riscrivere $E \{ \hat{G}_x(\omega) \}$ come

$$E \{ \hat{G}_x(\omega) \} = \int_{-\infty}^{+\infty} [R_x(\tau) u_{2T}(\tau)] e^{-j\omega\tau} d\tau \quad (4.30)$$

dove

$$u_{2T}(\tau) = \begin{cases} 1 & \text{per } -T \leq t \leq T \\ 0 & \text{altrove} \end{cases} \quad (4.31)$$

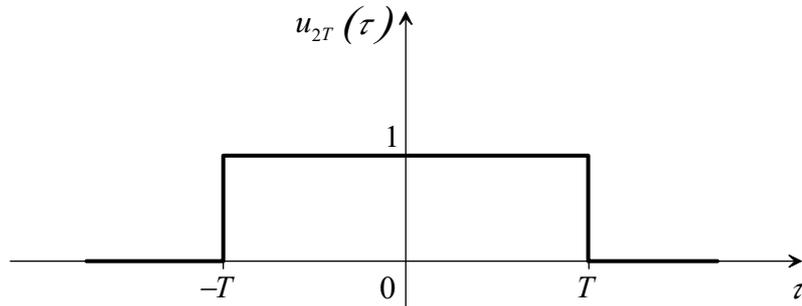


Figura 4.5: Funzione finestra $u_{2T}(\tau)$.

Per le proprietà della trasformata di Fourier, la trasformata del prodotto di due funzioni corrisponde alla convoluzione delle trasformate, pertanto si ottiene

$$E \{ \hat{G}_x(\omega) \} = G_x(\omega) * U_{2T}(\omega) \quad (4.32)$$

dove $U_{2T}(\omega) = \mathcal{F}(u_{2T}(\tau))$.

La trasformata della funzione finestra $u_{2T}(\tau)$ è una *sinc*, con una larghezza del lobo principale pari a $\frac{2\pi}{T}$

Poiché il valor medio della stima è diverso da ciò che vogliamo stimare, abbiamo a che fare con una stima polarizzata. In questo caso la polarizzazione non è dovuta a un fattore puramente proporzionale, ma è espressa attraverso una convoluzione. La polarizzazione si traduce in una *perdita di risoluzione*. Infatti a causa della convoluzione della densità spettrale di potenza con la *sinc*, viene effettuata una sorta di “media mobile pesata” dello spettro. Tale media provoca la perdita dei dettagli più fini dello spettro stesso, perdita dovuta principalmente al lobo principale della *sinc* (i lobi secondari operano un effetto di distorsione).

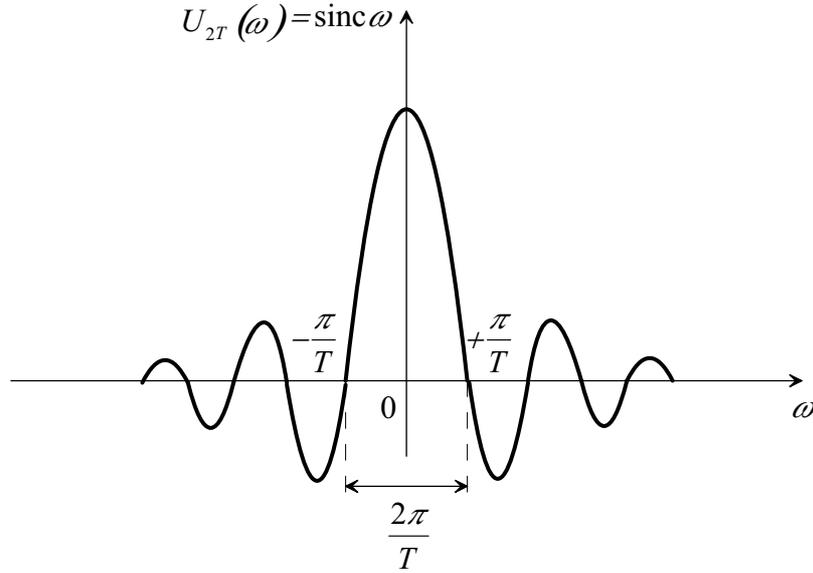


Figura 4.6: Funzione *sinc*, trasformata di Fourier della funzione finestra $u_{2T}(\tau)$.

In particolare si perdono tutti i dettagli inferiori a $\Delta\omega = \frac{2\pi}{T}$. La durata T dell'osservazione influenza quindi anche la risoluzione finale che si otterrà: piú lunga é il tempo di osservazione, maggiore sará la risoluzione ottenibile.

Esempio. Per preservare tutti i dettagli di un segnale al di sopra di 50Hz é necessario avere una durata dell'osservazione superiore a 20ms , infatti

$$2\pi 50 \geq \frac{2\pi}{T} \Rightarrow T \geq \frac{1}{50} = 20\text{ms} \quad (4.33)$$

Operativamente si stabilisce a priori quale é la risoluzione $\Delta\omega$ di cui si ha bisogno, e si determina T in modo tale che

$$T \geq \frac{2\pi}{\Delta\omega} \quad (4.34)$$

Con $T \gg \frac{2\pi}{\Delta\omega}$ si ottiene una stima non polarizzata e si dimostra che questa stima é anche stabile. Se invece si ha $T < \frac{2\pi}{\Delta\omega}$, la stima non viene effettuata, in quanto non ci sono le condizioni per ottenere una stima non polarizzata

Stima diretta

Si tratta di una stima della densità spettrale che non passa per la stima della funzione di autocorrelazione, ma si deduce direttamente dall'osservazione del segnale stocastico. La stima diretta prende il nome di *periodogramma* ed è espressa dalla seguente relazione:

$$\hat{G}_x(\omega) = \frac{1}{T} \left| \int_0^T x(t) e^{-j\omega t} dt \right|^2 \quad (4.35)$$

Il metodo diretto permette di utilizzare direttamente i campioni del segnale $x(t)$ e di utilizzare algoritmi di calcolo di trasformata di Fourier veloce (FFT) per ricavare la stima dello spettro di densità di potenza.

L'espressione (4.35) può essere riscritta integrando in un dominio simmetrico $[-T/2, +T/2]$, ottenendo

$$\hat{G}_x(\omega) = \frac{1}{T} \int_{-\frac{T}{2}}^{+\frac{T}{2}} x(\xi) e^{-j\omega\xi} d\xi \int_{-\frac{T}{2}}^{+\frac{T}{2}} x(\eta) e^{+j\omega\eta} d\eta \quad (4.36)$$

ossia

$$\hat{G}_x(\omega) = \frac{1}{T} \int_{-\frac{T}{2}}^{+\frac{T}{2}} \int_{-\frac{T}{2}}^{+\frac{T}{2}} x(\xi) x(\eta) e^{-j\omega(\xi-\eta)} d\xi d\eta \quad (4.37)$$

Verifichiamo ora la polarizzazione della stima. Consideriamo il valore atteso

$$\begin{aligned} E \left\{ \hat{G}_x(\omega) \right\} &= E \left\{ \frac{1}{T} \int_{-\frac{T}{2}}^{+\frac{T}{2}} \int_{-\frac{T}{2}}^{+\frac{T}{2}} x(\xi) x(\eta) e^{-j\omega(\xi-\eta)} d\xi d\eta \right\} = \\ &= \frac{1}{T} \int_{-\frac{T}{2}}^{+\frac{T}{2}} E \left\{ x(\xi) x(\eta) \right\} e^{-j\omega(\xi-\eta)} d\xi d\eta \end{aligned} \quad (4.38)$$

Sostituendo $R_x(\eta - \xi) = E \{ x(\xi) x(\eta) \}$ si ricava

$$E \left\{ \hat{G}_x(\omega) \right\} = \frac{1}{T} \int_{-\frac{T}{2}}^{+\frac{T}{2}} \int_{-\frac{T}{2}}^{+\frac{T}{2}} R_x(\eta - \xi) e^{-j\omega(\xi-\eta)} d\xi d\eta \quad (4.39)$$

Poniamo $\xi - \eta = \tau$. La figura 4.7 mostra come operare graficamente questo cambio di variabili.

Occorre ricavare l'area del trapezio colorato. Analizzando la figura 4.7 si ricava

$$Area = Area_{A'O'B'} - Area_{AOB} \quad (4.40)$$

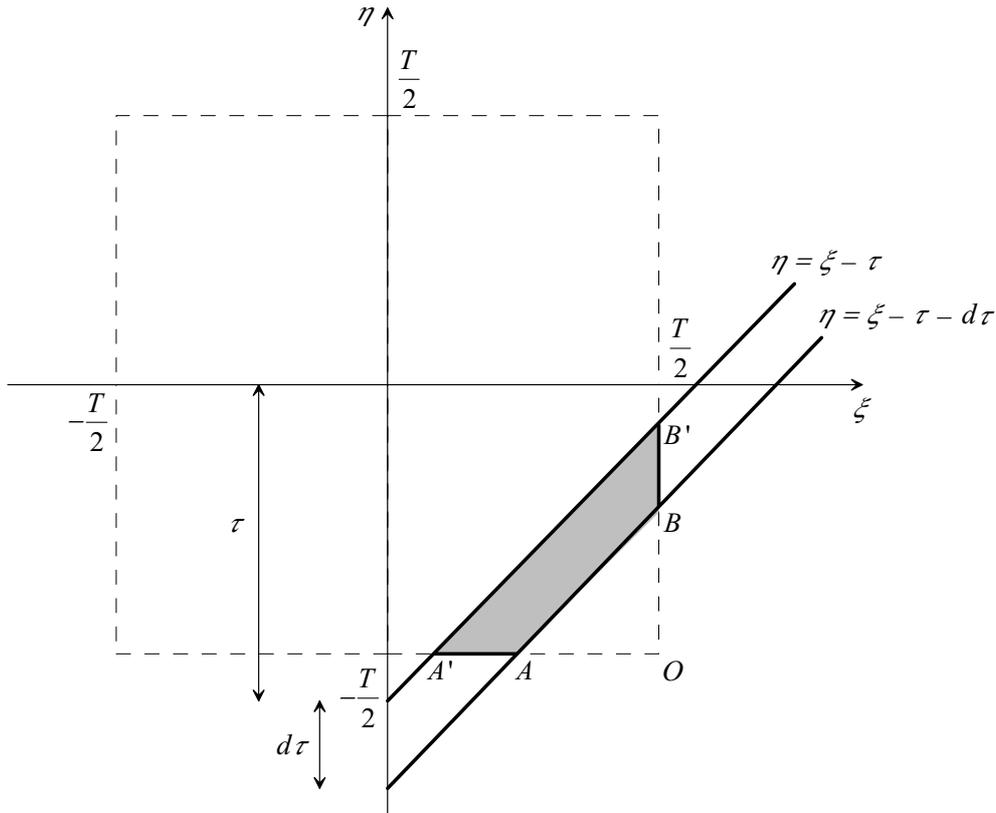


Figura 4.7: Cambio di variabili $\xi - \eta = \tau$ nell'espressione (4.39).

Analizzando la figura 4.39 si ricava

$$A'O = \frac{T}{2} - \left(\tau - \frac{T}{2} \right) = T - \tau \quad (4.41)$$

$$A'A = d\tau$$

Quindi l'area del trapezio risulta essere pari a

$$Area = \frac{(T - |\tau|)^2}{2} - \frac{[(T - |\tau|) - d\tau]^2}{2} = \frac{2(T - |\tau|) d\tau + (d\tau)^2}{2} \simeq \frac{2(T - |\tau|) d\tau}{2} \quad (4.42)$$

dove é stato trascurato il termine $(d\tau)^2$ in quanto infinitesimo di ordine superiore.

La relazione (4.39) diventa pertanto

$$E \left\{ \hat{G}_x(\omega) \right\} = \frac{1}{T} \int_{-T}^T R_x(\tau) (T - |\tau|) e^{-j\omega\tau} d\tau \quad (4.43)$$

ossia

$$E \left\{ \hat{G}_x(\omega) \right\} = \int_{-T}^T R_x(\tau) \frac{T - |\tau|}{T} e^{-j\omega\tau} d\tau = \int_{-\infty}^{+\infty} R_x(\tau) v_{2T}(\tau) e^{-j\omega\tau} d\tau \quad (4.44)$$

dove la funzione $v_{2T}(\tau)$ é la funzione triangolo riportata in figura 4.8.

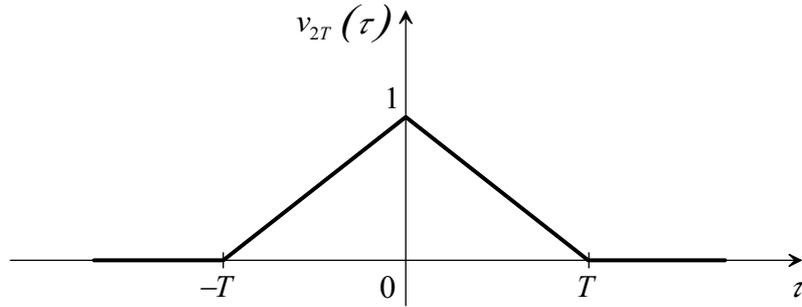


Figura 4.8: Funzione finestra triangolare $v_{2T}(\tau)$.

Il risultato trovato é simile a quello della stima indiretta, in quanto abbiamo a che fare anche in questo caso con una convoluzione

$$E \left\{ \hat{G}_x(\omega) \right\} = G_x(\omega) * V_{2T}(\omega) \quad (4.45)$$

dove $V_{2T}(\omega) = \mathcal{F} \{ v_{2T}(\tau) \}$. La trasformata della funzione finestra triangolare $v_{2T}(\tau)$ é una sinc^2 , con una larghezza del lobo principale pari a $\frac{4\pi}{T}$ (vedi figura 4.9)

Anche in questo caso la polarizzazione della stima si riduce a un problema di risoluzione, da imputare soprattutto al lobo principale. L'ampiezza del lobo principale prende il nome di *finestra di Bartlett*.

Poiché il lobo principale ha ampiezza doppia rispetto a quella che si ha nel metodo indiretto, il metodo diretto ha una risoluzione peggiore rispetto a quello indiretto. Per ottenere la stessa risoluzione del metodo indiretto, é necessario osservare il segnale stocastico per un intervallo almeno doppio rispetto all'intervallo richiesto dal metodo indiretto. La durata dell'intervallo é data dalla relazione

$$T \geq \frac{4\pi}{\Delta\omega} \quad (4.46)$$

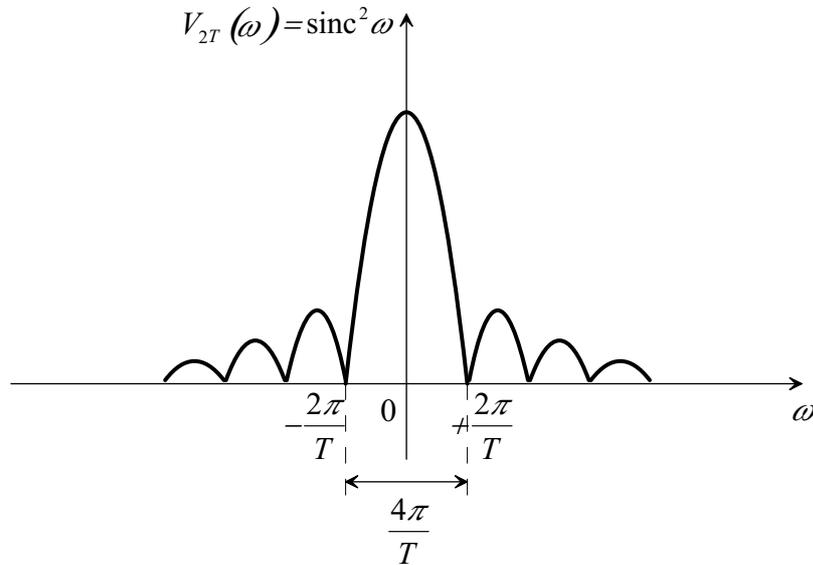


Figura 4.9: Funzione sinc^2 , trasformata di Fourier della funzione finestra triangolare $v_{2T}(\tau)$.

dove $\Delta\omega$ rappresenta la risoluzione desiderata in termini di frequenza.

Aumentando la durata dell'osservazione, la risoluzione migliora, mentre la stabilità rimane invariata, a differenza di quanto visto con il metodo indiretto. Nonostante tutti questi inconvenienti, in pratica si preferisce comunque usare il metodo diretto, in quanto permette di sfruttare la FFT nel calcolo della stima, con un conseguente minore costo computazionale.

Metodo della media periodogrammi

Si tratta di una tecnica applicabile sia al metodo diretto, sia al metodo indiretto, che permette di ridurre la varianza della stima. Supponiamo di aver osservato un segnale per un intervallo di durata T . In base a questa tecnica, non si utilizzano tutte le osservazioni per effettuare un'unica stima, ma si suddivide l'intervallo di osservazione in k sottointervalli di durata T_0 , e si effettua una stima della densità di potenza usando i valori all'interno di ciascun intervallo.

Se le stime sono indipendenti tra loro, la varianza risultante dalla media arit-

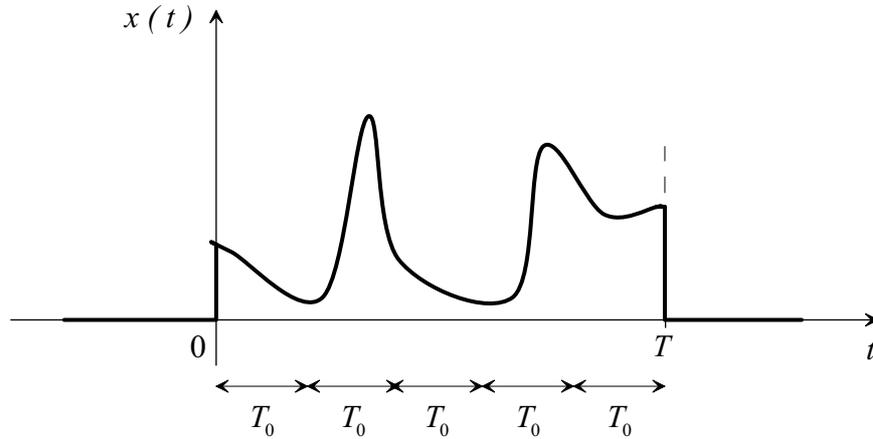


Figura 4.10: Processo aleatorio $x(t)$ osservato in un intervallo T suddiviso in n sottointervalli non sovrapposti di durata T_0 .

metica delle stime sui singoli sottointervalli risulta inferiore alla varianza delle singole stime, ossia

$$\frac{1}{k} \sum_{i=1}^k \sigma_{\hat{G}_x(\omega),i}^2 < \sigma_{\hat{G}_x(\omega),i}^2 \quad \forall i \quad (4.47)$$

dove k rappresenta il numero di sottointervalli

$$k = \left\lfloor \frac{T}{T_0} \right\rfloor \quad (4.48)$$

D'altra parte, considerando intervalli di durata inferiore, possono emergere dei problemi di risoluzione. Nel caso del metodo diretto, ad esempio, abbiamo visto come vengano persi i dettagli dello spettro inferiori a $\frac{4\pi}{T}$. Nel caso in cui $T_0 \ll T$ abbiamo $\frac{4\pi}{T_0} \gg \frac{4\pi}{T}$, ossia la risoluzione peggiora al diminuire di T_0 . E' quindi necessario scegliere la durata dei singoli sottointervalli in modo da garantire la risoluzione desiderata. Chiamando con T' la durata minima di un'osservazione del segnale necessaria per ottenere una risoluzione $\Delta\omega$. Nei paragrafi precedenti abbiamo ricavato

$$\begin{aligned} T' &= \frac{4\pi}{\Delta\omega} \text{ per la stima diretta} \\ T' &= \frac{2\pi}{\Delta\omega} \text{ per la stima indiretta} \end{aligned} \quad (4.49)$$

È possibile effettuare la stima solo nel caso in cui T' sia minore di T . Solitamente si richiede che $T' \ll T$ (ad esempio tramite un fattore dieci), e si sceglie T' come durata di ogni sottointervallo ($T_0 = T'$). Il metodo descritto si basa su una tecnica classica per ridurre la varianza della stima, basata sulla media aritmetica di numerose stime indipendenti. L'applicazione di questo concetto alla stima della densità spettrale di potenza è comunemente attribuita a Bartlett.

La tecnica descritta è principalmente usata per rendere più stabile la stima calcolata con il metodo diretto. L'applicazione al metodo indiretto è più rara, in quanto tale metodo non risente di problemi di stabilità (la stima è asintoticamente stabile), anzi, l'applicazione del metodo della media dei periodogrammi porta a una riduzione della risoluzione a parità di intervallo T .

Fino a questo momento abbiamo considerato stime indipendenti, dovute a sottosequenze adiacenti l'una all'altra (vedi figura 4.10). Consideriamo ora delle sottosequenze parzialmente sovrapposte, come mostrato in figura 4.11.

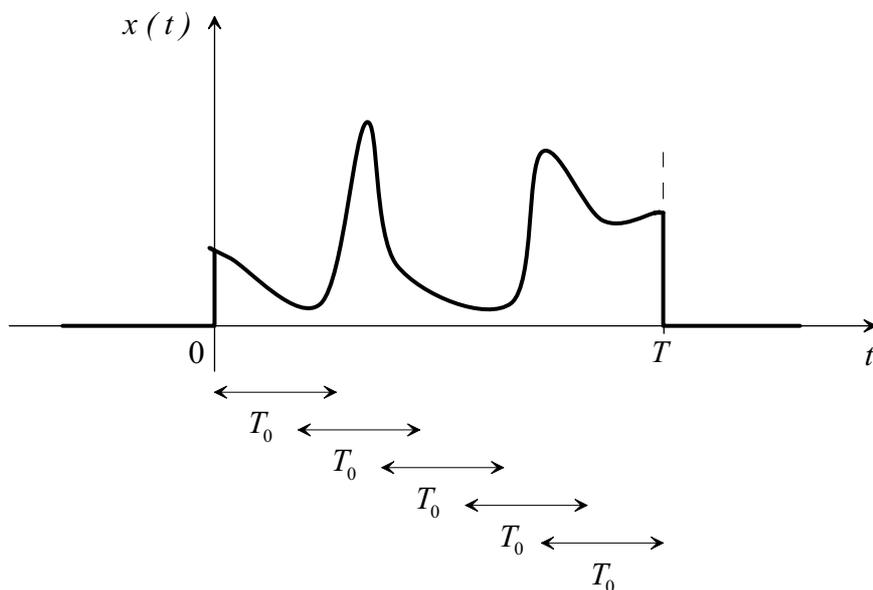


Figura 4.11: Processo aleatorio $x(t)$ osservato in un intervallo T suddiviso in n sottointervalli parzialmente sovrapposti di durata T_0 .

In questo modo è possibile aumentare il numero dei periodogrammi, ossia la

sommatoria considerata nella (4.47) é composta da un numero maggiore di termini. Tuttavia, dato che le varie stime non sono piú indipendenti a causa della parziale sovrapposizione delle sequenze usate per determinarle, non é garantita una effettiva riduzione della varianza. Il metodo della *media dei periodogrammi modificati* di Welch, é un esempio di come sia possibile ottenere la riduzione della varianza del metodo originale di Bartlett pur considerando stime non indipendenti, con l'ulteriore vantaggio di poter controllare la risoluzione della stima agendo sulla durata del sottointervallo T_0 .

4.2.4 Campionamento di segnali aleatori

Per il momento sono stati considerati esclusivamente segnali a tempo continuo. Nella pratica però le stime vengono effettuate con procedimenti numerici. Nasce quindi il problema del campionamento di segnali aleatori.

Consideriamo un segnale analogico *deterministico* $x(t)$ con trasformata di Fourier $X(\omega)$ e banda Ω_x . Estraiamo da $x(t)$ una sequenza $x_n = x(nT_c)$, campionando con intervallo di campionamento costante T_c . Detta ω_c la pulsazione

$$\omega_c = \frac{2\pi}{T_c} \quad (4.50)$$

Se si verifica

$$\omega_c \geq 2\Omega_x \quad (4.51)$$

allora il segnale é completamente specificato dai suoi campioni, e si puó ricostruire a partire da essi (teorema del campionamento).

É possibile definire gli stessi concetti anche nel caso in cui il segnale considerato sia un segnale stocastico. A tal proposito consideriamo la densità spettrale di un segnale $x(t)$.

Possiamo definire come larghezza di banda $\bar{\Omega}_x$, la misura dell'intervallo entro cui é contenuta una percentuale molto grande (ad esempio il 95%) della potenza

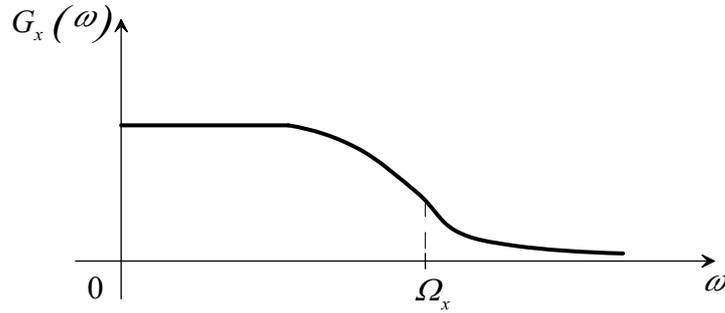


Figura 4.12: Esempio di spettro di densità di potenza di un segnale aleatorio.

totale del segnale, ossia

$$\int_{|\omega_c| > \Omega_x} G_x(\omega) d\omega \leq 5\% \text{ della potenza totale} \quad (4.52)$$

Definiamo l'intervallo di campionamento T_c come

$$T_c = \frac{2\pi}{\omega_c} \quad (4.53)$$

Si dimostra che campionando la funzione di autocorrelazione con un intervallo di campionamento $T_c \leq \pi/\Omega_x$, (*condizione di Nyquist*) siamo “quasi sempre” in grado di ricostruire la funzione di autocorrelazione originaria. In questo caso il teorema del campionamento continua a valere, ma in termini probabilistici, in quanto stiamo considerando quantità aleatorie.

Per poter fissare l'intervallo di campionamento è necessario conoscere la banda Ω_x , il che implica la conoscenza della densità spettrale di potenza. Poiché $G_x(\omega)$ rappresenta la quantità da stimare, non è possibile conoscere a priori l'intervallo T_c che soddisfa la condizione di Nyquist. In prima battuta è possibile effettuare un campionamento con T_c scelto casualmente, e verificare a posteriori se il limite di Nyquist è stato violato o meno. In generale, è buona norma filtrare mediante un filtro passa basso (LPF) il segnale $x(t)$ allo scopo di limitarlo in banda, e successivamente campionarlo con un intervallo di campionamento dipendente dalla frequenza di taglio del filtro (vedi figura 4.13).

In particolare, si possono presentare due casi relativamente al filtraggio di $x(t)$ con un LPF:

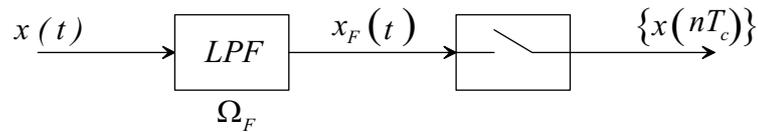


Figura 4.13: Rappresentazione schematica del passaggio da un segnale a tempo continuo al corrispondente a tempo discreto.

1. se $\Omega_F \gg \Omega_x$, il LPF permette di limitare in banda il segnale, senza tuttavia alterarne in modo vistoso le caratteristiche, dato che vengono attenuate solo le componenti armoniche piú elevate (vedi figura 4.14);
2. se $\Omega_F \ll \Omega_x$, il LPF limita in banda il segnale alla pulsazione Ω_x , eliminando parti dello spettro ancora significative. Il risultato é una alterazione sostanziale del segnale $x(t)$, per cui il processo di stima andrà a stimare uno spettro di densità di potenza di un segnale molto diverso da quello originario (vedi figura 4.15).

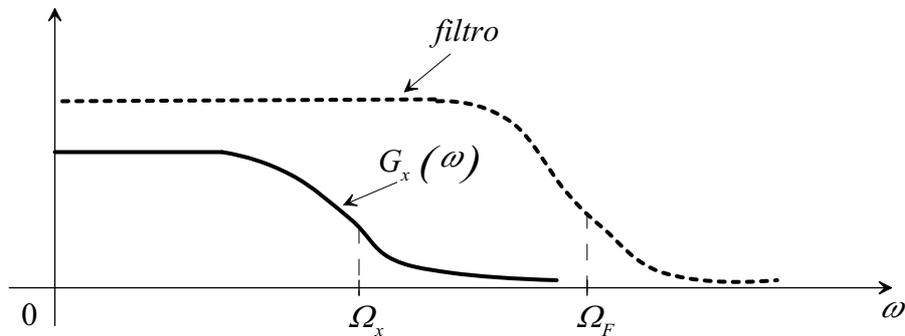


Figura 4.14: Spettro di densità di potenza del segnale e funzione di trasferimento di un filtro passa basso con $\Omega_F \gg \Omega_x$. L'operazione di filtraggio non modifica sostanzialmente le caratteristiche del segnale.

Per evitare di usare filtri con una banda troppo ristretta, si può procedere empiricamente mediante un apparato analogico. In particolare si può utilizzare una batteria di filtri passa basso in parallelo con frequenze di taglio via via crescenti (vedi figura 4.16).

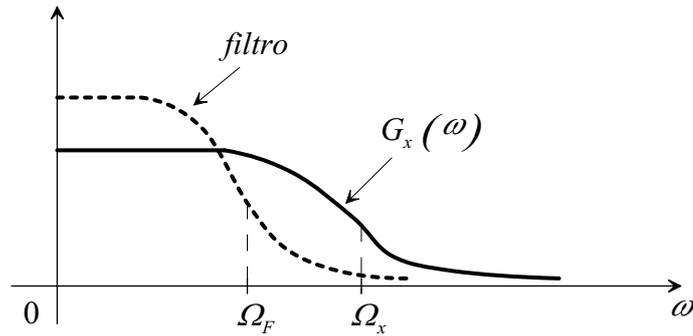


Figura 4.15: Spettro di densità di potenza del segnale e funzione di trasferimento di un filtro passa basso con $\Omega_F \ll \Omega_x$. L'operazione di filtraggio modifica sostanzialmente le caratteristiche del segnale.

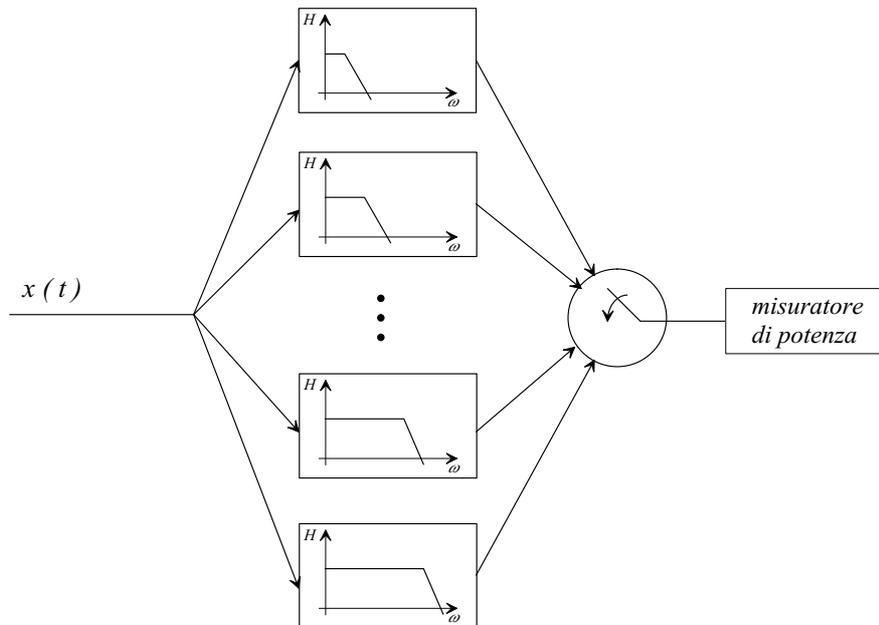


Figura 4.16: Batteria di filtri in parallelo con commutatore a valle che permette di scegliere l'uscita "migliore" sulla base di quanto indicato dal misuratore di potenza a valle.

Il segnale $x(t)$ passa attraverso tutti i filtri, e un selettore a valle collega di volta in volta l'uscita di ciascun filtro con un misuratore di potenza, avente lo scopo di misurare la potenza del segnale. E' ragionevole aspettarsi che i primi

filtri, quelli a banda piú stretta, tagliano componenti importanti dello spettro del segnale, e quindi che le quantità in uscita dai primi filtri abbiano una potenza limitata. Via via che la banda passante dei filtri aumenta, aumenterà anche la potenza del segnale in uscita, sino a un valore limite attorno a cui sembreranno stabilizzarsi tutte le misurazioni. Questo accade quando i filtri hanno raggiunto una larghezza di banda tale da attenuare solo le componenti meno significative dello spettro degli ingressi. Quindi in pratica si sceglie come frequenza di campionamento un valore maggiore del doppio della banda del filtro oltre il quale la potenza misurata non varia piú significativamente.

Una alternativa al procedimento illustrato consiste nel procedere per tentativi. Si campiona il segnale con un intervallo di campionamento via via crescente, e si effettua una stima della densità spettrale di potenza per ogni intervallo di campionamento. Quando la stima non cambia piú in modo apprezzabile, si fissa la frequenza corrispondente come frequenza di campionamento del segnale per una corretta valutazione della stima.

4.2.5 Formulazione numerica delle stime

E' possibile riscrivere le espressioni ricavate per la stima della funzione di autocorrelazione e della densità spettrale in tempo continuo anche in tempo discreto. Abbiamo definito la stima dell'autocorrelazione come

$$\hat{R}_x(\tau) = \frac{1}{T - |\tau|} \int_{|\tau|}^T x(t)x(t + \tau)dt \quad (4.54)$$

Si passa al tempo discreto ponendo

$$\begin{aligned} \tau &= kT_c \\ t &= nT_c \\ T &= NT_c \end{aligned} \quad (4.55)$$

La stima della funzione di autocorrelazione discreta é

$$\hat{R}_x(kT_c) = \frac{1}{T_c(N - k)} \sum_{n=k}^{N-1} x(nT_c)x[(n + k)T_c] T_c \quad \text{con } k = 0, \dots, N - 1 \quad (4.56)$$

dove N rappresenta il numero totale di campioni. La stima della densità spettrale discreta può essere ricavata con il metodo indiretto, tramite la trasformata di Fourier discreta della stima della funzione di autocorrelazione $\hat{R}_x(kT_c)$, ossia

$$\hat{G}_x(h) = \sum_{k=0}^{N-1} \hat{R}_x(k) e^{-jkh \frac{2\pi}{N}} \quad \text{con } h = 0, \dots, N-1 \quad (4.57)$$

Un possibile problema consiste nella scelta corretta del numero di campioni N da usare per garantire la stabilità della stima e una specifica risoluzione. Per quanto riguarda il metodo indiretto, la stabilità è garantita a patto di rispettare il teorema del campionamento. Per il metodo diretto invece, seguendo un procedimento analogo a quanto fatto nel tempo continuo, per ridurre la varianza della stima (e quindi aumentarne la stabilità), è possibile suddividere la sequenza di campioni in una serie di sottosequenze più brevi composte da N_0 campioni (vedi figura 4.17).

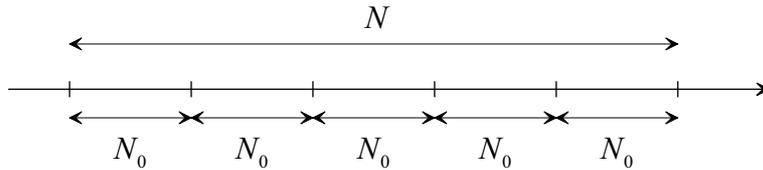


Figura 4.17: Suddivisione in sottosequenze di un'unica sequenza al fine di ridurre la varianza della stima.

Nel caso a tempo continuo abbiamo ricavato

$$\hat{G}_x(\omega) = \frac{1}{T} \left| \int_0^T x(t) e^{-j\omega t} dt \right|^2 \quad (4.58)$$

La corrispondente stima della densità spettrale discreta è

$$\hat{G}_{x,i}(h) = \frac{1}{N_0} \left| \sum_{k=0}^{N-1} x(k) e^{-jkh \frac{2\pi}{N_0}} \right|^2 \quad \text{con } h = 0, \dots, N-1 \quad i = 0, \dots, L \quad (4.59)$$

dove

$$L = \left\lfloor \frac{N}{N_0} \right\rfloor \quad (4.60)$$

rappresenta il numero di sottosequenze considerate.

Per quanto riguarda la risoluzione, si può dimostrare che la risoluzione del metodo diretto risulta essere

$$\frac{1}{N_0} \frac{2\pi}{T_c} \quad (4.61)$$

ossia pari alla differenza tra due valori consecutivi dell'indice h , mentre per il metodo indiretto risulta essere

$$\frac{1}{N} \frac{2\pi}{T_c} \quad (4.62)$$

Dato che $N_0 \ll N$, la risoluzione del metodo diretto risulta inferiore a quella del metodo indiretto. Nonostante questi problemi il metodo diretto è più usato perché ha un costo computazionale più basso.

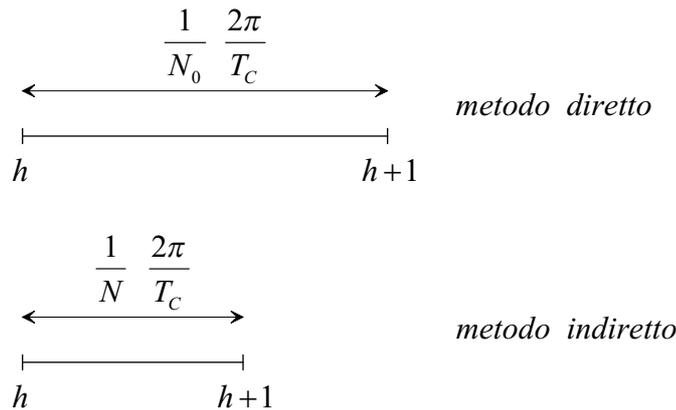


Figura 4.18: Differenze di risoluzione tra il metodo di stima diretto e quello indiretto.

4.2.6 Scelta della finestra temporale

Nelle stime effettuate fino a questo momento, è stata considerata una funzione di autocorrelazione definita in un intervallo $[-T, T]$ in quanto si hanno a disposizione solamente osservazioni del segnale in una finestra temporale limitata. Al di fuori dell'intervallo $[-T, T]$ la funzione di autocorrelazione non è nota.

Essendo sconosciuta, una possibile scelta consiste nel porla uguale a zero, e utilizzare poi questa funzione per stimare la densità spettrale di potenza attraverso la trasformata di Fourier.

In alternativa é possibile cercare un periodogramma “smussato”. Calcoliamo

$$\hat{G}_x(\omega) = \mathcal{F} \left\{ \hat{R}_x(\tau) \cdot w(\tau) \right\} \quad (4.63)$$

dove $w(\tau)$ é una opportuna funzione finestra scelta in modo tale da ridurre l'errore dovuto al troncamento di $R_x(\tau)$.

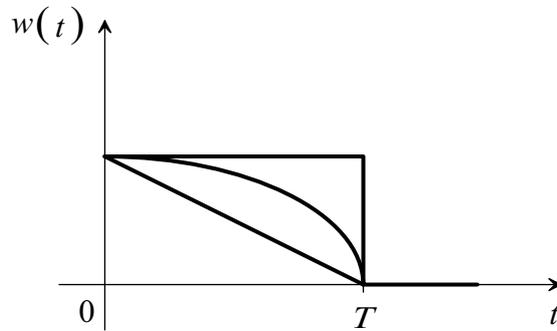


Figura 4.19: Esempi di possibili funzioni finestra $w(t)$.

	$w(\tau)$	$W(\omega)$
Bartlett	$v_{2T}(\tau)$	$\frac{4 \sin^2(\frac{\omega T}{2})}{\omega^2 T}$
Tukey	$\frac{1}{2} \left(1 + \cos\left(\frac{\pi}{T}\tau\right) \right)$	$\frac{\pi^2 \sin(\omega T)}{\omega(\pi^2 - \omega^2 T^2)}$
Parzen	$\frac{3}{2} v_T(\tau) * v_T(\tau)$	$\frac{3T}{4} \left(\frac{\sin(\frac{\omega T}{4})}{\omega \frac{T}{4}} \right)^4$

Tabella 4.1: Esempi di funzioni finestra.

La scelta della funzione finestra ha delle conseguenze sulla stima della densità spettrale, in particolare rispetto a

1. *risoluzione* (legata al lobo principale della $W(\omega)$);

2. *leakage* (fenomeno di interferenza che può provocare la perdita di informazioni relativamente piccole in termini di ampiezza nel dominio delle frequenze, ad una risoluzione minore di quella legata alla dimensione della finestra spettrale).

Come esposto nei paragrafi precedenti, infatti, moltiplicare la autocorrelazione per una funzione finestra equivale a effettuare la convoluzione nel dominio trasformato tra lo spettro di densità di potenza e la trasformata della finestra. Tale convoluzione provoca una polarizzazione della stima e una riduzione della risoluzione. La convoluzione con lo spettro della finestra (sia esso una *sinc*, una *sinc*² o altre forme) realizza una sorta di media mobile pesata della densità spettrale di potenza, che provoca la perdita dei dettagli al di sotto di un certo intervallo di frequenze. Si vuole dunque ricercare una funzione finestra $w(\tau)$ tale da ridurre i problemi di risoluzione e di *leakage* (problemi legati ai lobi laterali della trasformata della finestra temporale $W(\omega)$). Analizziamo ad esempio la situazione riportata in figura 4.20: lo spettro di densità di potenza “vero” è caratterizzato da due picchi di ampiezze molto diverse tra loro, ad una distanza $\omega_2 - \omega_1$ maggiore dell’ampiezza del lobo centrale di $W(\omega)$. A causa della convoluzione con la trasformata della finestra, il secondo picco non compare nella densità spettrale stimata perché troppo vicino al primo. Gli effetti di questa interferenza, legati alla dimensione dei lobi laterali di $W(\omega)$, possono essere ridotti se $W(\omega)$ si attenua rapidamente al crescere di ω . Alcuni esempi sono riportati in figura 4.19 e nella tabella 4.1: tutte sono caratterizzate dall’essere nulle per $|\tau| \geq T$.

4.3 Stima di un processo stocastico

Supponiamo di dover trovare una stima del segnale $s(t)$ a partire dalle osservazioni effettuate su un segnale $x(t)$, diverso da quello che si vuole stimare. Possono presentarsi due categorie di problemi, a seconda della stazionarietà o meno dei

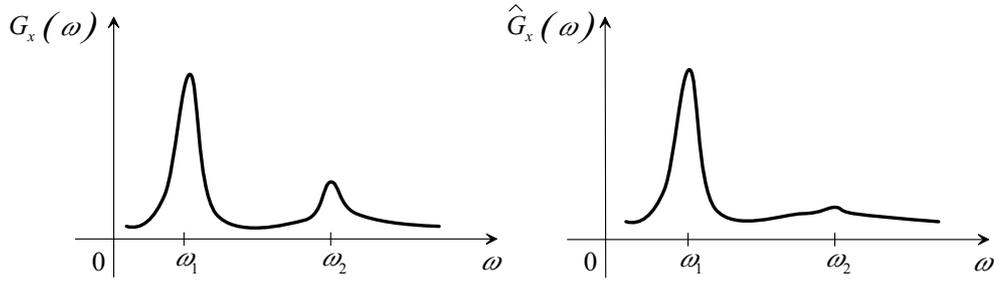


Figura 4.20: Perdita di risoluzione nello spettro di densità di potenza stimato a causa della convoluzione con lo spettro della funzione finestra.

processi in gioco. In particolare per i processi stocastici stazionari utilizzeremo la teoria sviluppata da Wiener, mentre nel caso di assenza di stazionarietà ci rifaremo all'analisi di Kalman.

4.3.1 Filtro di Wiener–Hopf

Come anticipato, la teoria di Wiener riguarda la stima di processi stocastici stazionari. Supponiamo di poter osservare per un certo intervallo di tempo $[a, b]$ un segnale $x(t)$ (*segnale osservabile*) e di voler utilizzare queste osservazioni per stimare un altro segnale $s(t)$. La stima $\hat{s}(t)$ di $s(t)$ dipenderà dalle osservazioni di $x(t)$ mediante un certo operatore, indicato nel seguito con g , ossia

$$\hat{s}(t) = [g \cdot x(\xi), \quad a \leq \xi \leq b] \quad (4.64)$$

Utilizzando la teoria della stima secondo il minimo errore quadratico medio, possiamo definire la stima come

$$\hat{s}(t) = E\{s(t)|x(\xi), \quad a \leq \xi \leq b\} \quad (4.65)$$

Trovare una espressione della stima a partire dalla relazione precedente è tuttavia un problema molto complesso. Per semplicità Wiener cercò soluzioni $\hat{s}(t)$ all'interno della classe di operatori g *lineari*.

E' ragionevole aspettarsi che in questa ipotesi semplificativa valga ancora il principio di ortogonalità enunciato nel paragrafo 3.2.3 per le stime di parametri.

Secondo questo principio l'errore sulla stima (nel caso del criterio MEQM), definito come differenza tra il segnale $s(t)$ e la sua stima $\hat{s}(t)$ risulta essere ortogonale ai dati, ossia in questo caso alle misurazioni $x(\xi)$. Come già affermato nei capitoli precedenti, se siamo nell'ipotesi di segnali gaussiani, allora la stima fornita dal valor medio condizionato (4.65) coincide con la stima lineare $\hat{E}\{\cdot\}$, ossia

$$E\{s(t)|x(\xi)\} = \hat{E}\{s(t)|x(\xi)\} \quad (4.66)$$

Sfruttando il principio di ortogonalità, cerchiamo il filtro ottimo lineare tale che

$$\hat{s}(t) = h(t) \diamond x(t) \quad (4.67)$$

dove $h(t)$ è una quantità caratteristica del filtro e “ \diamond ” indica un generico operatore lineare che agisce sul segnale osservabile $x(t)$. Ad esempio nel caso in cui l'operatore \diamond coincida con l'operatore di convoluzione, $h(t)$ rappresenta la risposta all'impulso del filtro.

Definito l'errore della stima come $\epsilon(t) = s(t) - \hat{s}(t)$, calcoliamone l'errore quadratico medio

$$P = E\{\epsilon^2(t)\} \quad (4.68)$$

Poiché la stima secondo il criterio del minimo errore quadratico medio è una stima non polarizzata, il valor medio dell'errore sarà pari a zero. L'espressione precedente quindi può essere riscritta come

$$P = E\{[\epsilon(t) - \bar{\epsilon}]^2\} \quad (4.69)$$

In questo modo la quantità P coincide con la varianza dell'errore. Cerchiamo una soluzione $h(t)$ tale da rendere minima la varianza dell'errore, ossia

$$h_{OTTIMO}(t) = \arg \min_{h(t)} E\{\epsilon^2(t)\} = \arg \min_{h(t)} E\{[s(t) - h(t) \diamond x(t)]^2\} \quad (4.70)$$

La quantità $h_{OTTIMO}(t)$ prende il nome di *filtro di Wiener*, ed è un filtro *adattato*.

Cerchiamo ora la stima $\hat{s}(t)$ utilizzando come operatore “ \diamond ” un integrale esteso all’intervallo di tempo in cui si effettuano le misurazioni. Si noti come un operatore di questo tipo sia effettivamente un operatore lineare, come ipotizzato in precedenza. Poniamo dunque

$$\hat{s}(t) = E\{s(t)|x(\xi), a \leq \xi \leq b\} = \int_a^b h(\alpha)x(\alpha)d\alpha \quad (4.71)$$

Passando al dominio discreto, supponendo di avere a che fare con n campioni di misura, l’espressione precedente diventa

$$\hat{s}(t) = \sum_{k=1}^n h(\alpha_k)x(\alpha_k)\Delta\alpha = \sum_{k=1}^n a_k x(\alpha_k) \quad (4.72)$$

dove $a_k = \Delta\alpha h(\alpha_k)$. La stima $\hat{s}(t)$ risulta essere una somma pesata delle osservazioni del segnale $x(t)$. Sfruttando il principio di ortogonalit , la stima ottima sar  quella in cui l’errore risulta ortogonale ai dati, ossia

$$E\left\{\left[s(t) - \sum_{k=1}^n h(\alpha_k)x(\alpha_k)\Delta\alpha\right]x(\xi_j)\right\} = 0 \quad \forall j = 1, \dots, n \quad (4.73)$$

Riarrangiando i termini dell’espressione precedente e sfruttando la linearit  dell’operatore *expectation* si ottiene

$$E\{s(t)x(\xi)\} - \sum_{k=1}^n h(\alpha_k)E\{x(\alpha_k)x(\xi_j)\} = 0 \quad j = 1, \dots, n \quad (4.74)$$

Introducendo la correlazione incrociata R_{sx} tra $s(t)$ e $x(t)$ e la funzione di autocorrelazione R_x del segnale osservabile, si ricava

$$R_{sx}(t, \xi_j) = \sum_{k=1}^n h(\alpha_k)R_x(\alpha_k, \xi_j)\Delta\alpha \quad j = 1, \dots, n \quad (4.75)$$

Per ritornare nel continuo, facciamo tendere $\Delta\alpha$ a zero, in modo che la sommatoria possa essere trasformata in un integrale. L’equazione precedente quindi si trasforma in

$$R_{sx}(t, \xi) = \int_a^b h(\alpha)R_x(\alpha, \xi)d\alpha \quad \text{con} \quad a \leq \xi \leq b \quad (4.76)$$

Si noti che era comunque possibile ricavare l'equazione (4.76) senza passare attraverso la discretizzazione, ma partendo direttamente dall'equazione (4.71).

La relazione (4.76) costituisce una equazione integrale di Fredholm di prima specie con incognita $h(\alpha)$. La soluzione generale di questa equazione é molto difficile da ricavare, pertanto si adottano spesso tecniche di risoluzione di tipo numerico. Soluzioni in forma chiusa possono essere ottenute facilmente solo in alcuni casi particolari.

Supponiamo ad esempio di conoscere il segnale $x(t)$ su tutto l'asse dei tempi. La stima $\hat{s}(t)$ può essere espressa come

$$\hat{s}(t) = E\{s(t)|x(\xi), -\infty < \xi < +\infty\} = \int_{-\infty}^{+\infty} h(\alpha)x(t-\alpha)d\alpha \quad (4.77)$$

Essendo $x(t)$ nota per ogni istante di tempo, l'operatore “ \diamond ” può coincidere con l'operatore di convoluzione “ $*$ ” in quanto $x(t)$ può essere valutato per qualunque valore di t . Si sottolinea come il filtro trovato potrebbe essere *non causale*, in quanto non si impone alcun vincolo sulla risposta all'impulso. Consideriamo ancora la stima MEQM e il principio di ortogonalit . In questo caso avremo

$$E \left\{ \left[s(t) - \int_{-\infty}^{+\infty} h(\alpha)x(t-\alpha)d\alpha \right] x(\xi) \right\} = 0 \quad (4.78)$$

con ξ indicante un generico istante di tempo. In particolare ponendo $\xi = t - \tau$ é possibile legare ξ all'istante corrente t , con τ che svolge il ruolo di un parametro. Osserviamo che imponendo $\tau > 0$ si impone la causalit  del filtro, in quanto questo equivale a considerare come dati per stimare $s(t)$ solo le osservazioni *precedenti* l'istante corrente. Se invece non si impone alcun vincolo sul valore di τ , é possibile ottenere soluzioni corrispondenti a filtri non causali (anticipativi).

Sostituendo $\xi = t - \tau$ (con τ qualunque) nella (4.78) si ha

$$E \left\{ \left[s(t) - \int_{-\infty}^{+\infty} h(\alpha)x(t-\alpha)d\alpha \right] x(t-\tau) \right\} = 0 \quad (4.79)$$

Ripetendo gli stessi ragionamenti affrontati in precedenza, riarrangiando i termini dell'equazione precedente si ottiene facilmente

$$R_{sx}(\tau) = \int_{-\infty}^{+\infty} h(\alpha)R_x(\tau-\alpha)d\alpha \quad (4.80)$$

Non avendo fissato alcun vincolo di τ , l'integrale precedente equivale a una convoluzione tra la risposta all'impulso del filtro e la funzione di autocorrelazione del segnale osservabile. Passando al dominio delle trasformate di Fourier, la convoluzione diventa un semplice prodotto, ossia l'equazione precedente diventa

$$\mathcal{F}\{R_{sx}(\tau)\} = \mathcal{F}\{h(\tau) * R_x(\tau)\} \quad \Rightarrow \quad G_{sx}(\omega) = H(\omega)G_x(\omega) \quad (4.81)$$

Pertanto in questo caso la funzione di trasferimento del filtro di Wiener può essere scritta in forma chiusa come

$$H(\omega) = \frac{G_{sx}(\omega)}{G_x(\omega)} \quad (4.82)$$

L'espressione precedente prende il nome di *filtro di Wiener non causale*. Nella pratica è impossibile realizzare un filtro anticipativo, quindi si ricorre spesso all'utilizzo di elementi di ritardo. Tali ritardi permettono di considerare nel processo di stima anche valori di $x(t)$ posteriori all'istante t in cui si vuole fare la stima. Il processo di stima risultante però non potrà più essere in tempo reale, ossia la stima di s all'istante t non sarà nota istantaneamente, ma dopo un certo periodo, legato al numero di informazioni successive a t che si vogliono usare nel processo di stima. Si può osservare come l'adozione di elementi ritardatori sia una prassi abbastanza comune nell'ambito delle telecomunicazioni, mentre in applicazioni di controllo di sistemi in tempo reale questa non è assolutamente utilizzabile.

Valutiamo ora la bontà della stima calcolando l'errore quadratico medio

$$P = E \left\{ \left[s(t) - \int_{-\infty}^{+\infty} h(\alpha)x(t - \alpha)d\alpha \right]^2 \right\} \quad (4.83)$$

Ricordando che la potenza del segnale può essere espressa dal valore dell'autocorrelazione nell'origine, indicato con $R_s(0)$, si ricava

$$P = R_s(0) - \int_{-\infty}^{+\infty} h(\alpha)R_{sx}(\alpha)d\alpha \quad (4.84)$$

Osserviamo che il secondo termine dell'equazione precedente corrisponde al prodotto scalare tra la funzione $h(t)$ e e la crosscorrelazione R_{sx} . Sfruttando la

definizione di prodotto scalare nel dominio delle frequenze, e interpretando la potenza del segnale come integrale dello spettro di densità di potenza, si ottiene

$$\begin{aligned} P &= \frac{1}{2\pi} \int_{-\infty}^{+\infty} G_s(\omega) d\omega - \frac{1}{2\pi} \int_{-\infty}^{+\infty} H^*(\omega) G_{sx}(\omega) d\omega = \\ &= \frac{1}{2\pi} \int_{-\infty}^{+\infty} [G_s(\omega) - H^*(\omega) G_{sx}(\omega)] d\omega \end{aligned} \quad (4.85)$$

4.3.2 Tipologie del processo di stima

Abbiamo visto come la stima $\hat{s}(t)$ si possa esprimere come

$$\hat{s}(t) = E\{s(t)|x(t)\}, \quad a \leq t \leq b \quad (4.86)$$

Si possono individuare tre diverse tipologie del processo di stima, che prendono il nome di “smoothing”, “predizione” e “filtraggio e predizione”. Vediamo di analizzarne le caratteristiche principali.

Processo di smoothing

In questo caso l'istante t in cui si vuole calcolare la stima è interno all'intervallo $[a, b]$, ossia è un istante di tempo compreso nella finestra di osservazione di $x(t)$. Un'applicazione pratica del processo di smoothing si ha nel caso in cui si osservi il segnale $s(t)$ corrotto dal rumore $n(t)$ (supposto additivo), ossia nel caso in cui

$$x(t) = s(t) + n(t) \quad (4.87)$$

L'operazione di stima è solo una operazione di ripulitura di $x(t)$ dal rumore.

Processo di predizione

In questo caso l'istante t in cui si vuole calcolare la stima è esterno all'intervallo $[a, b]$. Il caso maggiormente interessante è quello in cui $t > b$, ossia il caso in cui si utilizzano le osservazioni passate di $x(t)$ per effettuare una previsione del valore di s in un istante futuro. Un problema tipico di questo tipo è quello in cui

$$x(t) = s(t) \quad (4.88)$$

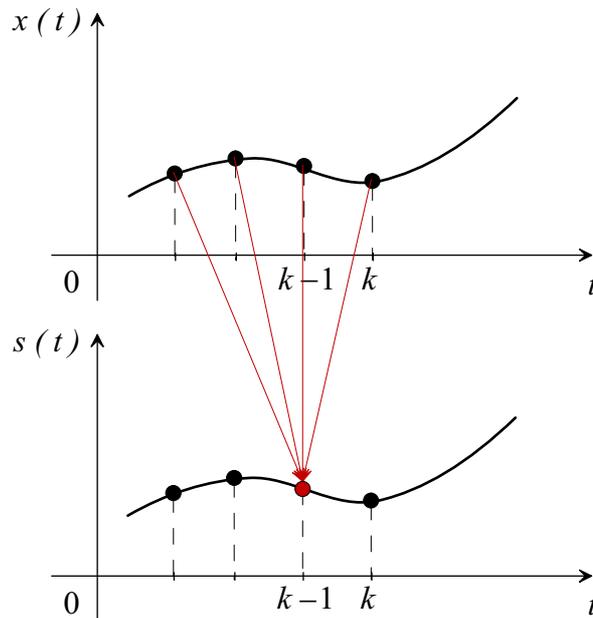


Figura 4.21: Rappresentazione grafica del problema di smoothing. Supponendo di essere all'istante k , le osservazioni sul segnale $x(t)$ vengono usate per stimare il segnale $s(t)$ in un istante *precedente* l'ultima misura effettuata.

ossia un problema in cui non vi sia rumore (o comunque, se presente, sia trascurabile) e si vogliono usare alcune osservazioni del segnale per prevedere i valori che esso assumerà in futuro.

Processo di filtraggio e predizione

In questo caso l'istante t in cui si vuole calcolare la stima è ancora esterno all'intervallo $[a, b]$, ossia si utilizzano le osservazioni di x a disposizione per stimare il valore del segnale s in un istante futuro. In questo caso solitamente

$$x(t) \neq s(t) \quad (4.89)$$

pertanto l'operazione di stima non è una semplice previsione, ma richiede anche una operazione di filtraggio per eliminare eventuali componenti di rumore.

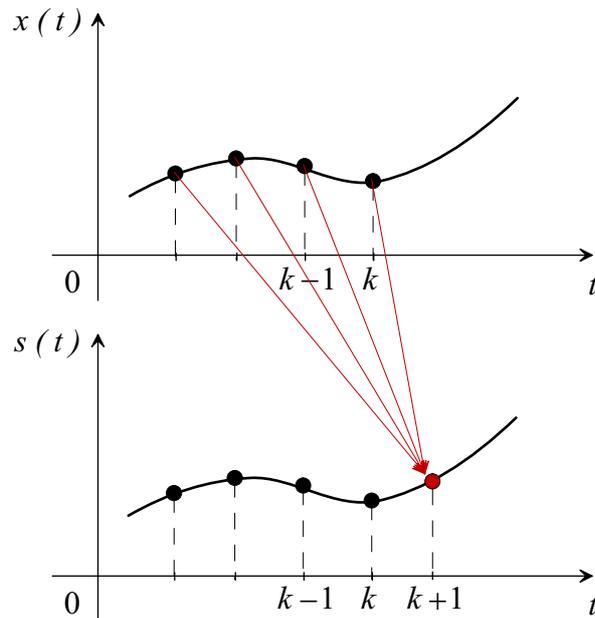


Figura 4.22: Rappresentazione grafica del problema di smoothing. Supponendo di essere all'istante k , le osservazioni sul segnale $x(t)$ vengono usate per stimare il segnale $s(t)$ in un istante *successivo* all'ultima misura effettuata.

4.3.3 Il problema di smoothing in dettaglio

Consideriamo il caso in cui osserviamo un segnale $x(t)$ pari a

$$x(t) = s(t) + n(t) \quad (4.90)$$

dove $n(t)$ é un rumore additivo casuale. Per ipotesi assumiamo che s e n siano tra loro indipendenti e a valor medio nullo. Nei paragrafi precedenti avevamo trovato un'espressione per la funzione di trasferimento del filtro ottimo di Wiener del tipo

$$H(\omega) = \frac{G_{sx}(\omega)}{G_x(\omega)} \quad (4.91)$$

a cui corrisponde una equazione integrale del tipo

$$R_{sx}(\tau) = \int_{-\infty}^{+\infty} h(\alpha) R_x(t - \alpha) d\alpha \quad (4.92)$$

Ricaviamo ora in modo esplicito l'espressione della crosscorrelazione R_{sx} . Si

ha

$$R_{sx}(\tau) = E\{x(t)s(t+\tau)\} = E\{[s(t) + n(t)]s(t+\tau)\} = R_s(\tau) + E\{s(t+\tau)n(t)\} \quad (4.93)$$

Il termine $E\{s(t+\tau)n(t)\}$ nell'equazione precedente é uguale a zero a causa dell'ipotesi di indipendenza dei segnali $n(t)$ e $s(t)$. Sfruttando i risultati precedenti e la linearit , l'equazione (4.91) diventa

$$H(\omega) = \frac{G_s(\omega)}{G_s(\omega) + G_n(\omega)} \quad (4.94)$$

L'interpretazione "operativa" della (4.94) é riportata nella figura 4.23.

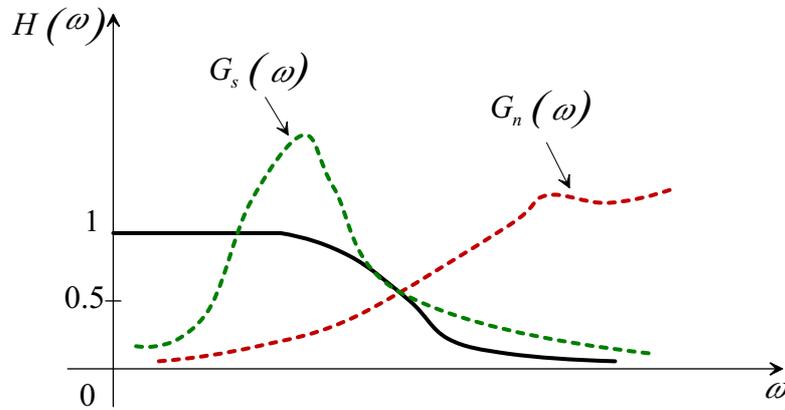


Figura 4.23: Rappresentazione grafica della funzione di trasferimento del filtro ottimo di Wiener rispetto alle densit  spettrali di potenza di $s(t)$ e del rumore.

Analizzando la figura si vede come per valori di ω in cui lo spettro di densit  di potenza del rumore é pressoch  nullo, la funzione di trasferimento assuma valore unitario in quanto non é necessaria alcuna operazione di "ripulitura". Al contrario in zone in cui il rumore ha uno spettro di potenza molto elevato il filtraggio deve essere notevole, e quindi la funzione di trasferimento tende a zero. Quando le densit  spettrali del rumore e del segnale $s(t)$ sono confrontabili la funzione di trasferimento tende ad assumere valori prossimi a $1/2$.

Si pu  dimostrare facilmente, calcolando l'errore quadratico medio P , che nel caso in cui $G_n(\omega)$ e $G_s(\omega)$ non siano sovrapposti la stima ha errore nullo.

4.3.4 Il problema di predizione in dettaglio

Consideriamo il caso dell'osservazione di un segnale $x(t)$ in assenza di rumore, ossia

$$x(t) = s(t) \quad (4.95)$$

Vogliamo ricavare una stima di s a un istante successivo a quello corrente t , ossia vogliamo ricavare

$$\hat{s}(t + \lambda) = E\{s(t + \lambda) | s(t - \tau), \tau \geq 0\} \quad (4.96)$$

A causa dell'ipotesi $\tau \geq 0$, la stima viene espressa come un integrale esteso solo al dominio $[0, +\infty]$, ossia

$$\hat{s}(t + \lambda) = \int_0^{+\infty} h(\alpha) s(t - \alpha) d\alpha \quad (4.97)$$

Questo equivale a cercare la soluzione $h(t)$ nell'ambito dei filtri lineari tempo invariati causali. Applichiamo ancora il principio di ortogonalit , ottenendo

$$E \left\{ \left[s(t + \lambda) - \int_0^{+\infty} h(\alpha) s(t - \alpha) d\alpha \right] s(t - \tau) \right\} = 0 \quad (4.98)$$

Seguendo ragionamenti simili ai precedenti si ricava

$$R_s(\lambda + \tau) = \int_0^{+\infty} h(\alpha) R_s(\tau - \alpha) d\alpha \quad (4.99)$$

valevole per $\tau \geq 0$. L'equazione precedente prende il nome di *equazione di Wiener-Hopf*. La soluzione di questa equazione non si pu  ricavare in forma chiusa per via del vincolo $\tau \geq 0$.

4.3.5 Inconvenienti del filtro di Wiener

Il limite maggiore della trattazione di Wiener risiede nel fatto che essa non   valida nel caso in cui i processi stocastici in gioco non siano stazionari. Inoltre tutte le equazioni ricavate nei paragrafi precedenti considerano tutte le misurazioni effettuate fino a un dato istante: nel caso in cui si effettuasse una nuova

misura e si volesse utilizzare anche quest'ultima per migliorare il processo di stima, bisognerebbe ripetere dall'inizio tutto il procedimento considerando anche la nuova misura. In altri termini non é possibile sfruttare i risultati ottenuti per migliorare la stima con nuove misurazioni, l'unica soluzione é svolgere nuovamente tutti i calcoli. Per ovviare a questo problema si puó cercare di fornire una relazione *ricorsiva*, che permetta ad ogni passo di raffinare (in seguito a un nuovo campione) la stima raggiunta al passo precedente senza dover ripetere i calcoli. É pertanto desiderabile sviluppare una procedura in grado di aggiornare la stima ricorsivamente

$$\hat{s}(k) = a_k \hat{s}(k-1) + b_k x(k) \quad (4.100)$$

dove a_k e b_k sono coefficienti indipendenti dai dati. La stima all'istante k é una somma pesata della stima al passo precedente e della nuova misurazione $x(k)$. La soluzione ricorsiva del processo di stima é stata formulata da Kalman nel 1960, che propose un filtro, valido anche per la stima di processi stocastici non stazionari, che puó essere considerato come generalizzazione di quello di Wiener. In queste prospettive il filtro di Wiener appare come un caso particolare del filtro di Kalman ottenuto quando il filtro di Kalman raggiunge lo stato stazionario.

4.4 Filtro di Kalman (cenni)

Il filtro venne sviluppato alla fine degli anni cinquanta dall'ingegnere di origine ungherese Rudolf E. Kalman mentre lavorava presso la NASA allo sviluppo del programma spaziale americano. La motivazione applicativa fu lo sviluppo di algoritmi che permettessero di determinare, con la massima precisione possibile, posizione e velocità dei satelliti lanciati nell'ambito del programma spaziale. Nello sviluppo del filtro, grande attenzione venne dedicata agli aspetti implementativi cercando di ottenere una soluzione che fosse computazionalmente economica e numericamente robusta. L'interesse del filtro di Kalman (e la fama del

suo inventore) non risiede tuttavia tanto nell'aver risolto un importante problema applicativo, quanto nell'aver inquadrato brillantemente un problema di stima all'interno della Teoria dei Sistemi che, in quegli anni, per merito dello stesso Kalman, compiva passi decisivi, scoprendo nozioni quali controllabilità, osservabilità, controllo ottimo, etc.

Il punto di partenza del ragionamento di Kalman é la stima secondo il criterio del minimo errore quadratico medio di processi stocastici *non stazionari*.

Poiché il sistema é tempo variante, occorre dare una espressione ricorsiva del segnale, ossia

$$s_k = \Phi_k s_{k-1} + \beta_{k-1} \quad (4.101)$$

Il segnale al passo k é legato a quello al passo precedente attraverso un operatore Φ_k (detto operatore di transizione di stato), che nella nostra trattazione é un semplice coefficiente. Il termine β_{k-1} é un rumore gaussiano a media nulla $N(0, \sigma_\beta^2)$ che rappresenta l'incertezza associata al modello del segnale. La relazione (4.101), infatti, rappresenta un modello del segnale, detto *autoregressivo*.

Analogamente, abbiamo bisogno di un modello delle osservazioni, esprimibile come

$$x_k = \rho_k s_k + n_k \quad (4.102)$$

dove n_k rappresenta il rumore della misurazione, e x_k rappresenta il segnale osservato e ρ_k l'operatore di misura.

Confrontiamo il modello trattato con la rappresentazione nello spazio degli stati di un sistema dinamico lineare. La tipica rappresentazione di stato di un sistema lineare é

$$\begin{cases} x_k = Ax_{k-1} + Bu_{k-1} + w_{k-1} \\ y_k = Cx_k + Bv_k \end{cases} \quad (4.103)$$

dove x rappresenta lo stato, y l'uscita, u gli ingressi, w il disturbo di processo e v il disturbo di misura.

Confrontando le equazioni (4.101) e (4.102) con la (4.103), si può notare una stretta somiglianza. In particolare, l'equazione di stato

$$x_k = Ax_{k-1} + Bu_{k-1} + w_{k-1} \quad (4.104)$$

é l'equivalente dell'equazione di modello del segnale, mentre l'equazione di uscita

$$y_k = Cx_k + Bv_k \quad (4.105)$$

é l'equivalente dell'equazione di modello di misura. La similitudine appare ancor piú evidente se i modelli di Kalman sono considerati come sistemi unicamente con ingressi di rumore (ossia con $u \equiv 0$).

4.4.1 Soluzione del problema di Kalman (cenni)

Supponiamo di avere k osservazioni x_1, \dots, x_k del segnale $x(t)$. A partire da queste osservazioni si vuole trovare una stima \hat{s} all'istante j dato k secondo il criterio del minimo errore quadratico medio. In altri termini indichiamo con $\hat{s}(j|k)$ la stima MEQM di s_j basata sulle osservazioni x_1, \dots, x_k .

Vediamo di definire anche in questo caso una quantificazione dell'errore. Poiché stiamo considerando un problema in piú dimensioni, avremo a che fare con una matrice di covarianza. La matrice di covarianza dell'errore di stima risulta essere

$$P(j|k) = E \{ \epsilon(j|k) \epsilon^T(j|k) \} \quad (4.106)$$

dove

$$\epsilon(j|k) = s(j) - \hat{s}(j|k) \quad (4.107)$$

Le osservazioni fino a k dato $k - 1$ possono essere ottenute tramite le osservazioni precedenti, piú altri termini.

Si può ricavare una soluzione in forma ricorsiva, che nel caso scalare assume la forma

$$\hat{s}(k|k-1) = \Phi_k \hat{s}(k-1|k-2) + P_k (x(k) - \Phi_k \hat{s}(k-1|k-2)) \quad (4.108)$$

dove

$$P_k = P(k|k-1) = \frac{\Gamma + \Phi_k^2 P(k-1|k-2)}{1 + \Gamma + \Phi_k^2 P(k-1|k-2)} \quad (4.109)$$

con

$$\Gamma = \sigma_\beta^2 / \sigma_n^2 \quad (4.110)$$

Anche l'espressione di P_k é espressa in termini ricorsivi.

Inoltre, si puó osservare che é possibile ricavare P_k con un procedimento ricorsivo *indipendente dalle misure*.

4.4.2 Struttura ricorsiva del filtro

Per determinare l'errore all'istante k si utilizza il valore dell'errore \hat{P} calcolato al passo precedente. Noto l'errore all'istante corrente é possibile stimare s all'istante corrente utilizzando anche la stima di s al passo precedente e la nuova misurazione. Questi dati permettono anche di fare previsioni sul futuro.

La struttura ricorsiva mostrata in figura 4.24 rappresenta il nucleo del filtro di Kalman. Si articola in due aggiornamenti:

1. *aggiornamento "alla misura"*: la stima predittiva (e la rispettiva matrice di covarianza dell'errore) viene aggiornata per l'effetto di una nuova misura, diventando una stima filtrata (a posteriori);
2. *aggiornamento temporale*: la stima filtrata viene aggiornata per inseguire l'evoluzione dello stato diventando nuovamente stima predittiva, ma all'istante successivo.

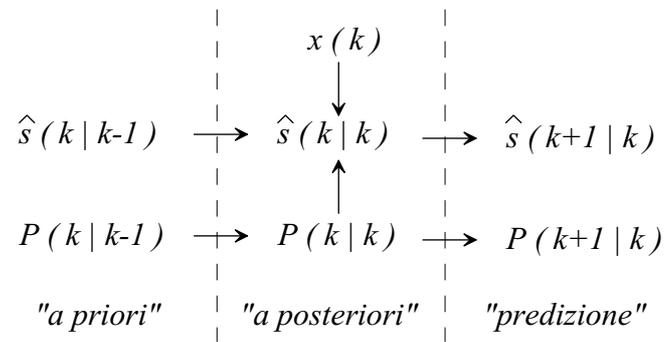


Figura 4.24: Struttura ricorsiva del filtro di Kalman.

Capitolo 5

Microsistemi elettromeccanici

5.1 Definizione di microsistema

Un microsistema é un sistema in cui sono presenti sensori, elaboratori e attuatori con lo scopo di fornire una efficiente interazione tra il mondo esterno (origine delle informazioni da elaborare) e il sistema stesso. Rispetto a un semplice sistema di elaborazione, in cui la parte di interfaccia é spesso secondaria rispetto all'elaborazione vera e propria, i microsistemi sono caratterizzati proprio dalla grande importanza data all'interfacciamento con l'esterno. Lo scopo principale di un microsistema é acquisire informazione dal mondo esterno, elaborarla secondo opportuni algoritmi e produrre come risultato una modifica del mondo che lo circonda.

Un microsistema é composto generalmente da cinque macro-blocchi:

1. acquisizione e conversione del segnale;
2. gestione della potenza;
3. elaborazione dei dati acquisiti;
4. immagazzinamento dati;

5. attuazione dei risultati dell'elaborazione.

La fase di acquisizione e conversione del segnale é costituita prevalentemente da sensori che prelevano l'informazione desiderata dal mondo esterno. Questi sensori sono tipicamente di tipo analogico, in quanto le grandezze del mondo che ci circonda sono quantità prettamente analogiche. Al fine di ottenere una elaborazione piú efficiente e maggiore robustezza al rumore molto spesso conviene convertire tali grandezze in corrispondenti grandezze digitali mediante opportuni convertitori. Dette grandezze vengono poi “inviare” al blocco di elaborazione (sia esso costituito da dispositivi programmabili o hardware dedicato) che ha lo scopo di effettuare su di esse i calcoli necessari al raggiungimento dei risultati desiderati. Strettamente connesso al blocco di elaborazione dati é ovviamente il blocco di immagazzinamento degli stessi, nelle forme piú diverse a seconda delle necessità. I risultati prodotti dalla fase di elaborazione non sono fruibili direttamente dal mondo esterno, ma richiedono l'intervento del blocco di attuazione. Questo blocco é composto, come del resto dice il nome stesso, da attuatori che hanno lo scopo di trasformare i risultati dell'elaborazione in grandezze o azioni in grado di agire sul mondo esterno e di modificarlo. Il blocco di gestione della potenza ha lo scopo di alimentare l'intero sistema, riducendo possibilmente i consumi, ma anche quello di proteggere il sistema da eventuali interferenze provenienti dall'esterno o cattive connessioni che potrebbero danneggiarlo. La figura 5.1 mostra schematicamente l'interazione tra queste cinque parti e la tecnologia che solitamente viene impiegata per realizzarle.

5.2 Tecnologie alla base dei microsistemi

I microsistemi elettromeccanici (*MEMS, MicroElectroMechanical Systems*) sono realizzati con tecnologie che si possono classificare in due categorie principa-

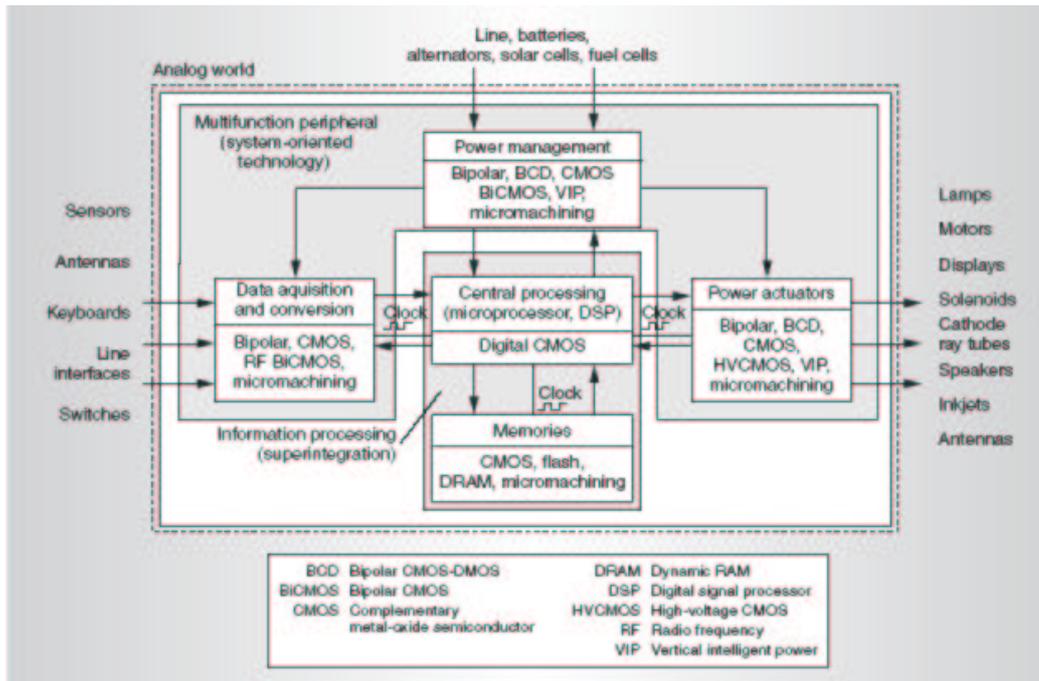


Figura 5.1: Suddivisione schematica in cinque blocchi delle funzionalità tipiche di un microsistema e tecnologie di realizzazione più comuni.

li: *tecnologie singole* (per realizzare funzioni specifiche) e *tecnologie miste* (per consentire l'integrazione di diverse funzionalità).

La figura 5.2 mette in evidenza il ruolo delle tecnologie di base (CMOS, DRAM, Flash etc.) nella realizzazione di microsistemi: per poter realizzare le diverse funzionalità richieste a un *MEMS* occorre utilizzare diverse tecnologie.

5.2.1 Microsistemi in silicio

Tra le diverse possibilità offerte dalla moderna tecnologia, ci soffermeremo sui *MEMS* basati sul silicio. La scelta del silicio si spiega facilmente in quanto questo materiale combina importanti proprietà termo-meccaniche (illustrate in figura 5.3) con processi di lavorazione ormai consolidati grazie ad anni di esperienza accumulati nella fabbricazione di circuiti integrati.

Analizzando i dati riportati in figura 5.3 si nota come il silicio abbia caratte-

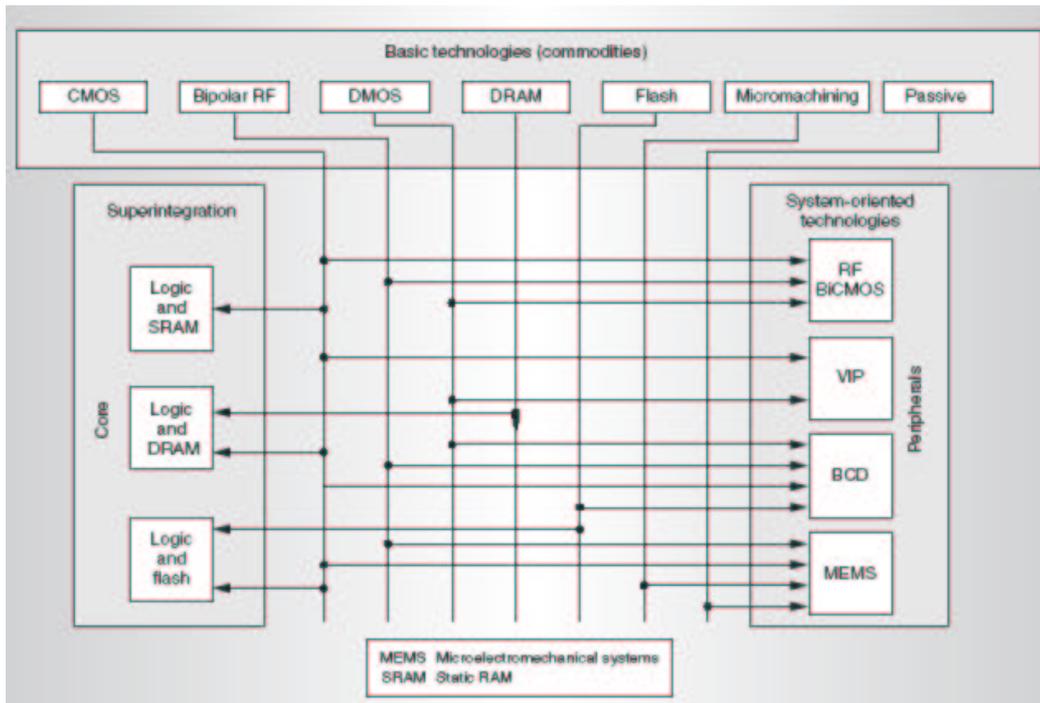


Figura 5.2: Combinando tra loro le varie tecnologie di base (elencate nel rettangolo in alto) é possibile realizzare le piú svariate funzionalità di calcolo e interfacciamento con l'esterno.

Material	Yield strength (GPa)	Knoop hardness (kg/mm ²)	Young's modulus (GPa)	Density (g/cm ³)	Thermal conductivity [kW/(cm-K)]	Thermal expansion (ppm/K)
Silicon	7	850	190	2.3	1.57	2.33
Carbon	53	7,000	1,030	3.5	20	1
Silicon carbide	21	2,480	700	3.2	3.5	3.3
Silicon nitride	14	3,486	390	3.1	0.19	0.8
Iron	12.6	400	196	7.8	0.8	12
Tungsten	4	485	410	19.3	1.78	4.5
Steel	2.1	660	200	7.9	0.32	17.3
Molybdenum	2.1	275	343	10.3	1.38	5
Aluminum	0.17	130	70	2.7	2.36	25

Figura 5.3: Principali proprietà termo-meccaniche del silicio.

ristiche abbastanza simili a quelle dell'acciaio. Ad esempio il silicio può tollerare ripetute deformazioni elastiche senza riportare alcun danno, in quanto possiede uno *sforzo di snervamento* (*Yield Strength*) simile all'acciaio. Tale parametro rappresenta il valore della pressione applicata a un corpo che segna il passaggio

da un comportamento elastico a un comportamento plastico. In altre parole una pressione di valore inferiore a tale limite permette al corpo di assumere nuovamente la forma originaria una volta che la forza deformante é rimossa, mentre una pressione maggiore deforma il corpo in maniera permanente.

5.3 Tipologie di microsistemi

I microsistemi elettromeccanici si possono suddividere in cinque grandi famiglie, corrispondenti a cinque diverse aree di applicazione:

1. dispositivi microfluidodinamici;
2. dispositivi inerziali;
3. dispositivi a radiofrequenza;
4. dispositivi microattuatori;
5. dispositivi ottici.

La figura 5.4 mostra alcune applicazioni tipiche corrispondenti alle varie famiglie appena elencate.

5.3.1 Dispositivi microfluidodinamici

Questi sistemi sono caratterizzati dalla capacità di manipolare piccolissime quantità di fluido. Le applicazioni più comuni sono ad esempio la realizzazione di stampanti a getto d'inchiostro e dispositivi per la diagnostica medica (i cosiddetti DNA-chip).

Il silicio rappresenta il materiale ideale per queste applicazioni, soprattutto grazie alle sue caratteristiche termiche, che permettono un rapido riscaldamento di alcune porzioni di materiale, e alla sua resistività termica ridotta, che permette di mantenere vaste porzioni di materiale a una temperatura costante. Le

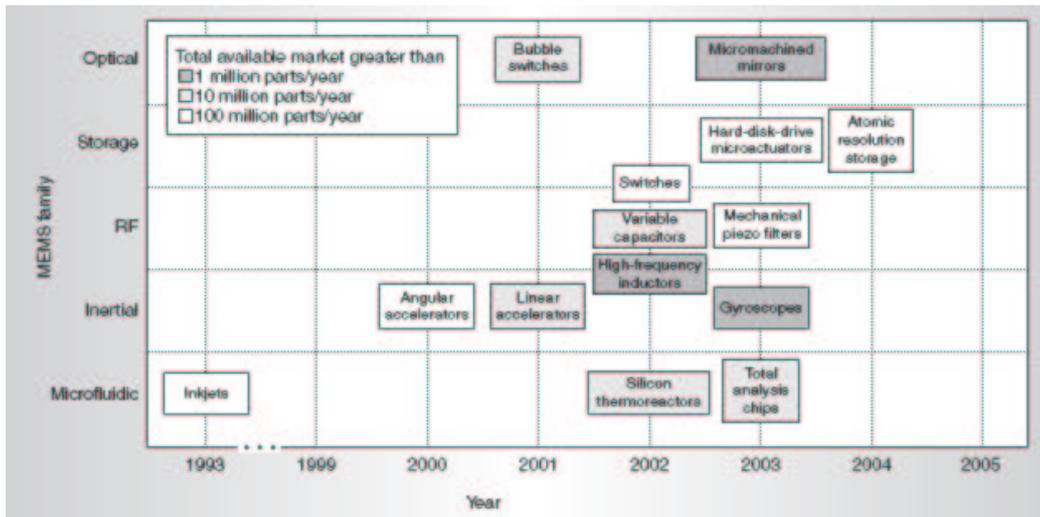


Figura 5.4: Applicazioni principali della cinque grandi famiglie di microsistemi.

piccole quantità di fluido da gestire vengono quindi controllate spesso attraverso la temperatura.

5.3.2 Dispositivi inerziali

Questi dispositivi sfruttano l'inerzia dei corpi in movimento. Solitamente sono costituiti da una parte fissa e da una parte in movimento: il sistema é in grado di percepire l'accelerazione di una parte rispetto all'altra misurando la variazione di capacità tra le due parti. I sensori inerziali possono essere sia lineari (in questo caso viene misurata l'accelerazione di traslazione) sia rotazionali (in questo caso viene misurata l'accelerazione angolare attorno a un asse di rotazione). La figura 5.5 mostra un esempio di accelerometro rotazionale.

5.3.3 Dispositivi a radiofrequenza

Questi dispositivi sono costituiti per la maggior parte da interruttori, condensatori variabili, risonatori e filtri.

Per quanto riguarda gli interruttori (vedi figura 5.6), essi sono comandati

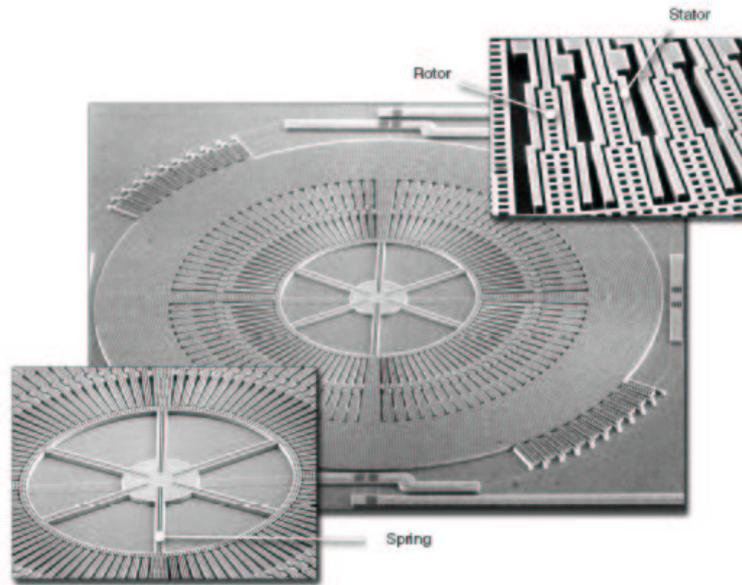


Figura 5.5: Esempio di accelerometro rotazionale.

da una azione di attrazione elettrostatica che permette di stabilire o meno il collegamento semplicemente deformando una lamina (*suspended beam*) in modo da stabilire o meno un contatto con degli elettrodi posti sotto di essa.

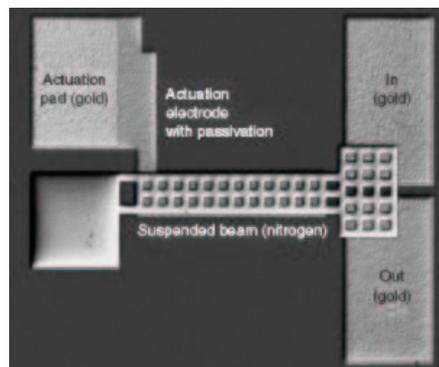


Figura 5.6: Esempio di interruttore utilizzato nei microsistemi a radiofrequenza.

Di solito gli interruttori appena descritti vengono utilizzati per comandare una batteria di condensatori, di diverso valore, a essi collegati. Infine si utilizzano risonatori e filtri di tipo meccanico, che hanno una stabilità e una selettività in

frequenza maggiore rispetto alle corrispondenti soluzioni elettroniche, soprattutto a causa della dissipazione di potenza di queste ultime.

5.3.4 Dispositivi microattuatori

Questa classe di dispositivi riguarda sostanzialmente il movimento controllato elettrostaticamente (ad esempio attraverso dispositivi piezoelettrici) su scala nanometrica. Una applicazione tipica di tali dispositivi risiede nei supporti di memorizzazione di massa: in un disco rigido, ad esempio, riuscire a controllare meglio piccoli spostamenti della testina permette di aumentare la densità dei dati e quindi il numero di dati immagazzinati a parità di superficie occupata.

5.3.5 Dispositivi ottici

Questi sistemi fanno largo uso di specchi mobili (per la realizzazione di interruttori di rete ottici) e interruttori a bolla (utilizzando la stessa tecnologia di base delle stampanti a getto d'inchiostro). Nel secondo caso una goccia (*bubble*) di fluido viene fatto evaporare in modo che essa modifichi le sue caratteristiche dielettriche, rifrangendo in diversi modi i raggi ottici che la colpiscono.

5.4 Aspetti economici che influenzano la realizzazione di microsistemi

Una breve analisi del mercato dei microsistemi e dei circuiti integrati (come detto infatti i microsistemi fanno largo uso delle tecnologie di integrazione) porta ad alcune considerazioni di tipo economico che influenzano le scelte di tipo più strettamente tecnico. Innanzitutto sono favoriti i progetti che realizzano dispositivi che saranno venduti in grandi volumi, in quanto un gran numero di vendite permette di ammortizzare gli elevati costi di progetto. Di conseguenza

i progettisti vengono notevolmente influenzati nella decisione di partizionare o meno il dispositivo su piú chip. Infatti anche quando un sistema potrebbe essere realizzato su un unico chip, economie di scala o vincoli di costo possono portare a scegliere una soluzione multi-chip. Queste considerazioni comportano anche la necessità di porre una grande attenzione sui sistemi di assemblaggio (*packaging*). Il risultato é la ricerca di nuovi metodi di interconnessione e posizionamento dei chip, realizzati con diverse tecnologie e poi sistemati all'interno di un unico package.

Appendice A

Richiami di calcolo delle probabilità

A.1 Concetto di variabile aleatoria

Dato uno spazio degli eventi

$$\Omega = \{\omega_1, \omega_2, \dots, \omega_n\} \quad (\text{A.1})$$

definiamo una corrispondenza che associ univocamente ciascun risultato ω_i dell'esperimento con un numero reale. Tale corrispondenza é una *variabile aleatoria* X se, per qualunque x reale, l'insieme dei risultati dell'esperimento per i quali $X \leq x$ é un evento a cui é possibile associare una probabilità.

Per una variabile aleatoria X si può definire una funzione di *distribuzione cumulativa di probabilità* (*cdf*)

$$P_X(x) = Pr\{X \leq x\} \quad (\text{A.2})$$

e una funzione *densità di probabilità* (*pdf*)

$$p_X(x) = \frac{dP_X(x)}{dx} \quad (\text{A.3})$$

Le due funzioni sono legate dalla relazione

$$P_X(x) = \int_{-\infty}^x p_X(\xi) d\xi \quad (\text{A.4})$$

e godono delle seguenti proprietà

- $0 \leq P_X(x) \leq 1$
- $\lim_{x \rightarrow +\infty} P_X(x) = 1 \Rightarrow \int_{-\infty}^{+\infty} p_X(\xi) d\xi$
- $\lim_{x \rightarrow -\infty} P_X(x) = 0$

Estendiamo ora questi concetti al caso di piú variabili aleatorie. Date n variabili aleatorie X_1, X_2, \dots, X_n , definiamo *funzione di distribuzione di densità congiunta* la quantità

$$P_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n) = Pr\{X_1 \leq x_1 \cap X_2 \leq x_2 \cap \dots \cap X_n \leq x_n\} \quad (\text{A.5})$$

e *funzione densità di probabilità congiunta* la quantità

$$p_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n) = \frac{\partial^n P_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n)}{\partial x_1 \partial x_2 \dots \partial x_n} \quad (\text{A.6})$$

A.2 Expectation, varianza e scarto tipo

Definiamo *expectation* della variabile aleatoria X la quantità

$$\bar{X} = E\{X\} = \int_{-\infty}^{+\infty} x p_X(x) dx \quad (\text{A.7})$$

mentre si definisce *varianza* la quantità

$$\sigma_X^2 = E\{(X - \bar{X})^2\} = \int_{-\infty}^{+\infty} (x - \bar{X})^2 p_X(x) dx \quad (\text{A.8})$$

Infine definiamo *deviazione standard* o *scarto tipo* della variabile aleatoria X la quantità

$$\sigma_X = \sqrt{\int_{-\infty}^{+\infty} (x - \bar{X})^2 p_X(x) dx} \quad (\text{A.9})$$

Questi valori non sono sempre facilmente calcolabili in quanto richiedono la conoscenza della funzione densità di probabilità. E' possibile definire allora degli *stimatori* di queste quantità che possono essere facilmente calcolati a partire da una serie di osservazioni X_1, X_2, \dots, X_n della variabile aleatorie X . In particolare definiamo la *media campione* come stimatore dell'expectation:

$$x = \frac{1}{n} \sum_{k=1}^n X_k \quad (\text{A.10})$$

E' intuitivo aspettarsi che questa sia la migliore stima possibile che possiamo fornire del misurando.

Come stimatore della varianza introduciamo la *varianza campione*

$$s_X^2 = \frac{1}{n-1} \sum_{k=1}^n (X_k - x)^2 \quad (\text{A.11})$$

Le stime sono esse stesse variabili aleatorie, in quanto dipendenti dalla scelta delle osservazioni X_1, X_2, \dots, X_n con cui sono calcolate. Come tali sono dotate a loro volta di un'expectation e di una varianza. In particolare, per la linearità dell'operatore $E\{\cdot\}$ risulta

$$E\{x\} = E\left\{\frac{1}{n} \sum_{k=1}^n X_k\right\} = \frac{1}{n} \sum_{k=1}^n E\{X_k\} = \bar{X} \quad (\text{A.12})$$

Notiamo che l'expectation della media campione coincide con il valore di cui essa stimatrice. Tutti gli stimatori che godono di questa proprietà si dicono *non polarizzati*. Si potrebbe facilmente dimostrare che anche lo stimatore della varianza é non polarizzato.

Si noti che per un numero molto elevato di osservazioni, le stime così definite tendono a coincidere con le quantità da stimare, ossia

- $x \rightarrow \bar{X}$ per $n \rightarrow \infty$,
- $s_X^2 \rightarrow \sigma_X^2$ per $n \rightarrow \infty$.

A.3 Teorema del limite centrale

Si tratta di un teorema di grande importanza, piú volte richiamato nei capitoli precedenti. Qui di seguito ci limitiamo a riportarne l'enunciato senza fornirne la dimostrazione.

Teorema A.3.1. *Dato un numero di variabili aleatorie indipendenti con densità di probabilità qualunque (purché non impulsive), la loro somma é una variabile aleatoria con una densità di probabilità gaussiana.*

Appendice B

La disuguaglianza di Schwartz

Riportiamo in questa appendice la dimostrazione della disuguaglianza di Schwartz

$$|s_{X_1 X_2}| \leq s_{X_1} s_{X_2} \quad (\text{B.1})$$

introdotta nel paragrafo 1.6.

Si consideri la seguente quantità definita positiva

$$\begin{aligned} A(t) &= \frac{1}{n} \sum_{i=1}^n [(X_{1,i} - \bar{X}_1) + t(X_{2,i} - \bar{X}_2)]^2 = \\ &= \frac{1}{n} \sum_{i=1}^n (X_{1,i} - \bar{X}_1)^2 + \frac{1}{n} t^2 \sum_{i=1}^n (X_{2,i} - \bar{X}_2)^2 + \\ &+ \frac{2}{n} t \sum_{i=1}^n [(X_{1,i} - \bar{X}_1) \cdot (X_{2,i} - \bar{X}_2)] \end{aligned} \quad (\text{B.2})$$

Calcoliamo ora per quali valori di t questa espressione raggiunge il minimo uguagliando a zero la derivata rispetto a t .

$$\begin{aligned} \frac{\partial}{\partial t} \left[\frac{1}{n} \sum_{i=1}^n (X_{1,i} - \bar{X}_1)^2 \right] + \frac{\partial}{\partial t} \left[\frac{t^2}{n} \sum_{i=1}^n (X_{2,i} - \bar{X}_2)^2 \right] + \\ + \frac{\partial}{\partial t} \left[\frac{2t}{n} \sum_{i=1}^n [(X_{1,i} - \bar{X}_1)(X_{2,i} - \bar{X}_2)] \right] = 0 \end{aligned} \quad (\text{B.3})$$

Svolgendo i calcoli della derivata si ottiene:

$$t \sum_{i=1}^n (X_{2,i} - \bar{X}_2)^2 + \sum_{i=1}^n [(X_{1,i} - \bar{X}_1)(X_{2,i} - \bar{X}_2)] = 0 \quad (\text{B.4})$$

ossia

$$t_{min} = - \frac{\sum_{i=1}^n [(X_{1,i} - \bar{X}_1)(X_{2,i} - \bar{X}_2)]}{\sum_{i=1}^n (X_{2,i} - \bar{X}_2)^2} \quad (\text{B.5})$$

Poiché $A(t)$ é una quantità definita positiva, anche il valore minimo A_{min} sarà sicuramente positivo:

$$\begin{aligned} A_{min} &= \frac{1}{n} \sum_{i=1}^n (X_{1,i} - \bar{X}_1)^2 + \frac{1}{n} t_{min}^2 \sum_{i=1}^n (X_{2,i} - \bar{X}_2)^2 + \\ &+ \frac{2}{n} t_{min} \sum_{i=1}^n [(X_{1,i} - \bar{X}_1)(X_{2,i} - \bar{X}_2)] \geq 0 \end{aligned} \quad (\text{B.6})$$

Sostituendo il valore di t_{min} espresso dalla relazione (B.5) si ha:

$$\begin{aligned} A_{min} &= \frac{1}{n} \sum_{i=1}^n (X_{1,i} - \bar{X}_1)^2 + \\ &+ \frac{1}{n} \left[- \frac{\sum_{i=1}^n [(X_{1,i} - \bar{X}_1)(X_{2,i} - \bar{X}_2)]}{\sum_{i=1}^n (X_{2,i} - \bar{X}_2)^2} \right]^2 \sum_{i=1}^n (X_{2,i} - \bar{X}_2)^2 + \\ &+ \frac{2}{n} \left[- \frac{\sum_{i=1}^n [(X_{1,i} - \bar{X}_1)(X_{2,i} - \bar{X}_2)]}{\sum_{i=1}^n (X_{2,i} - \bar{X}_2)^2} \right] \sum_{i=1}^n [(X_{1,i} - \bar{X}_1)(X_{2,i} - \bar{X}_2)] \geq 0 \end{aligned} \quad (\text{B.7})$$

ossia

$$\begin{aligned} A_{min} &= \frac{1}{n} \sum_{i=1}^n (X_{1,i} - \bar{X}_1)^2 + \frac{1}{n} \frac{\sum_{i=1}^n [(X_{1,i} - \bar{X}_1)(X_{2,i} - \bar{X}_2)]^2}{\sum_{i=1}^n (X_{2,i} - \bar{X}_2)^2} + \\ &- \frac{2}{n} \frac{\sum_{i=1}^n [(X_{1,i} - \bar{X}_1)(X_{2,i} - \bar{X}_2)]}{\sum_{i=1}^n (X_{2,i} - \bar{X}_2)^2} \sum_{i=1}^n [(X_{1,i} - \bar{X}_1)(X_{2,i} - \bar{X}_2)] \geq 0 \end{aligned} \quad (\text{B.8})$$

Quindi alla fine abbiamo ricavato

$$A_{min} = \frac{1}{n} \sum_{i=1}^n (X_{1,i} - \bar{X}_1)^2 - \frac{1}{n} \frac{\sum_{i=1}^n [(X_{1,i} - \bar{X}_1)(X_{2,i} - \bar{X}_2)]^2}{\sum_{i=1}^n (X_{2,i} - \bar{X}_2)^2} \geq 0 \quad (\text{B.9})$$

Posto ora

$$\begin{aligned} s_{X_1 X_2} &= \frac{1}{n} \sum_{i=1}^n [(X_{1,i} - \bar{X}_1)(X_{2,i} - \bar{X}_2)] \\ s_{X_1}^2 &= \frac{1}{n} \sum_{i=1}^n (X_{1,i} - \bar{X}_1)^2 \\ s_{X_2}^2 &= \frac{1}{n} \sum_{i=1}^n (X_{2,i} - \bar{X}_2)^2 \end{aligned} \tag{B.10}$$

si ottiene quindi

$$\begin{aligned} s_{X_1}^2 - \frac{s_{X_1 X_2}^2}{s_{X_2}^2} &\geq 0 \\ s_{X_1}^2 s_{X_2}^2 &\geq s_{X_1 X_2}^2 \end{aligned} \tag{B.11}$$

da cui la disuguaglianza finale

$$|s_{X_1 X_2}| \leq s_{X_1} s_{X_2} \tag{B.12}$$

Appendice C

Il limite di Cramer-Rao

Ci proponiamo di dimostrare il limite di Cramer-Rao. Si parte da uno stimatore non polarizzato, pertanto il valor medio dell'errore sarà per definizione nullo:

$$\int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} (\hat{\alpha} - \alpha) p(\mathbf{X}|\alpha) d\mathbf{X} = 0 \quad (\text{C.1})$$

Derivando l'equazione precedente rispetto ad α e uguagliando il tutto a zero si ottiene

$$\begin{aligned} \frac{d}{d\alpha} \left[\int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} (\hat{\alpha} - \alpha) p(\mathbf{X}|\alpha) d\mathbf{X} \right] = \\ \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} (\hat{\alpha} - \alpha) \frac{d}{d\alpha} p(\mathbf{X}|\alpha) d\mathbf{X} - \underbrace{\int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} p(\mathbf{X}|\alpha) d\mathbf{X}}_1 = 0 \end{aligned} \quad (\text{C.2})$$

quindi

$$\begin{aligned} \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} (\hat{\alpha} - \alpha) \frac{d}{d\alpha} p(\mathbf{X}|\alpha) d\mathbf{X} = 1 \\ \frac{d}{d\alpha} p(\mathbf{X}|\alpha) = p(\mathbf{X}|\alpha) \frac{d}{d\alpha} \ln p(\mathbf{X}|\alpha) \\ \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} (\hat{\alpha} - \alpha) p(\mathbf{X}|\alpha) \frac{d}{d\alpha} \ln p(\mathbf{X}|\alpha) d\mathbf{X} = 1 \end{aligned} \quad (\text{C.3})$$

Per dimostrare il limite si ricorre a una maggiorazione. Usando la disugua-

gianza di Schwartz (si veda al riguardo l'appendice B), si ottiene:

$$\left| \int_{-\infty}^{+\infty} U(X)V(X)dX \right|^2 \leq \left(\int_{-\infty}^{+\infty} |U(X)|^2 dX \right) \cdot \left(\int_{-\infty}^{+\infty} |V(X)|^2 dX \right) \quad (C.4)$$

Posto

$$U(X) = (\hat{\alpha} - \alpha) \cdot p^{\frac{1}{2}}(\mathbf{X}|\alpha) \quad (C.5)$$

$$V(X) = \frac{d}{d\alpha} \ln p(\mathbf{X}|\alpha) \cdot p^{\frac{1}{2}}(\mathbf{X}|\alpha) \quad (C.6)$$

e sostituendo nell'equazione (C.4) si ha:

$$\begin{aligned} & \left| \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} (\hat{\alpha} - \alpha) p(\mathbf{X}|\alpha) \frac{d}{d\alpha} \ln p(\mathbf{X}|\alpha) d\mathbf{X} \right|^2 \leq \\ & \leq \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} (\hat{\alpha} - \alpha)^2 p(\mathbf{X}|\alpha) d\mathbf{X} \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} \left[\frac{d}{d\alpha} \ln p(\mathbf{X}|\alpha) \right]^2 p(\mathbf{X}|\alpha) d\mathbf{X} \end{aligned} \quad (C.7)$$

Quindi

$$\int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} (\hat{\alpha} - \alpha)^2 p(\mathbf{X}|\alpha) g\mathbf{X} \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} \left[\frac{d}{d\alpha} \ln p(\mathbf{X}|\alpha) \right]^2 p(\mathbf{X}|\alpha) d\mathbf{X} \geq 1 \quad (C.8)$$

Poiché lo stimatore é non polarizzato, la stima $\hat{\alpha}$ coincide con il valor medio di α , quindi

$$\sigma_{\hat{\alpha}}^2 = \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} (\alpha - \bar{\alpha})^2 p(\mathbf{X}|\alpha) g\mathbf{X} = \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} (\hat{\alpha} - \alpha)^2 p(\mathbf{X}|\alpha) g\mathbf{X} \quad (C.9)$$

da cui

$$\sigma_{\hat{\alpha}}^2 \geq \frac{1}{\int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} \left[\frac{d}{d\alpha} \ln p(\mathbf{X}|\alpha) \right]^2 p(\mathbf{X}|\alpha) dX_0 \cdots dX_{N-1}} \quad (C.10)$$

L'uguaglianza nella relazione precedente si ottiene solamente nel caso in cui si abbia

$$U(X) = \lambda V(X) \quad (C.11)$$

ossia quando tra $U(X)$ e $V(X)$ é presente un legame di proporzionalitá attraverso una quantitá λ che non dipende da X .

Appendice D

Il teorema di De Moivre–Laplace

Questo teorema permette di dimostrare come per $n \rightarrow \infty$ la distribuzione binomiale $b_{n,p}(\nu)$ tenda a una distribuzione gaussiana con stesso valor medio e stessa varianza.

Definiamo innanzitutto la funzione di densità normale come

$$\psi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \quad (\text{D.1})$$

e la funzione di distribuzione normale come il suo integrale

$$\Psi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{1}{2}y^2} dy \quad (\text{D.2})$$

Come mostrato nella (2.6) il *termine centrale* della distribuzione binomiale é l'unico numero intero che può essere scritto nella forma

$$m = np + \delta \quad \text{con} \quad -q < \delta \leq p \quad (\text{D.3})$$

Definiamo la quantità

$$a_k = b_{n,p}(m+k) = \binom{n}{m+k} p^{m+k} q^{n-m-k} \quad (\text{D.4})$$

Per semplicità consideriamo solo il caso $k > 0$. Ragionamenti analoghi possono essere applicati anche al caso $k < 0$ (in alternativa é possibile ricondursi al caso $k > 0$ scambiando p con q).

A partire dalla definizione (D.4) si ricava che il termine a_k può essere scritto in funzione di a_0 in questo modo:

$$a_k = a_0 \frac{(n-m)(n-m-1)\cdots(n-m-k+1)p^k}{(m+1)(m+2)\cdots(m+k)q^k} \quad (\text{D.5})$$

Effettuando la sostituzione

$$t_j = \frac{j + \delta + q}{(n+1)pq} \quad (\text{D.6})$$

la (D.5) può essere riscritta come

$$a_k = a_0 \frac{(1-pt_0)(1-pt_1)\cdots(1-pt_{k-1})}{(1+qt_0)(1+qt_1)\cdots(1+qt_{k-1})} \quad (\text{D.7})$$

Nel seguito utilizzeremo la (D.7) per valori di k tali per cui t_k sia sufficientemente piccolo, ad esempio minore di $\frac{1}{2}$.

Sfruttando l'espansione di Taylor per il logaritmo

$$\ln(1+x) = x - \frac{1}{2}x^2 + \frac{1}{3}x^3 - \frac{1}{4}x^4 + \cdots \quad (\text{D.8})$$

si ricava

$$\frac{1-pt_j}{1+qt_j} = e^{-t_j+\cdots} \quad (\text{D.9})$$

dove la quantità omissa è in valore assoluto minore di t_j^2 . Pertanto è possibile scrivere

$$a_k = a_0 e^{-(t_0+\cdots+t_{k-1})+\cdots} \quad (\text{D.10})$$

dove è stata omissa una quantità che è in valore assoluto minore di $kt_{k-1}^2 < \frac{k^3}{(npq)^2}$.

Poiché

$$t_0 + t_1 + \cdots + t_{k-1} = \frac{\frac{1}{2}k(k-1) + k(\delta+q)}{(n+1)pq} \quad (\text{D.11})$$

sostituendo il secondo membro con $\frac{k^2}{2npq}$ commettiamo un errore minore di $\frac{2k}{npq}$.

Effettuando questa approssimazione la (D.10) diventa

$$a_k = a_0 e^{\frac{-k^2}{(2npq)+\rho_k}} \quad (\text{D.12})$$

dove il termine di errore ρ_k soddisfa la diseguaglianza

$$|\rho_k| < \frac{k^3}{(npq)^2} + \frac{2k}{npq} \quad (D.13)$$

Mostreremo che

$$a_0 = \frac{n!}{m!(n-m)!} p^m q^{n-m} \sim \frac{1}{\sqrt{2\pi npq}} \quad (D.14)$$

Nel caso in cui $p = \frac{m}{n}$ la stima (D.14) é una conseguenza immediata della formula di Stirling

$$n! \simeq \sqrt{2\pi n} n^{n+\frac{1}{2}} e^{-n} \quad \text{per} \quad n \rightarrow \infty \quad (D.15)$$

Una semplice differenziazione mostra che il termine centrale nella (D.14) assume il valore massimo quando $p = \frac{m}{n}$. Per un dato m dobbiamo considerare solo i valori di p per i quali la relazione (D.3) é verificata; pertanto il minimo di a_0 é assunto essere agli estremi, ossia per $p = \frac{m}{n+1}$ o $p = \frac{m+1}{n+1}$. Con questi valori per p , una applicazione diretta della formula di Stirling conduce alla (D.14) a patto di sostituire n con $n+1$. Ne segue che la (D.14) é valida per tutti i possibili valori di p .

Posto

$$h = \frac{1}{\sqrt{2\pi npq}} \quad (D.16)$$

la (D.12) mostra che

$$a_k \sim h\psi(kh) \quad (D.17)$$

Teorema D.0.2. *Se $n \rightarrow \infty$ e $k < K_n$ con K_n tale che $\frac{K_n^2}{n^2} \rightarrow 0$, allora la (D.17) é valida uniformemente in k . Pertanto, per ogni $\epsilon > 0$ e n sufficientemente grande si ha*

$$1 - \epsilon < \frac{a_k}{h\psi(kh)} < 1 + \epsilon \quad (D.18)$$

La principale applicazione del teorema precedente consiste nell'ottenere approssimazioni di probabilitá nella forma

$$\Pr\{\alpha \leq S_n \leq \beta\} = \sum_{\nu=\alpha}^{\beta} b_{n,p}(\nu) = \sum_{k=\alpha-m}^{\beta-m} a_k \quad (D.19)$$

Nei limiti di applicabilità del teorema precedente, si ottiene una buona approssimazione sostituendo il termine a_k con il termine $h\psi(kh)$. Questa quantità può essere interpretata come l'area di un rettangolo di altezza $\psi(kh)$ la cui base è un intervallo di lunghezza h centrato in kh (vedi figura D.1). L'approssimazione consiste nel rimpiazzare l'area del rettangolo con la corrispondente area sottesa dalla curva $\psi(x)$. L'errore commesso si riduce sempre più al diminuire di h .

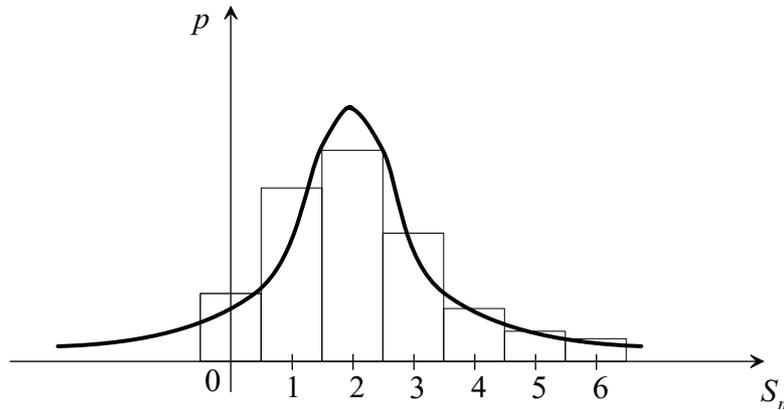


Figura D.1: La distribuzione binomiale approssimata con una distribuzione normale. La funzione a gradino fornisce la probabilità $b_{10,0.2}(k)$ di k successi in dieci prove con $p = 0.2$. La curva continua fornisce per ogni intero k la corrispondente approssimazione normale.

Quando α e β sono interi si ottiene

$$\Pr\{\alpha \leq S_n \leq \beta\} \simeq \Psi \left[\left(\alpha - m + \frac{1}{2} \right) h \right] - \Psi \left[\left(\beta - m - \frac{1}{2} \right) h \right] \quad (\text{D.20})$$

Posto $z_1 = \alpha - np$ e $z_2 = \beta - np$ si ottiene la formulazione definitiva del teorema di De Moivre-Laplace:

Teorema D.0.3 (Teorema limite di De Moivre-Laplace). *Dati due valori z_1 e z_2 , per $n \rightarrow \infty$ si ha*

$$\Pr\{np + z_1\sqrt{npq} \leq S_n \leq np + z_2\sqrt{npq}\} \rightarrow \Psi(z_2) - \Psi(z_1) \quad (\text{D.21})$$

Posto

$$S_n^* = \frac{S_n - np}{\sqrt{npq}} \quad (\text{D.22})$$

(dove np é il valor medio della variabile aleatoria S_n e npq la sua varianza) la relazione (D.21) diventa

$$\Pr\{z_1 \leq S_n^* \leq z_2\} \rightarrow \Psi(z_2) - \Psi(z_1) \quad (\text{D.23})$$

Questa forma mostra come per valori grandi di n , il valore della probabilità $\Pr\{z_1 \leq S_n^* \leq z_2\}$ é praticamente indipendente da p .

Bibliografia

- [1] William Feller. *An Introduction to Probability Theory and Its Application*. John Wiley & Sons, 1968.
- [2] Bruno Murari. Integrating nonelectronic components into electronic microsystems. *IEEE Micro*, May–June:34–44, 2003.
- [3] Athanasios Papoulis. *Probability, random variables and stochastic processes*. McGraw Hill International Editions, 1991.
- [4] John R. Taylor. *Introduzione all'analisi degli errori: lo studio delle incertezze nelle misure fisiche*. Zanichelli Editore, 2000.