

Appunti del corso di Fondamenti di Telecomunicazioni

Maurizio di Bisceglie

Università degli Studi del Sannio
Piazza Roma 21
Benevento
Italy

19 marzo 2018

Prefazione

Prefazione

Indice

Chapter 1	Introduzione ai sistemi di trasmissione numerica	1
1.1	Introduzione	1
1.2	Modello di un sistema di trasmissione numerica	2
Chapter 2	Elementi di Teoria della Probabilità	13
2.1	Esperimento aleatorio	14
2.1.1	Spazio dei campioni	14
2.1.2	Eventi	16
2.1.3	Gli assiomi della probabilità	21
2.1.4	Definizioni alternative di probabilità	27
2.2	Probabilità condizionata e indipendenza	29
2.2.1	Probabilità condizionata	29
2.2.2	Leggi Fondamentali	31
2.3	Variabili aleatorie	33
2.4	Funzione di Distribuzione Cumulativa	35
2.4.1	CDF empirica	38
2.4.2	Proprietà della CDF	39
2.5	Funzione di distribuzione di probabilità	39
2.6	Funzione di densità di probabilità	42
2.6.1	La funzione $Q(\cdot)$	48
2.7	Media e momenti di una variabile aleatoria	53
2.8	Caratterizzazione congiunta	56
2.8.1	CDF congiunta di due variabili aleatorie	56
2.8.2	PDF e PMF congiunta di due variabili aleatorie	57
2.9	Distribuzioni e densità condizionate	57
2.9.1	CDF condizionata ad un evento	58
2.9.2	CDF di X condizionata all'evento $\{Y \leq y\}$	59
2.9.3	PMF condizionata	59
2.9.4	PDF condizionata	59

2.9.5 Leggi della probabilità per la PDF	60
2.10 Variabili aleatorie indipendenti	60
2.11 Momenti congiunti di due variabili aleatorie	61
2.11.1 Incorrelazione e indipendenza	61
2.12 Stime basate su sequenze di osservazioni	64
2.12.1 Media campionaria	65
2.12.2 Legge dei grandi numeri	67
2.13 Esercizi proposti	69
 Chapter 3 Segnali e sistemi nel dominio del tempo	 77
3.1 Introduzione	77
3.2 Classificazione dei segnali	78
3.3 Segnali deterministici elementari	79
3.3.1 Impulso o finestra rettangolare	79
3.3.2 Gradino unitario	80
3.3.3 Fasore (tempo continuo)	81
3.3.4 Segnale sinusoidale (tempo continuo)	81
3.3.5 Sequenza esponenziale (tempo discreto)	82
3.3.6 Fasore (tempo discreto)	82
3.3.7 Impulso di tipo sinc	83
3.3.8 Impulso discreto	85
3.3.9 Impulso continuo (impulso di Dirac)	86
3.4 Medie temporali	89
3.4.1 Rappresentazione dei segnali	94
3.4.2 Funzioni di correlazione	95
3.5 Sistemi nel dominio del tempo	98
3.6 Proprietà	100
3.6.1 Dispersività	101
3.6.2 Causalità	101
3.6.3 Invertibilità	101
3.6.4 Invarianza temporale	102
3.6.5 Stabilità	102
3.6.6 Linearità	103
3.7 Somma e integrale di convoluzione	103
3.8 Legami ingresso-uscita per le funzioni di correlazione	114
 Chapter 4 Segnali e sistemi nel dominio della frequenza	 121
4.1 Risposta in frequenza di un sistema LTI	123
4.2 Trasformata di Fourier	126
4.3 Esempi di trasformate	128
4.4 Trasformata di Fourier al limite	137
4.5 Proprietà della trasformata di Fourier	139
4.5.1 Linearità	140

4.5.2	Proprietà di simmetria della trasformata di Fourier	140
4.5.3	Cambiamento di scala	140
4.5.4	Traslazione nel dominio del tempo	143
4.5.5	Traslazione nel dominio della frequenza (modulazione)	145
4.5.6	Convoluzione	149
4.5.7	Replicazione e campionamento	151
4.5.8	Spettro di segnali periodici	155
4.6	Spettri di energia	157
4.6.1	Densità spettrale di potenza	162
4.6.2	Spettri mutui di energia e di potenza	163
4.6.3	Processi bianchi	164
4.6.4	Teorema di Wiener-Kintchine	165
4.7	Legami ingresso-uscita per le \mathcal{PSD}	166
Chapter 5	Il campionamento dei segnali	171
5.1	Campionamento ideale	171
5.2	Campionamento reale	176
5.3	CONVERSIONE ANALOGICO/NUMERICA	183
5.3.1	Conversione t/n	183
5.3.2	Quantizzazione e codifica	184
Chapter 6	Ricezione ottima su canale AWGN	191
6.1	Spazio dei segnali trasmessi	191
6.1.1	Costellazione dei segnali	197
6.2	Procedura di ortonormalizzazione di Gram-Schmidt	197
6.2.1	Ortonormalizzazione in \mathbb{R}^2	197
6.3	Ricezione ottima in rumore AWGN	202
6.4	Segnalazione binaria antipodale	211
6.5	Segnalazione PAM M-aria	214
6.5.1	Probabilità di errore in funzione del contrasto di energia	217
6.5.2	Probabilità di errore in funzione della distanza	220
6.6	Phase Shift Keying (PSK)	223
6.6.1	Calcolo della probabilità di errore	225
6.7	Segnalazione ortogonale	227
6.8	Parametri di un sistema di trasmissione	236

Capitolo 1

Introduzione ai sistemi di trasmissione numerica

"The single biggest problem in communication is the illusion that it has taken place."

George Bernard Shaw

1.1 INTRODUZIONE

L'informazione è presente nella vita quotidiana in molte forme diverse: voce, musica, video, file di dati, pagine web, e viene trasferita da un punto all'altro in molteplici modi e con velocità diversa a seconda delle esigenze. Altrettanto importante, anche se meno tradizionale, è l'esigenza di conservare l'informazione nel tempo mediante la memorizzazione e il recupero di dati da supporti magnetici e ottici.

L'informazione trasmessa è in genere associata ad segnale elettrico variabile nel tempo ed è quindi rappresentabile mediante una funzione del tempo $s(t)$. È opportuno sottolineare come al giorno d'oggi sia possibile trasformare una qualsiasi informazione in un segnale elettrico, per cui considerare come informazione da trasmettere un segnale elettrico non è riduttivo e/o limitativo. È, invece, particolarmente interessante considerare il modo in cui l'informazione può essere rappresentata, specificatamente, in forma analogica o in forma numerica. La rappresentazione è analogica quando l'informazione è associata ad un parametro del segnale che assume valori in un insieme continuo: ad esempio, il segnale in uscita da un microfono è analogico in quanto esso può assumere tutti i valori appartenenti ad un determinato intervallo di valori. L'informazione è invece numerica quando è associata a parametri che possono assumere solo un numero finito di valori, cioè appartenenti ad un insieme discreto. Ad esempio, se si considera l'informazione contenuta in una sequenza di 4 bit e si associano alle possibili configurazioni 00-01-10-11 le ampiezze -3A,-A,A,*3A di una forma d'onda sinusoidale si ottiene un insieme di segnali numerici. Analogamente, l'informazione associata al lancio di un dado è costituita dall'insieme dei numeri da uno a sei, e tale informazione potrà essere associata ad un segnale $s(t)$ costante a tratti la cui ampiezza può assumere sei possibili valori oppure ad una sinusoide la cui frequenza può assumere sei possibili valori. Gli ultimi 20 anni hanno visto una progressiva sostituzione di tutti gli apparati di trasmissione analogica

con sistemi di trasmissione numerica; le motivazioni di ciò posso essere riassunte nei seguenti punti.

1. L'informazione analogica può essere convertita in forma numerica mediante operazioni di campionamento e quantizzazione. Tale processo di conversione comporta una perdita irreversibile di informazione per cui non è più possibile ricostruire il segnale originario $s(t)$ a partire dalla sua versione numerica. Tuttavia, il processo di conversione analogico-numerica introduce una perdita di informazione che può essere resa arbitrariamente piccola o comunque determinata attraverso specifiche di progetto. I sistemi di trasmissione numerica possono essere quindi utilizzati per la trasmissione di informazioni originariamente analogiche che sono state sottoposte ad un processo di conversione analogico-numerica. La rete di telefonia fissa, inizialmente progettata e realizzata come un sistema di comunicazione analogico è oggi quasi universalmente stata convertita in un sistema di trasmissione numerica; in particolare, il segnale vocale viene convertito in un flusso di dati numerico con un bit-rate di 64Kbit/s. Anche la rete di telefonia cellulare GSM si basa sull'utilizzo di una modulazione numerica con un bit-rate 13Kbit/s.
2. sorgenti analogiche di natura diverse possono essere indistintamente convertite in sequenze di bit, per cui differenti sorgenti di informazione possono essere trasmesse simultaneamente (ad esempio, nella multiplazione a divisione di tempo, i bit appartenenti a diversi messaggi vengono alternati prima del processo di modulazione).
3. Il fatto che ogni sorgente analogica possa essere convertita in una sequenza di simboli binari, unito al fatto che i simboli binari sono le unità di informazione elementari utilizzate dai computer, fa sì che molte delle operazioni realizzate in una catena di trasmissione e ricezione di un sistema di trasmissione numerica possano essere effettuate ricorrendo a sistemi di elaborazione di dati. Questi ultimi, inoltre, sono sempre più spesso realizzati mediante sistemi hardware programmabili (DSP e FPGA) di largo consumo.
4. in presenza di rumore, i sistemi di trasmissione numerica possono essere resi di gran lunga più affidabili dei sistemi di comunicazione analogica.

1.2 MODELLO DI UN SISTEMA DI TRASMISSIONE NUMERICA

Cominciamo ad illustrare lo schema a blocchi di un sistema di trasmissione numerica, illustrato in figura 1.1.

1.2.1

Sorgente numerica

La Sorgente numerica rappresenta la fonte dell'informazione che si vuole trasmettere. I simboli emessi dalla sorgente possono rappresentare già per se l'informazione in forma numerica (ad esempio, un file di dati), o provenire da una sorgente analogica che è stata convertita in forma numerica. I simboli emessi dalla sorgente appartengono, in genere, ad un

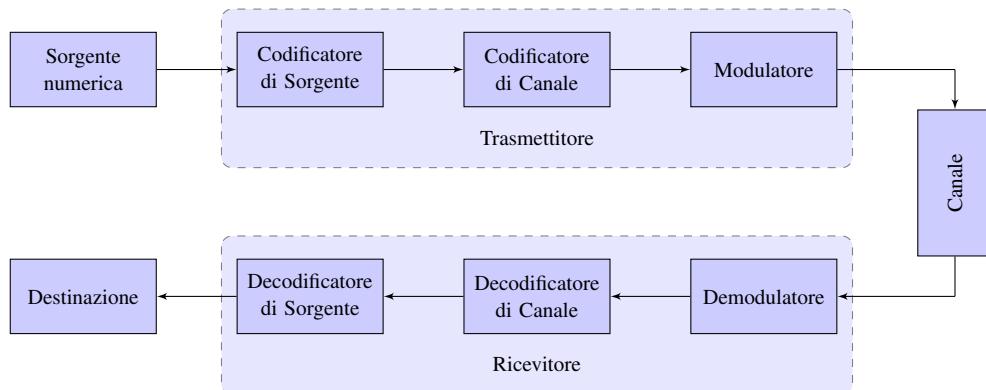


Figura 1.1: Sistema di comunicazione numerico.

alfabeto binario.

1.2.2

Codificatore di sorgente

Ha il compito di rappresentare la sequenza di simboli emessi dalla sorgente mediante una nuova sequenza di simboli avente la minima lunghezza possibile. In altre parole, il codificatore di sorgente ha il compito di rimuovere la ridondanza presente nella sorgente di informazione, in maniera tale che questa possa essere trasmessa con la massima efficienza e senza perdita di informazione. Per chiarire le idee, supponiamo che la sorgente emetta quattro possibili simboli con probabilità differente: i simboli (00 01 10 11) vengono emessi, rispettivamente, con probabilità (1/2 1/4 1/8 1/8). Evidentemente conviene associare ai simboli 00 e 01 sequenze più brevi (0 e 10 per esempio) e ai simboli 10 e 11 sequenze più lunghe (110 e 111 per esempio). Questo consente di ridurre la lunghezza media della parola codice, data dalla somma delle lunghezze delle parole codice pesate per la propria probabilità di essere emesse, che nel primo caso è pari a $2 \times 0.5 + 2 \times 0.25 + 2 \times 0.125 + 2 \times 0.125 = 2$ e nel secondo caso a $1 \times 0.5 + 2 \times 0.25 + 3 \times 0.125 + 3 \times 0.125 = 1.75$ (*compressione lossless*).

Quando è possibile tollerare una perdita di informazione si parla di compressione con distorsione o *lossy*. Questo tipo di compressione consente di ottenere un livello di compattazione più elevato nel caso in cui l'utente finale non può distinguere o può accettare un certo livello di degrado dell'informazione. Ciò avviene comunemente nel caso di sorgenti musicali oppure di sorgenti video.

Codificatore di canale

1.2.3

$$H = -\sum_{i=1}^n p_i \cdot \log_2 p_i$$

H è il valore medio
di parola massimo per
avere una trasmissione
Lossless

Il codificatore di canale aggiunge alla sequenza di bit in uscita al codificatore di sorgente dei bit di ridondanza al fine di rendere l'informazione meno vulnerabile agli errori che possono verificarsi durante la fase di ricezione dell'informazione effettuata alla destinazione. Un esempio di codifica di canale si ha con la semplice aggiunta di bit di parità. Aggiungendo dei bit di parità in modo tale che sequenze di bit di opportuna lunghezza abbiano un numero pari di valori 1 aiuta ad individuare le sequenze di bit che sono affette da errori dovuti alla trasmissione del segnale sul canale. Sebbene possa sembrare contraddittorio rimuovere la ridondanza tramite la codifica di sorgente e poi aggiungerla tramite la codifica di canale, va notato che il codificatore aggiunge una ridondanza strutturata e nota al ricevitore, che può quindi avvalersi di tale conoscenza per rivelare gli errori e in molti casi correggerli. Si noti infine che l'insieme dei blocchi "sorgente", "codificatore di sorgente" e "codificatore di canale" costituisce una sorgente binaria equivalente, così come illustrato in figura 1.2.

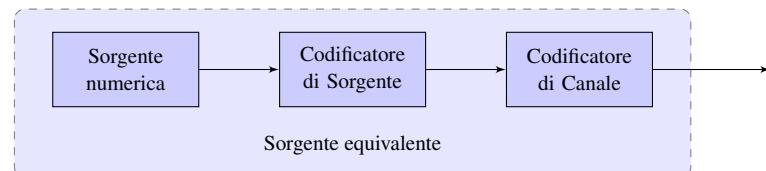


Figura 1.2: Sorgente numerica equivalente.

1.2.4

Modulatore

Nella sua forma più semplice, il modulatore è un dispositivo che genera una sequenza di forme d'onda con cadenza T_s , associando ad ognuna tra le possibili $M = 2^k$ sequenze di k simboli binari una tra M possibili forme d'onda di durata T . Il modulatore è definito dai seguenti parametri:

- L'intervallo di simbolo T_s , cioè il tempo che intercorre tra l'emissione di un simbolo di k cifre binarie ed il successivo. L'inverso di T_s è il tasso o frequenza di simbolo (*symbol-rate*) indicato con R_s , che misura il numero di simboli emessi nell'unità di tempo; R_s si misura in simboli al secondo o anche baud. Per $k = 1$ si ottiene la semplice emissione di simboli binari. Il tempo intercorrente tra l'emissione di due bit è detto *intervallo di bit* e si indica con T_b , mentre il suo inverso è comunemente detto *bit-rate* e si indica con R_b . Il bit-rate esprime il numero di bit al secondo emessi dalla sorgente.
- La cardinalità M . È un numero intero, usualmente una potenza di 2, $M = 2^k$. Se $M = 2$ il modulatore si dice binario, altrimenti si parlerà di modulatore M -ario.

- Un insieme di M forme d'onda $\mathcal{S} = \{s_1(t), s_2(t), \dots, s_M(t)\}$. Si assumerà che tali forme d'onda abbiano durata limitata, cioè siano non nulle solo per t appartenente all'intervallo $[0, T]$, e che siano di quadrato sommabile, ovvero siano segnali di energia. In particolare, indicheremo con

$$\mathcal{E}_i = \int_{-\infty}^{\infty} s_i^2(t) dt = \int_0^T s_i^2(t) dt \quad (1.1)$$

l'energia associata all' i -esima forma d'onda $s_i(t)$ del modulatore.

- Una funzione biunivoca che associa a sequenze binarie di lunghezza $k = \log_2 M$ una tra le M forme d'onda a disposizione del modulatore.

Se $M = 4$, il modulatore ha a disposizione 4 forme d'onda $s_1(t), \dots, s_4(t)$, e può effettuare la seguente corrispondenza

$$00 \rightarrow s_1(t), \quad 01 \rightarrow s_2(t), \quad 10 \rightarrow s_3(t), \quad 11 \rightarrow s_4(t). \quad (1.2)$$

Se il codificatore di canale produce la sequenza di bit 011100, il modulatore trasmette sul canale il segnale

$$s_2(t) + s_4(t - T) + s_1(t - 2T) \quad (1.3)$$

Affinché il sistema di trasmissione possa funzionare con continuità, è necessario che la durata T delle forme d'onda del modulatore (usualmente indicata col termine di intervallo di segnalazione) sia non maggiore del tempo che il codificatore di canale impiega a produrre i simboli binari. In tal modo, quando in uscita al codificatore di canale si presenta una sequenza di k simboli binari, il modulatore invia sul canale una delle forme d'onda a sua disposizione, e tale trasmissione termina prima che una nuova sequenza sia a nuovamente disposizione. È opportuno notare che tra modulatore e canale sono presenti usualmente dei blocchi aggiuntivi, quali ad esempio amplificatori, convertitori a radio frequenza e, nel caso di trasmissione sul canale radio, un'antenna trasmittente. Nel seguito daremo per scontata la presenza di tali dispositivi la cui definizione e progettazione non è trattata in queste note. È opportuno invece citare che il modulatore numerico esercita una funzione di adattamento della forma d'onda generata al canale di trasmissione; tale operazione si realizza modificando i parametri di un segnale sinusoidale, detto portante.

Numerosi sono i motivi che rendono opportuna o necessaria la modulazione mediante una portante:

1. Per irradiare i segnali in modo efficiente è necessario che le antenne abbiano dimensioni confrontabili con la lunghezza d'onda della radiazione da trasmettere; quindi per segnali di tipo passa-basso (ad esempio il segnale vocale la cui banda va dai 300 Hz ai 3400 Hz), occorrerebbero antenne di dimensioni non fisicamente realizzabili. Con la modulazione, lo spettro del segnale di informazione viene traslato nell'intorno della frequenza portante; l'operazione di modulazione consente, quindi di utilizzare antenne efficienti e fisicamente realizzabili.

Mezzo trasmittivo	Banda
Linea bifilare	1 - 300 kHz
Cavo coassiale	300 kHz - 1 GHz
Guida d'onda	1 - 300 GHz
Fibra ottica	0.6-1.6 μm

Tabella 1.1

Mezzi trasmittivi che realizzano la propagazione guidata delle onde elettromagnetiche e intervalli di frequenze in cui è possibile utilizzarli (valori indicativi)

2. La banda passante dei dispositivi utilizzati in un sistema di comunicazione (amplificatori, filtri, canale di comunicazione, etc.) deve contenere la banda del segnale da elaborare. D'altro canto la pratica realizzazione di un sistema di amplificazione necessita di imporre che esso sia a banda stretta. Quindi, per trasmettere su bande larghe senza violare l'ipotesi di banda stretta del sistema è necessario che la frequenza della portante f_c sia sufficientemente elevata.
3. La modulazione consente la condivisione del canale tra più utenti; stazioni radio differenti possono, ad esempio, coprire la stessa area geografica se trasmettono utilizzando porzioni diverse dello spettro elettromagnetico; i segnali possono essere separati in ricezione con l'ausilio di appositi filtri. Più in generale, le modulazioni analogiche e quelle numeriche consentono di trasmettere sullo stesso canale (ed eventualmente con lo stesso trasmettitore) più segnali contemporaneamente (nel caso si utilizzi lo stesso trasmettitore si costruisce un segnale multiplex).

1.2.5

Canale di trasmissione

Il canale di trasmissione è il mezzo fisico che realizza la connessione tra la sorgente e la destinazione. Possiamo distinguere tra canali di tipo wireline (come la linea bifilare o intrecciata, il cavo coassiale, la guida d'onda la fibra ottica) e canali di tipo wireless (atmosfera o spazio libero). Nel caso del canale wireless il segnale elettrico viene convertito in una radiazione elettromagnetica ad alta frequenza da un'antenna trasmittente ed un'altra antenna, in ricezione, esegue l'operazione inversa. Le due antenne sono usualmente considerate parte integrante del canale. La maggior parte dei canali di trasmissione sono utilizzabili in preassegnati intervalli di frequenze. In Tabella 1.1 sono riportati alcuni esempi di canali che realizzano la propagazione guidata delle onde elettromagnetiche e le frequenze a cui essi sono comunemente utilizzati. Si noti che, per le fibre ottiche, si è soliti far riferimento alle lunghezze d'onda piuttosto che alle frequenze (nello spazio libero una lunghezza d'onda di 3 μm corrisponde ad una frequenza di 100 THz = 10^{14} Hz).

La propagazione delle onde elettromagnetiche nell'atmosfera può avvenire secondo (almeno) tre diversi meccanismi; si parla, a seconda dei casi, di propagazione per onda di superficie (ground-wave), per riflessione ionosferica (sky-wave) e di collegamento in visibilità (line of sight). La propagazione in banda MF (Medium Frequency, 0.3-3 MHz) avviene prevalentemente per onde di superficie dove la rifrazione tende a curvare l'onda che segue il profilo della superficie terrestre (oltre l'orizzonte visivo). Le medie frequenze sono adoperate

per la radiodiffusione di segnali DSB-SC AM (o semplicemente AM, ottenuti ricorrendo alla modulazione di ampiezza). La propagazione ionosferica sfrutta la rifrazione dell'onda stessa da parte della ionosfera, che è formata da diversi strati di particelle ionizzate posizionati ad altezze comprese tra i 70 e i 400 km sopra la superficie terrestre. La propagazione ionosferica, che rappresenta il modo dominante delle onde radio HF (High Frequency, 3-30 MHz), diviene trascurabile oltre i 30 MHz. Al di sopra dei 30 MHz i segnali radio non subiscono una significativa attenuazione nel passaggio attraverso la ionosfera e questo rende possibile realizzare, a partire dalle frequenze in banda VHF (Very High Frequency, 30-300 MHz) collegamenti tra trasmettitori/ricevitori terrestri e satellitari. A partire dalla banda VHF la forma di propagazione predominante delle onde elettromagnetiche è quella in visibilità (LOS). La propagazione in visibilità tra due stazioni terrestri è limitata dalla curvatura terrestre: se l'antenna trasmittente è montata ad un'altezza h in metri sulla superficie terrestre, la distanza dell'orizzonte radio, in assenza di ostruzioni fisiche, montagne, edifici, etc., è approssimativamente pari a $d = 4.122 \sqrt{h}$ km.

Canale rumoroso non distorto

Ogni canale introduce una certa attenuazione per cui la potenza del segnale si riduce al crescere della distanza dal trasmettitore. Nel seguito descriveremo il canale trasmissivo attraverso un sistema. Se si denotano con $u(t)$ e $v(t)$ i segnali, rispettivamente, in ingresso e in uscita al sistema e con P_T e P_R le corrispondenti potenze, espresse tipicamente in Watt, è possibile definire l'attenuazione o perdita in trasmissione in potenza subita dal segnale nella trasmissione dalla sorgente alla destinazione (e in ipotesi di adattamento per il massimo trasferimento di potenza dall'ingresso all'uscita) come

$$L = \frac{P_T}{P_R} \quad (1.4)$$

L'attenuazione è il reciproco del guadagno G ($G = 1/L$) ed entrambe le quantità sono usualmente espresse in decibel (dB); per esempio, il valore dell'attenuazione in unità logaritmiche è

$$L_{dB} = 10 \log_{10} \frac{P_T}{P_R} \quad (1.5)$$

A seconda del tipo di canale l'attenuazione in dB può dipendere dalla distanza tra la sorgente e la destinazione con legge lineare (come accade nei mezzi che realizzano la propagazione guidata come, ad esempio, le linee bifilari, il cavo coassiale, la guida d'onda, la fibra ottica) oppure quadratica, come accade per la perdita dovuta alla propagazione nello spazio libero (quando non siano presenti perdite aggiuntive di tipo dissipativo). L'attenuazione chilometrica si può esprimere mediante la relazione

$$L(dB) = \alpha d^\beta \quad (1.6)$$

dove α è l'attenuazione introdotta e β un coefficiente compreso tra 1 e 3. In Tabella 1.2 sono riportati i valori indicativi di α in dB per km per alcuni mezzi trasmissivi. Dalla tabella si

Mezzo trasmissivo	Frequenza	attenuazione (dB/km)
Linea bifilare (\varnothing 0.3 cm)	1 kHz	0.05
Linea bifilare intrecciata	10 kHz	2
Linea bifilare intrecciata	100 kHz	3
Linea bifilare intrecciata	300 kHz	6
Cavo coassiale (\varnothing 1 cm)	100 kHz	1
Cavo coassiale (\varnothing 1 cm)	1 MHz	2
Cavo coassiale (\varnothing 1 cm)	3 MHz	4
Cavo coassiale (\varnothing 15 cm)	100 MHz	1.5
Guida d'onda	10 GHz	5
Fibra ottica	$0.82 \mu\text{m}$	3.5

Si evince che ciascuno dei mezzi trasmissivi è utilizzabile solo in un certo intervallo di frequenze; ad esempio, il cavo coassiale è utilizzabile dai 100 kHz a qualche GHz. Al di fuori di questo intervallo di frequenze è più conveniente ricorrere ad altri mezzi trasmissivi.

Il ricevitore opera sul segnale in uscita al canale amplificandolo, in modo da compensare l'attenuazione introdotta dal canale, e filtrandolo (l'opportunità di filtrare il segnale in uscita al canale sarà chiara nel seguito). In aggiunta all'attenuazione diversi effetti indesiderati si manifestano nella trasmissione del segnale di informazione. Il segnale viene, infatti, distorto e corrotto dalla presenza di interferenza e rumore che producono nell'insieme un'alterazione della forma del segnale. Nonostante tali alterazioni possano manifestarsi in ogni punto del sistema, è opportuno associarle al canale trasmissivo modellando il trasmettitore e il ricevitore come dispositivi ideali (il rumore generato nei primi stadi del ricevitore può essere infatti riportato in ingresso al ricevitore stesso, cioè in uscita al canale).

Il progetto e l'analisi di un sistema di comunicazione richiedono di rappresentare il canale fisico attraverso un modello matematico. In molti casi un attento progetto preliminare dei segnali in gioco consente di ipotizzare che il canale sia non distorcente; se, il rumore presente è solo quello termico esso può essere modellato come gaussiano bianco additivo e si parla di canale AWGN (dall'inglese Additive White Gaussian Noise). Il rumore termico viene prodotto dal moto di agitazione termica dei portatori di carica nei dispositivi resistivi o a stato solido relativi ai primi stadi del ricevitore che realizzano le funzioni di amplificazione e filtraggio del segnale in uscita dal canale. Da un punto di vista matematico il canale AWGN è descritto dal legame ingresso/uscita (Figura 1.3)

$$r(t) = Ax(t - \tau) + n(t); \quad (1.7)$$

dove $x(t)$ ed $r(t)$ denotano, rispettivamente, il segnale in ingresso e quello in uscita al canale fisico, A è un fattore di attenuazione, τ il ritardo di propagazione, ed $n(t)$ il rumore.

Il canale AWGN modella in maniera accurata molti canali di comunicazione di tipo wired, e alcuni tipi di canale wireless tra cui alcuni canali satellitari e il canale spaziale che si frappone nelle comunicazioni tra la Terra e i sistemi di comunicazione tra satelliti. L'uso del

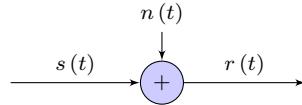


Figura 1.3: Canale AWGN

semplice modello AWGN, d'altra parte, è utile alla comprensione e allo sviluppo dei risultati fondamentali della trasmissione numerica.

Canale rumoroso distorcente

La distorsione è il risultato di una imperfetta risposta del sistema al segnale desiderato. La distorsione può risultare più o meno dannosa in relazione al tipo di applicazione che si considera. L'orecchio umano è sensibile alla distorsione di ampiezza, ma poco sensibile alla distorsione di fase; per questo motivo nei sistemi telefonici ci si deve preoccupare soprattutto degli effetti della (eventuale) distorsione di ampiezza. Infatti, la rapidità di variazione nel tempo del segnale in ingresso dipende dalla sua banda (per esempio, l'estensione dell'intervallo di frequenze che contiene il 98% dell'energia del segnale). D'altro canto la capacità del sistema di seguire le variazioni del segnale in ingresso (senza distorcerlo) dipende dalla sua larghezza di banda. I canali fisici hanno una banda finita che limita il tasso di variazione dell'eventuale segnale analogico o il numero di bit per unità di tempo del segnale numerico che, come vedremo, è direttamente legata alla larghezza di banda del segnale a tempo continuo ad esso associato. Ad esempio, la larghezza di banda dei mezzi trasmittivi che realizzano la propagazione guidata delle onde elettromagnetiche varia dalle centinaia di kHz della linea bifilare ai MHz dei cavi coassiali, per arrivare ai GHz delle fibre ottiche. In tal caso, il segnale ricevuto è una versione distorta del segnale trasmesso (figura 1.4). Si assume che il canale di trasmissione possa essere schematizzato mediante un filtro lineare tempo-invariante (LTI) di risposta impulsiva $c(t)$, ragion per cui il segnale ricevuto è espresso come

$$r(t) = s(t) * h(t) + n(t) = \int s(\tau) c(t - \tau) d\tau + n(t) \quad (1.8)$$

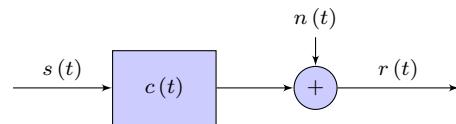


Figura 1.4: Canale AWGN distorcente

Un esempio di canale distorcente è quello in cui ad un utente viene assegnata una banda limitata di frequenze $[f_1, f_2]$ poiché le bande adiacenti sono già occupate da altri canali

di comunicazione. In tal caso, ammesso che non vi siano distorsioni di altra natura, $c(t)$ può considerarsi la risposta impulsiva di un filtro passa-banda ideale con frequenza di taglio inferiore f_1 e frequenza di taglio superiore f_2 . Si parla in tal caso anche di “modello di canale AWGN a banda limitata”.

Altri canali di trasmissione

Se il canale, oltre ad essere distorcente, è temporalmente variante, l’uscita del sistema sarà ancora una volta una versione distorta dell’ingresso ma la distorsione dipenderà dall’istante di tempo in cui il canale viene usato. Esempi di canali tempo varianti sono il canale radiomobile e il canale ionosferico, in cui il segnale perviene al ricevitore attraverso cammini multipli caratterizzati da differenti valori dell’attenuazione e del ritardo di propagazione (multipath). Poichè i differenti contributi ottenuti attraverso i cammini multipli arrivano al ricevitore con ritardi diversi e variabili nel tempo essi possono sommarsi in modo costruttivo o distruttivo dando luogo al fenomeno del fading, una forma di attenuazione tempo-variante del segnale ricevuto.

Un’altra forma di distorsione è legata alla presenza di altre sorgenti di segnale, come accade, ad esempio, nei casi di utilizzo di un canale da parte di più utenti. Il filtraggio del segnale in uscita dal canale e l’utilizzo di bande di guardia consente di sopprimere l’interferenza e i disturbi allocati in bande di frequenze diverse da quelle occupate dal segnale utile. Nei sistemi wireless possono essere presenti altre sorgenti di rumore quali il rumore cosmico, quello atmosferico e il rumore prodotto dai dispositivi di elettronici urbani (sistemi di iniezione delle automobili e altri dispositivi). Il rumore cosmico, quello atmosferico e quello di origine antropica sono ricevuti dall’antenna e si sommano al rumore termico.

Demodulatore

Il demodulatore elabora il segnale ricevuto $r(t)$ al fine di individuare quale tra le M forme d’onda è stata trasmessa sul canale. Anche nel caso di canale AWGN, l’aggiunta del rumore termico modifica in modo non prevedibile il segnale trasmesso ed è lecito aspettarsi che il demodulatore sia affetto da un certo numero di errori. Gli errori corrispondono a decisioni errate, in cui il demodulatore decide per un segnale $s_j(t)$, diverso da quello effettivamente trasmesso. Il problema di individuare il “demodulatore ottimo”, ove il criterio di ottimalità è la minimizzazione della probabilità di errore sarà affrontato nei prossimi capitoli.

Decodificatore di canale

A causa degli errori che inevitabilmente saranno introdotti nel processo di demodulazione, la sequenza di bit in ingresso al decodificatore di canale potrà differire da quella prodotta dal codificatore di canale. Compito del decodificatore di canale è quindi quello di utilizzare la ridondanza introdotta dal codificatore di canale al fine di individuare e correggere parte degli errori prodotti nella fase di decisione.

Decodificatore di sorgente

Il decodificatore di sorgente esegue l’operazione inversa a quella eseguita dal codificatore di sorgente. A partire dai dati compressi, reintroduce la ridondanza statistica al fine di restituire

ai dati il loro significato semantico e renderli quindi fruibili al destinatario. Si noti che la cascata dei blocchi modulatore, canale e demodulatore, come rappresentata in figura 1.5, è un sistema che accetta in ingresso simboli binari e produce in uscita simboli binari che, con una certa probabilità di errore p_e , possono essere diversi dai corrispondenti simboli in ingresso.

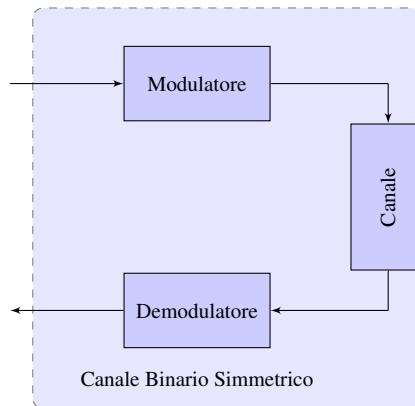


Figura 1.5: Modello esploso di un canale binario simmetrico

Capitolo 2

Elementi di Teoria della Probabilità

“Quando è stato eliminato tutto ciò che è impossibile, ciò che rimane, sebbene improbabile, deve essere la verità.”

Sherlock Holmes (The adventure of the blanched soldier)

Nello studio dei fenomeni fisici è di fondamentale importanza avere a disposizione un modello matematico che renda possibile la descrizione o la predizione di alcune caratteristiche di interesse. La relazione $v = gt$ fornisce, ad esempio, il modello matematico per descrivere la velocità di un corpo in caduta libera nel vuoto. Questo modello è di tipo deterministico, in quanto ripetendo più volte lo stesso esperimento si osserveranno, essenzialmente, gli stessi valori di velocità in determinati punti dello spazio. In condizioni non ideali, tuttavia, questa legge può fornire risultati completamente inadeguati: si pensi, ad esempio, al moto di una particella di polvere soggetta, oltre all’accelerazione di gravità, anche all’azione di disturbi ambientali difficilmente predicibili, quali gradienti di temperatura, pressione e umidità.

Lo scopo della teoria della probabilità è quello di fornire modelli matematici per trattare situazioni non predicibili in maniera deterministica. I primi studi sulla teoria della probabilità risalgono al 1650 quando il Cavalier de Mèrè chiese agli amici Blaise Pascal e Pierre de Fermat di sviluppare un modello matematico per descrivere alcune “ricorrenze del gioco d’azzardo”. Più avanti, verso gli inizi del 1900, basandosi sulle idee di Emile Borel circa la teoria della misura, André Kolmogorov elaborò un insieme di assiomi tramite i quali la teoria della probabilità poteva essere formalizzata mediante la teoria della misura.

 Italian writers of the fifteenth and sixteenth centuries, had discussed the problem of the division of a stake between two players whose game was interrupted before its close. The problem was proposed to Pascal and Fermat, probably in 1654, by the Chevalier de Mèrè, a gambler who is said to have had unusual ability “even for the mathematics”. The correspondence which ensued between Fermat and Pascal, was fundamental in the development of modern concepts of probability.

Gamblers were used to bet on the event of getting at least one 1 (ace) in four rolls of a dice. As a more trying variation, two die were rolled 24 times with a bet on having at least one double ace. According to the reasoning of Chevalier de Mèrè, two aces in two rolls are 1/6 as likely as 1 ace in one roll. (Which is correct.) To compensate, de Mèrè thought, the two die should be rolled 6 times. And to achieve the probability of 1 ace in four rolls, the number of the rolls should be increased four fold - to 24. Thus reasoned Chevalier de Mèrè who expected a couple of aces to turn up in 24 double rolls with the frequency of an ace in 4 single rolls. However, he lost consistently.

*[*Oeuvres de Fermat* (ed. Tannery and Henry, Vol. II, pp. 288–314, Paris 1894)]*

2.1 ESPERIMENTO ALEATORIO

In questo paragrafo viene illustrata la definizione formale di un esperimento aleatorio. A tal fine sarà necessario introdurre tre elementi: l'insieme dei possibili risultati dell'esperimento considerato; gli eventi che si possono costruire a partire dai risultati dell'esperimento e una misura di probabilità da applicare sugli eventi.

2.1.1 Spazio dei campioni

Ogni volta che si effettua un esperimento si osserva la risposta dell'ambiente ad una data sollecitazione e se ne ricava un risultato sperimentale. La singola esecuzione di un esperimento si chiama *prova* e ad ogni prova corrisponde un *risultato* ω . Per un dato esperimento l'insieme Ω di tutti i possibili risultati si chiama spazio dei campioni o spazio delle prove e può essere finito, infinito numerabile (cioè indicizzabile mediante l'insieme degli interi positivi), o infinito non numerabile. Se lo spazio dei campioni è finito o numerabile lo diremo anche discreto.

Esempio: sorgente binaria

Si consideri una sorgente che emetta simboli binari. Se l'esperimento consiste nell'emissione da parte della sorgente di un solo simbolo, lo spazio dei campioni è l'insieme $\Omega = \{0, 1\}$. Se invece si osserva l'emissione di un pacchetto di N simboli, il generico risultato è un' N -pla ordinata di cifre binarie del tipo

$$\omega = (b_1, b_2, \dots, b_N), \quad b_i \in \{0, 1\}, \quad i = 1, 2, \dots, N$$

ed il corrispondente spazio dei campioni è: $\Omega = \{0, 1\}^N$, cioè l'insieme delle N -ple binarie. Anche in questo caso lo spazio delle prove è finito, precisamente ha cardinalità 2^N , e può essere rappresentato in forma tabellare come illustrato in tabella 2.1 per $N = 3$. Se, infine, si osserva l'emissione della sorgente fino a quando non si ottiene un 1, il risultato è del tipo $\omega = (0, 0, \dots, 0, 1)$. Il corrispondente spazio dei campioni è

$$\Omega = \{(1), (0, 1), (0, 0, 1), \dots, \}$$

Tabella 2.1
Spazio dei campioni per pacchetti di lunghezza 3.

ω	(b_1, b_2, b_3)
ω_0	$(0, 0, 0)$
ω_1	$(0, 0, 1)$
ω_3	$(0, 1, 1)$
ω_4	$(1, 0, 0)$
ω_5	$(1, 0, 1)$
ω_6	$(1, 1, 0)$
ω_7	$(1, 1, 1)$

ed è infinito numerabile.

Esempio: lancio di dadi

Si consideri il lancio di un dado: per tale esperimento lo spazio dei campioni è:

$$\Omega = \{f_1, f_2, f_3, f_4, f_5, f_6\}$$

cioè l'insieme delle 6 facce del dado. Se poi si lancia il dado due volte o, equivalentemente, si lancia una coppia di dadi lo spazio dei campioni è:

$$\Omega = \{f_1, f_2, f_3, f_4, f_5, f_6\}^2$$

ed è illustrato in Tabella 2.2.

Tabella 2.2
Spazio dei campioni relativo al lancio di una coppia di dadi

(f_1, f_1)	(f_1, f_2)	(f_1, f_3)	(f_1, f_4)	(f_1, f_5)	(f_1, f_6)
(f_2, f_1)	(f_2, f_2)	(f_2, f_3)	(f_2, f_4)	(f_2, f_5)	(f_2, f_6)
(f_3, f_1)	(f_3, f_2)	(f_3, f_3)	(f_3, f_4)	(f_3, f_5)	(f_3, f_6)
(f_4, f_1)	(f_4, f_2)	(f_4, f_3)	(f_4, f_4)	(f_4, f_5)	(f_4, f_6)
(f_5, f_1)	(f_5, f_2)	(f_5, f_3)	(f_5, f_4)	(f_5, f_5)	(f_5, f_6)
(f_6, f_1)	(f_6, f_2)	(f_6, f_3)	(f_6, f_4)	(f_6, f_5)	(f_6, f_6)

Esempio: coppie di arrivi in $[0, T]$

Si consideri l'arrivo di un viaggiatore e di un treno in una certa stazione nell'intervallo $[0, T]$. Il generico risultato è la coppia

$$\omega = (t_V, t_T)$$

degli istanti di arrivo del viaggiatore e del treno ed il corrispondente spazio dei campioni è

$$\Omega = [0, T]^2$$

Pertanto, in questo caso, lo spazio delle prove è il quadrato di figura 2.1.1 ed è infinito non numerabile, precisamente è continuo.

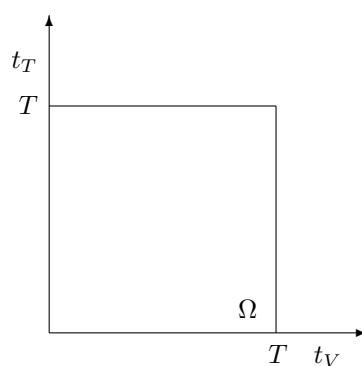


Figura 2.1: Spazio delle prove relativo ad una coppia di arrivi.

2.1.2 Eventi

Considerato un esperimento, siamo interessati a sapere se i risultati soddisfino o meno determinate condizioni relative ai possibili risultati dell'esperimento. Tali condizioni definiscono un sottoinsieme E dello spazio dei campioni Ω . L'evento E si verifica in una generica prova dell'esperimento se il particolare risultato ω di tale prova appartiene ad E , cioè se $\omega \in E$.

Si noti che lo spazio dei campioni Ω è esso stesso un evento così come si include tra gli eventi anche l'insieme vuoto \emptyset . L'evento Ω si verifica in ogni prova, per cui viene detto evento certo, mentre l'evento \emptyset è l'evento impossibile in quanto non si verifica mai. Infine gli eventi del tipo $\{\omega\}$, costituiti cioè da un unico risultato, sono detti eventi elementari. Gli eventi sono sottoinsiemi dello spazio dei campioni e le operazioni sugli eventi corrispondono pertanto alle usuali operazioni tra sottoinsiemi. Così ad esempio, la disgiunzione di due eventi A e B , cioè l'evento A o B , è l'evento che si verifica quando si verifica almeno uno tra i due eventi, eventualmente entrambi: essa coincide quindi con l'unione dei sottoinsiemi A e B , cioè col sottoinsieme costituito dai risultati che appartengono ad A oppure a B , eventualmente ad entrambi, e sarà pertanto denotata con $A \cup B$, oppure $A + B$, (figura 2.2). Analogamente la congiunzione di due eventi A e B è l'evento che si verifica quando si verificano sia A che B ; coincide quindi con l'intersezione dei sottoinsiemi A e B , cioè col sottoinsieme costituito dai risultati che appartengono sia ad A che a B e la si denota pertanto con $A \cap B$, oppure AB , (figura 2.2). Due eventi A e B la cui intersezione è non vuota ($A \cap B \neq \emptyset$) si dicono

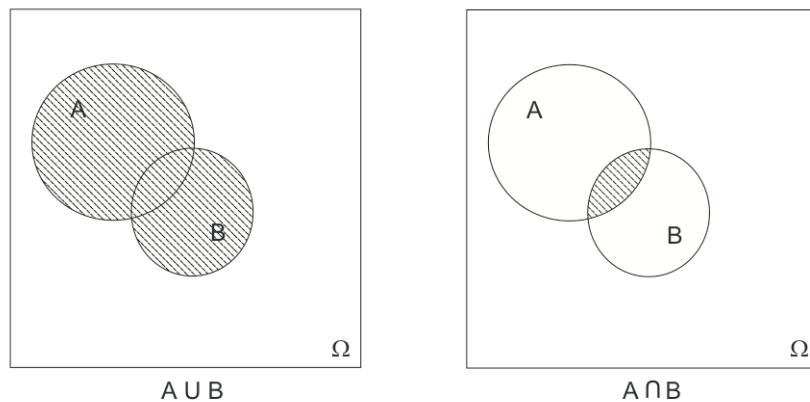


Figura 2.2: Disgiunzione e congiunzione di eventi

compatibili in quanto, in una prova dell'esperimento in esame, possono verificarsi entrambi. Viceversa, due eventi A e B che non possono mai verificarsi contemporaneamente, cioè per i quali $A \cap B = \emptyset$, si dicono mutuamente esclusivi o incompatibili (figura 2.3). La negazione

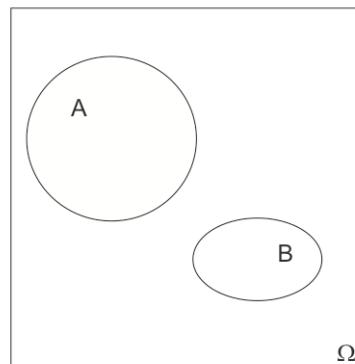


Figura 2.3: Eventi mutuamente esclusivi.

di un evento A è l'evento che si verifica ogni volta che non si verifica A ; essa è pertanto il complemento \bar{A} del sottoinsieme A , cioè il sottoinsieme costituito da tutti i risultati che non appartengono ad A (figura 2.4). Ovviamente un evento A ed il suo negato \bar{A} sono sempre mutuamente esclusivi. Infine la differenza $A - B$ tra due eventi è l'evento che si verifica ogni volta che si verifica A , ma non B ; essa è pertanto il sottoinsieme costituito dai risultati che

appartengono ad A e non appartengono a B (figura 2.4); in altri termini si ha:

$$A - B = A \cap \bar{B}$$

Per comodità del lettore nella tabella 2.3 sono riportate le più comuni proprietà di unione,

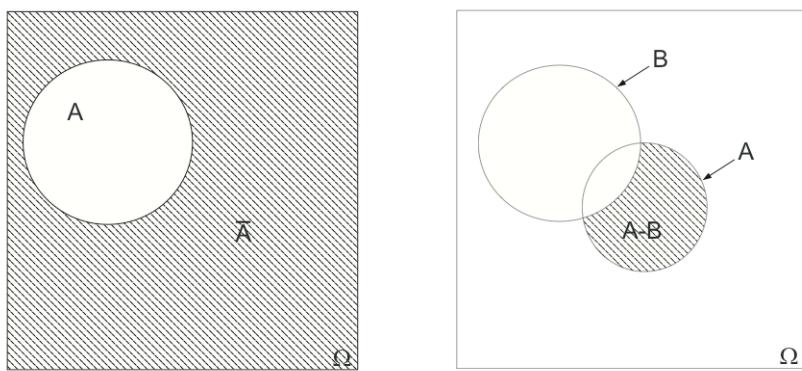


Figura 2.4: Negazione e Differenza di eventi.

intersezione e complementazione.

Esempio: sorgente binaria (continuazione)

Con riferimento all'emissione di un pacchetto di lunghezza 3, si riprendano in esame gli eventi

$$\begin{aligned} A_1 &\equiv \text{"il numero di uno nel pacchetto è pari o nullo"} \\ &\equiv \{(0, 0, 0), (0, 1, 1), (1, 0, 1), (1, 1, 0)\} \end{aligned}$$

$$\begin{aligned} A_2 &\equiv \text{"il numero di uno è minore di due"} \\ &\equiv \{(0, 0, 0), (0, 0, 1), (0, 1, 0), (1, 0, 0)\} \end{aligned}$$

$$\begin{aligned} A_3 &\equiv \text{"il numero di uno è uguale a due"} \\ &\equiv \{(0, 1, 1), (1, 0, 1), (1, 1, 0)\} \end{aligned}$$

La disgiunzione $A_1 \cup A_2$ degli eventi A_1 ed A_2 è l'evento

$$\begin{aligned} A_1 \cup A_2 &\equiv \text{"il numero di uno nel pacchetto è minore od uguale a due"} \\ &= \{(0, 0, 0), (0, 0, 1), (0, 1, 0), (1, 0, 0), (0, 1, 1), (1, 0, 1), (1, 1, 0)\} \end{aligned}$$

mentre la loro congiunzione $A_1 \cap A_2$ è l'evento

$$\begin{aligned} A_1 \cap A_2 &= \equiv \text{"il numero di uno nel pacchetto è nullo"} \\ &= \{(0, 0, 0)\} \end{aligned}$$

Tabella 2.3
Principali proprietà di unione, intersezione e complementazione

<i>Idempotenza</i>	$A \cup A = A$ $A \cap A = A$
<i>Associatività</i>	$A \cap (B \cap C) = (A \cap B) \cap C = A \cap B \cap C$ $A \cup (B \cup C) = (A \cup B) \cup C = A \cup B \cup C$
<i>Commutatività</i>	$A \cap B = B \cap A$ $A \cup B = B \cup A$
<i>Distributività</i>	$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$ $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$
<i>Leggi di De Morgan</i>	$\overline{A \cup B} = \overline{A} \cap \overline{B}$ $\overline{A \cap B} = \overline{A} \cup \overline{B}$
<i>Complementazione</i>	$\overline{\overline{A}} = A$ $A \cap \overline{A} = \emptyset$ $A \cup \overline{A} = \Omega$
Ω e \emptyset	$A \cup \Omega = \Omega$ $A \cap \Omega = A$ $A \cup \emptyset = A$ $A \cap \emptyset = \emptyset$

Quindi A_1 ed A_2 sono compatibili ed in ogni prova si verificano o meno entrambi a seconda che il risultato sia o no $(0, 0, 0)$. Analogamente la congiunzione di A_1 ed A_3 è l'evento

$$\begin{aligned} A_1 \cap A_3 &\equiv \text{"il numero di uno nel pacchetto è due"} \\ &= \{(0, 1, 1), (1, 0, 1), (1, 1, 0)\} \end{aligned}$$

Pertanto anche A_1 ed A_3 sono compatibili: in particolare se il risultato della prova è $(0, 1, 1)$ (oppure $(1, 0, 1)$ o $(1, 1, 0)$) si verificano entrambi, mentre se il risultato è $(0, 0, 0)$ A_1 si verifica e non A_3 ; infine se il risultato è $(1, 1, 1)$ (oppure $(0, 0, 1)$, $(0, 1, 0)$ o $(1, 0, 0)$) non si verifica nessuno dei due eventi. Viceversa, avendosi $A_2 \cap A_3 = \emptyset$, A_2 e A_3 sono mutuamente esclusivi. La differenza $A_1 - A_2$ è l'evento

$$\begin{aligned} A_1 - A_2 &\equiv \text{"il numero di uno nel pacchetto è pari o nullo e non minore di due"} \\ &= \{(0, 1, 1), (1, 0, 1), (1, 1, 0)\} \end{aligned}$$

Infine l'evento

$$\begin{aligned} \overline{A_1} &\equiv \text{"il numero di uno nel pacchetto è dispari"} \\ &= \{(0, 0, 1), (0, 1, 0), (1, 0, 0), (1, 1, 1)\} \end{aligned}$$

è il complemento di A_1 .

Esempio: coppie di arrivi in $[0, T]$ (continuazione)
Sono eventi (fig. 2.5):

$$\begin{aligned} E_1 &\equiv \text{"il viaggiatore ed il treno arrivano contemporaneamente"} \\ &\equiv \{(t_V, t_T) \in \Omega : t_V = t_T\} \end{aligned}$$

$$\begin{aligned} E_2 &\equiv \text{"il viaggiatore arriva prima del treno"} \\ &\equiv \{(t_V, t_T) \in \Omega : t_V < t_T\} \end{aligned}$$

$$\begin{aligned} E_3 &\equiv \text{"il viaggiatore arriva tra } 0.2T \text{ e } 0.4T\text{"} \\ &\equiv \{(t_V, t_T) \in \Omega : 0.2T \leq t_V \leq 0.4T\} \end{aligned}$$

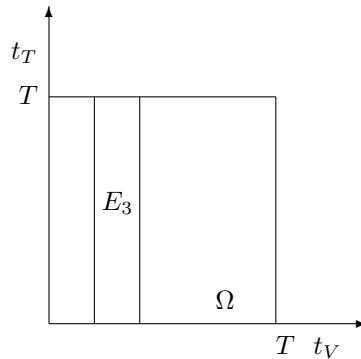
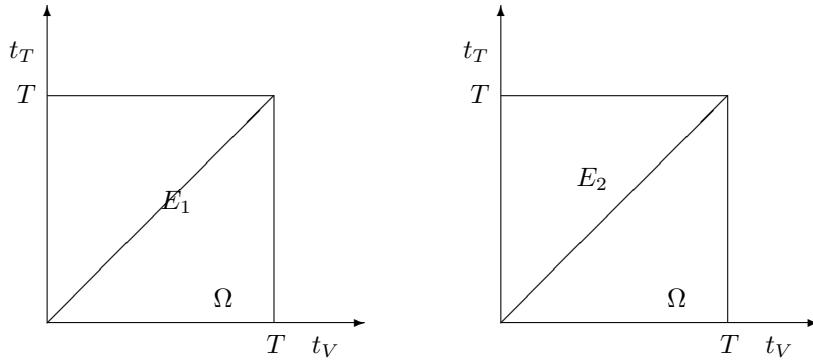


Figura 2.5: Alcuni eventi relativi alle coppie di arrivi in $[0, T]$.

Quando è stata introdotta la nozione di evento, volutamente, non si è discusso se tutti i sottoinsiemi dello spazio delle prove siano eventi. Invero ciò non è sempre conveniente o possibile. Infatti non sempre è conveniente considerare eventi tutti i possibili sottoinsiemi dello spazio delle prove; ad esempio, con riferimento alla sorgente binaria, se siamo interessati non alla struttura del pacchetto di bit ma solo alla sua parità, a sapere cioè se il numero di bit costituenti il pacchetto è o meno pari, è conveniente limitarsi a considerare solo tale evento

(più in generale solo gli eventi di interesse) e quelli che si ottengono operando con le usuali operazioni insiemistiche a partire da tale evento. È quindi necessario garantirsi che operando su tali sottoinsiemi si ottenga ancora un evento: ciò porta ad imporre alcune condizioni sull'insieme degli eventi \mathcal{E} che si sintetizzano dicendo che \mathcal{E} deve essere un'algebra. Formalmente una famiglia non vuota \mathcal{E} di sottoinsiemi di Ω è un'algebra di eventi se

- A1:** $A \in \mathcal{E} \implies \overline{A} \in \mathcal{E}$
A2: $A, B \in \mathcal{E} \implies A \cup B \in \mathcal{E}$

Dalle condizioni **A1** ed **A2** segue che sono soddisfatte anche le seguenti proprietà:

- P1:** $\Omega \in \mathcal{E}; \quad (\Omega = A \cup \overline{A})$
P2: $\emptyset \in \mathcal{E}; \quad (\emptyset = \overline{\Omega})$
P3: $A, B \in \mathcal{E} \implies A \cap B \in \mathcal{E} \quad (A \cap B = \overline{\overline{A} \cup \overline{B}})$
P4: $A, B \in \mathcal{E} \implies A - B \in \mathcal{E} \quad (A - B = A \cap \overline{B})$

Un'algebra di eventi è pertanto chiusa rispetto alle operazioni di complementazione, unione ed intersezione eseguite su un numero finito di eventi. Talvolta è però necessario operare dei ragionamenti al limite per cui occorre prendere in esame successioni di eventi: in tal caso si richiede che comunque si operi su una successione di eventi si abbia ancora un evento. Ciò porta ad imporre che \mathcal{E} non solo sia un'algebra, ma una σ -algebra. In altri termini una σ -algebra è chiusa rispetto alle operazioni di complementazione, unione ed intersezione eseguite su un numero finito o un'infinità numerabile di eventi.

2.1.3 Gli assiomi della probabilità

Ad ogni evento A occorre associare un numero $P(A)$ che ne misura la probabilità; a tal fine consideriamo la funzione

$$P : A \in \mathcal{E} \longrightarrow P(A) \in [0, 1]$$

ove \mathcal{E} è l'insieme degli eventi. Tale corrispondenza non può essere arbitraria, ma deve soddisfare le seguenti condizioni:

Assiomi della probabilità

A1. Non negatività

$$P(A) \geq 0 \quad \forall A \in \mathcal{E}$$

A2. Normalizzazione

$$P(\Omega) = 1$$

A3a. Finita additività

$$P(A \cup B) = P(A) + P(B) \quad \forall A, B : A \cap B = \emptyset$$

A3b: Numerabile additività

$$P(\bigcup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} P(A_n) \quad \forall n, m \in \mathbb{N} : n \neq m \text{ e } A_n \cap A_m = \emptyset$$

Gli assiomi della probabilità sono anche noti come assiomi di Kolmogorov. Essi affermano che la probabilità di un qualsiasi evento è non negativa (**A1**), che l'evento certo ha probabilità uno (**A2**) ed inoltre la probabilità di due eventi mutuamente esclusivi è la somma delle probabilità dei singoli eventi (**A3a**); tale proprietà vale anche per un'infinità numerabile di eventi a due a due mutuamente esclusivi (**A3b**). Dagli assiomi di Kolmogorov segue che la probabilità gode anche delle seguenti proprietà:

P1: $P(\emptyset) = 0$

P2: $P(\bar{A}) = 1 - P(A) \quad \forall A \in \mathcal{E}$

P3: $P(A - B) = P(A) - P(A \cap B)$

P4: $P(A \cup B) = P(A) + P(B) - P(A \cap B) \leq P(A) + P(B) \quad (\text{subadditività})$

P5: $A \subseteq B \implies P(A) \leq P(B) \quad (\text{monotonicità})$

ove le varie relazioni valgono qualunque siano gli eventi A o B . In altri termini la probabilità dell'evento impossibile è nulla; la probabilità di una negazione è semplicemente il complemento ad uno della probabilità dell'affermazione; la probabilità di una differenza è la differenza tra la probabilità del primo evento e quella della congiunzione; la probabilità di una disgiunzione è, per eventi non necessariamente mutuamente esclusivi, pari alla somma delle probabilità degli eventi costituenti meno quella della loro congiunzione; infine la probabilità è crescente rispetto alla relazione d'inclusione. Per comodità del lettore gli assiomi e le

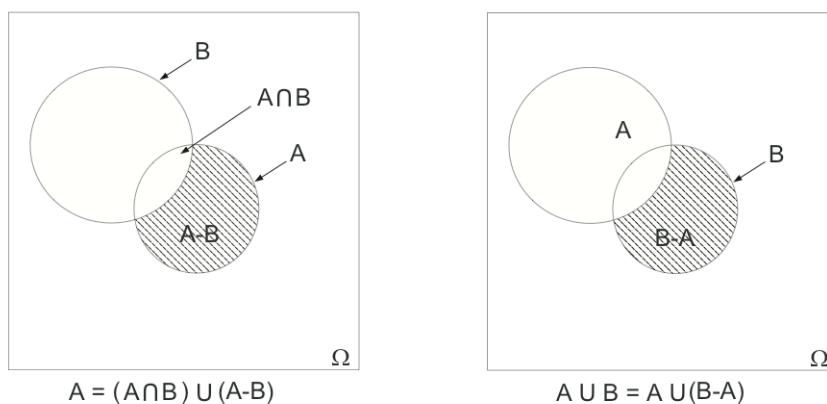
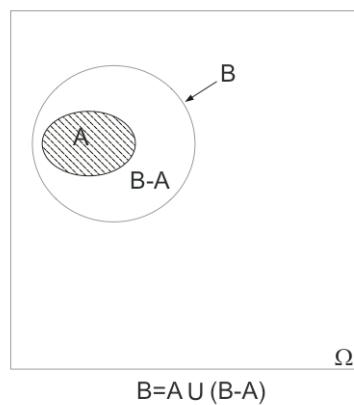


Figura 2.6: Diagrammi di Venn per $A = (A \cap B) \cup (B - A)$ $A \cup B = A \cup (B - A)$

proprietà della probabilità sono riassunte nella tabella 2.4.

In conclusione, si ha la seguente definizione.

Figura 2.7: Diagramma di Venn di $B = A \cup (B - A)$

Esperimento aleatorio

Un esperimento aleatorio è caratterizzabile mediante la terna (Ω, \mathcal{E}, P) dove:

1. Ω è l'insieme di tutti i possibili risultati sperimentali (spazio dei campioni).
2. \mathcal{E} è la σ -algebra degli eventi.
3. $P(\cdot)$ è la legge di probabilità definita attraverso gli assiomi di Kolmogorov.

Esempio: Lancio di una coppia di dadi

Si riprenda in esame il lancio di una coppia di dadi: ricordiamo che lo spazio dei campioni è l'insieme

$$\Omega = \{f_1, f_2, f_3, f_4, f_5, f_6\}^2$$

illustrato in tabella 2.2. Come famiglia di eventi consideriamo l'insieme delle parti¹ di Ω , cioè

$$\mathcal{E} = \mathcal{P}(\Omega) .$$

In questo caso lo spazio dei campioni, e quindi anche l'insieme degli eventi, è finito. Conseguentemente è sufficiente assegnare le probabilità agli eventi elementari, per definire la probabilità di un qualunque altro evento. Infatti ogni evento può sempre essere riguardato come l'unione degli eventi elementari che lo costituiscono e, quindi, essendo gli eventi elementari mutuamente esclusivi e dovendo essere la probabilità additiva, si ha:

$$P(A) = \sum_{\omega \in A} P(\{\omega\}) . \quad (2.1)$$

1 L'insieme delle parti di un insieme A è l'insieme costituito da tutti i sottoinsiemi di A e si denota con $\mathcal{P}(A)$. Se A ha cardinalità n , allora l'insieme delle parti ha cardinalità 2^n .

Tabella 2.4
Assiomi e proprietà della probabilità.

<i>A1</i>	Non negatività	$P(A) \geq 0 \quad \forall A \in \mathcal{E}$
<i>A2</i>	Normalizzazione	$P(\Omega) = 1$
<i>A3a</i>	Finita additività	$P(A \cup B) = P(A) + P(B) \quad \forall A, B : A \cap B = \emptyset$
<i>A3b</i>	Numerabile additività	$P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n)$ $\forall n, m \in \mathbb{N}, n \neq m : A_n \cap A_m = \emptyset$
<i>P1</i>	Evento impossibile	$P(\emptyset) = 0$
<i>P2</i>	Complemento	$P(\bar{A}) = 1 - P(A) \quad \forall A \in \mathcal{E}$
<i>P3</i>	Differenza	$P(A - B) = P(A) - P(A \cap B) \quad \forall A \in \mathcal{E}$
<i>P4a</i>	Unione	$P(A \cup B) = P(A) + P(B) - P(A \cap B)$
<i>P4b</i>	Subadditività	$P(A \cup B) \leq P(A) + P(B)$
<i>P5</i>	Monotonicità	$A \subseteq B \implies P(A) \leq P(B)$

Si noti che la relazione (2.1) vale se il numero di eventi elementari costituenti A è finito in virtù della finita additività (**A3**), o se è numerabile, in virtù della numerabile additività (**A3b**). Nel caso in esame non vi è alcuna ragione per cui alcuni risultati dovrebbero verificarsi preferenzialmente rispetto agli altri, a meno che i dadi non siano truccati; conseguentemente gli eventi elementari possono essere assunti equiprobabili (*principio di ragion insufficiente*), cioè si pone:

$$\{(f_i, f_j)\} \subseteq \Omega \implies P(\{(f_i, f_j)\}) = \frac{1}{36}.$$

Conseguentemente in virtù della (2.1), per un qualsiasi evento si ha

$$P(E) = \frac{|E|}{|\Omega|} \tag{2.2}$$

ove $|\cdot|$ denota la cardinalità di un insieme. La quantità così definita, come è immediato verificare, è una probabilità nel senso che soddisfa gli assiomi. Infatti essa è non negativa (**A1**); inoltre si ha (**A2**):

$$P(\Omega) = \frac{|\Omega|}{|\Omega|} = 1$$

Infine, avendosi

$$A \cap B = \emptyset \implies |A \cup B| = |A| + |B|$$

vale anche la finita additività (**A3**).

Definita la legge di probabilità calcoliamo la probabilità dei seguenti eventi:

$$A \equiv \{"facce uguali"\}$$

$$B \equiv \{"la somma è compresa fra 7 e 10, estremi inclusi"\}$$

$$C \equiv \{"la somma è 2, 7 o 8"\}$$

Tali eventi sono i seguenti sottoinsiemi di Ω :

$$\begin{aligned} A &= \{(f_1, f_1), (f_2, f_2), (f_3, f_3), (f_4, f_4), (f_5, f_5), (f_6, f_6)\} \\ B &= \{(f_1, f_6), (f_2, f_5), (f_3, f_4), (f_4, f_3), (f_5, f_2), (f_6, f_1), \\ &\quad (f_2, f_6), (f_3, f_5), (f_4, f_4), (f_5, f_3), (f_6, f_2), (f_3, f_6), \\ &\quad (f_4, f_5), (f_5, f_4), (f_6, f_3), (f_4, f_6), (f_5, f_5), (f_6, f_4)\} \\ C &= \{(f_1, f_1), (f_1, f_6), (f_2, f_5), (f_3, f_4), (f_4, f_3), (f_5, f_2), \\ &\quad (f_6, f_1), (f_2, f_6), (f_3, f_5), (f_4, f_4), (f_5, f_3), (f_6, f_2)\} \end{aligned}$$

le cui cardinalità sono

$$|A| = 6 \quad |B| = 18 \quad |C| = 12.$$

Conseguentemente le loro probabilità valgono:

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

Esempio: Sorgente binaria senza memoria

Nel caso in cui la sorgente binaria emetta un solo simbolo, lo spazio dei campioni è:

$$\Omega = \{0, 1\}$$

mentre gli eventi sono gli elementi di:

$$\mathcal{E} = \mathcal{P}(\Omega) = \{\emptyset, \{0\}, \{1\}, \Omega\}$$

Anche in questo caso l'insieme dei possibili risultati e quello degli eventi sono finiti. Conseguentemente, sulla scorta di quanto detto nel paragrafo precedente, per assegnare una legge di probabilità è sufficiente definire la probabilità dell'evento $\{1\}$. Posto infatti $P(\{1\}) = p$, con $0 \leq p \leq 1$, necessariamente deve aversi $P(\{0\}) = q$, ove $q = 1 - p$. Conseguentemente resta definita la probabilità di tutti gli eventi elementari e, quindi, la probabilità di un qualsiasi altro evento.

Esempio: Coppia di arrivi in $(0, T]$

Si riprenda in esame l'esperimento dell'esempio 3, relativo all'arrivo *a caso* di un viaggiatore e di un treno in una certa stazione nell'intervallo $[0, T]$, il cui spazio delle prove è il quadrato:

$$\Omega = [0, T]^2$$

di figura 2.1.1. Tale spazio è infinito non numerabile, precisamente è continuo. Nel caso di spazi continui, come famiglia di eventi, si considera la più piccola σ -algebra contenente gli

intervalli (sotto-rettangoli del quadrato nel caso in esame): essa è costituita dai sottoinsiemi misurabili, per i quali cioè è definita l'area, del quadrato $[0, T]^2$. Dai dati del problema, cioè *arrivi a caso*, è ragionevole ipotizzare che la probabilità di avere arrivi in un certo intervallo non dipenda dagli estremi dell'intervallo, ma solo dalla sua durata, e quindi, più in generale, che la probabilità di un evento sia proporzionale alla sua area. In altri termini, nel caso in esame, una possibile legge di probabilità è:

$$P : E \subseteq \Omega \longrightarrow \frac{\mu_2(E)}{\mu_2(\Omega)} \quad (2.3)$$

ove $\mu_2(E)$ denota l'area di E , ed è illustrata in fig. 2.8, . Utilizzando la legge di probabilità

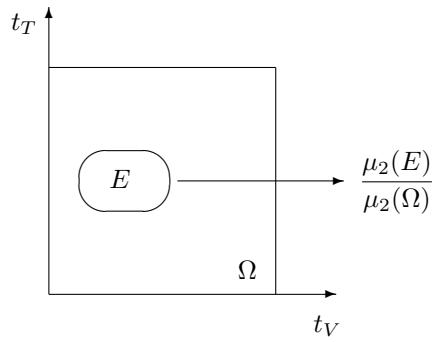


Figura 2.8: Legge di probabilità per lo spazio delle prove dell'esempio 3.

(2.3), con l'ausilio della fig. 2.5, si ha:

$$\begin{aligned} P(E_1) &= P(\{\text{"viaggiatore e treno arrivano contemporaneamente"}\}) = 0 \\ P(E_2) &= P(\{\text{"il viaggiatore arriva prima del treno"}\}) = 0.5 \\ P(E_3) &= P(\{\text{"il viaggiatore arriva tra } 0.2T \text{ e } 0.4T\}) = 0.2 \end{aligned}$$

Si noti che l'evento

$$E_1 \equiv \{\text{"viaggiatore e treno arrivano contemporaneamente"}\}$$

ha probabilità nulla pur non essendo l'evento impossibile, analogamente l'evento complementare

$$E_4 \equiv \{\text{"viaggiatore e treno non arrivano contemporaneamente"}\}$$

ha probabilità uno pur non essendo l'evento certo. Si osservi infine che la corrispondenza definita dalla (2.3) è una possibile legge di probabilità ogni volta che lo spazio delle prove è un sottoinsieme di \mathbb{R}^n di misura finita. Qualora si utilizzi la probabilità definita dalla (2.3), si dice che la probabilità è stata assegnata a caso; tale assegnazione è l'equivalente, per spazi delle prove continui, dell'equiprobabilità degli eventi elementari considerata nel caso discreto.

2.1.4 Definizioni alternative di probabilità

Storicamente, l'approccio assiomatico non è l'unico utilizzato per definire la probabilità: nel presente paragrafo si considerano brevemente il cosiddetto approccio classico e quello frequentistico.

Definizione classica

Si riprenda in esame la relazione (2.2) che definisce la probabilità nel caso del lancio di una coppia di dadi. Tale relazione la si interpreta dicendo che la probabilità di un evento è il rapporto tra il numero di casi favorevoli al verificarsi dell'evento E (numero degli eventi elementari costituenti E) ed il numero dei casi possibili (numero degli eventi elementari costituenti Ω). La legge di probabilità definita dalla (2.2) è nota come definizione *classica* di probabilità. Essa però è applicabile solo ad esperimenti il cui spazio dei campioni sia finito e, implicitamente, ipotizza che gli eventi elementari siano equiprobabili. Tale definizione è pertanto circolare nel senso che nel dare la definizione utilizza il concetto che si vuole definire. Tuttavia, con riferimento a spazi delle prove finiti, l'equiprobabilità degli eventi elementari è comunemente assunta quando, dai dati del problema, non risulti alcuna informazione che porti a considerare un evento elementare più o meno probabile di un altro. In altri termini la perfetta simmetria tra gli eventi elementari giustifica l'assunzione della loro equiprobabilità.

Definizione frequentistica

In alternativa all'approccio assiomatico si può definire la probabilità sulla scorta della frequenza relativa di un evento. A tal fine si definisce frequenza relativa di un evento E in N prove il rapporto:

$$F_E(N) = \frac{N(E)}{N}$$

tra il numero di volte $N(E)$ che si verifica l'evento E ed il numero complessivo N delle prove. La probabilità di E è poi definita come il limite della frequenza relativa in N prove, *indipendenti* ed effettuate tutte in *identiche condizioni*, al divergere di N (**Legge empirica del caso**): in altri termini, si pone:

$$P(E) = \lim_{N \rightarrow \infty} \frac{N(E)}{N} \quad (2.4)$$

Tale definizione però postula l'esistenza del limite ed è circolare in quanto il concetto di indipendenza è esso stesso un concetto probabilistico (vedi sezioni successive) così come l'effettuare le prove in identiche condizioni. Nell'ambito dell'approccio assiomatico, come si vedrà in un prossimo capitolo, è possibile però dimostrare che la frequenza relativa, in

opportune ipotesi che traducono in termini precisi l'affermazione "prove indipendenti ed effettuate in identiche condizioni", converge alla probabilità (**Legge dei grandi numeri**). Si osservi tuttavia che la (2.4) ha notevole valore operativo in quanto, in ipotesi di norma soddisfatte in pratica, risulta:

$$P(E) \approx \frac{N(E)}{N} \quad N \gg 1$$

Tale relazione fornisce la base per la stima di una probabilità.

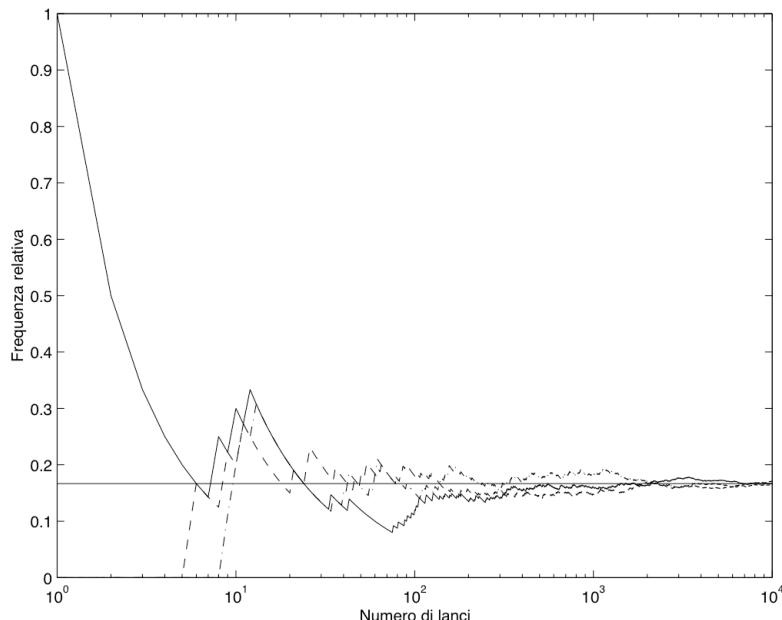


Figura 2.9: Andamento della frequenza relativa dell'evento $E \equiv \{ \text{"facce uguali"} \}$.

A scopo illustrativo si riprenda in esame il lancio di una coppia di dadi, precedentemente analizzato, ed in particolare l'evento:

$$E \equiv \{ \text{"facce uguali"} \}$$

la cui probabilità vale $1/6$. In Fig. 2.9 è riportato l'andamento della frequenza relativa di tale evento in funzione del numero delle prove: l'andamento della frequenza di successo è molto irregolare se il numero delle prove è basso, ma, mano a mano che tale numero cresce, lo scostamento dal valore precedentemente calcolato diventa trascurabile. Nella stessa figura l'andamento della frequenza di successo è riportato per tre diverse serie di prove: i vari

andamenti sono inizialmente molto diversi, ma la loro differenza è trascurabile se il numero delle prove è sufficientemente elevato. Si osservi infine che gli andamenti riportati in figura 2.9 si possono anche interpretare come la conferma sperimentale delle ipotesi che sono alla base dell'analisi precedentemente effettuata.

2.2 PROBABILITÀ CONDIZIONATA E INDIPENDENZA

Da un punto di vista intuitivo due eventi sono indipendenti se non si influenzano: se cioè il verificarsi dell'uno non altera le aspettative che si hanno sul verificarsi o meno dell'altro. Tale punto di vista intuitivo, come sarà meglio chiarito nel prossimo paragrafo quando verrà introdotta la probabilità condizionata, viene formalizzato con la seguente definizione: due eventi A e B si dicono indipendenti se:

$$P(A \cap B) = P(A)P(B) \quad (2.5)$$

Si noti che tale definizione è simmetrica nel senso che se A è indipendente da B anche B è indipendente da A . Eventi non indipendenti sono anche detti correlati.

La definizione di indipendenza non va confusa con quella di eventi mutuamente esclusivi. Infatti se gli eventi A e B sono indipendenti allora la loro probabilità congiunta è data dalla (2.5); conseguentemente possono essere anche mutuamente esclusivi solo se almeno uno dei due ha probabilità nulla. Pertanto eventi mutuamente esclusivi, aventi probabilità non nulla, sono correlati. Tale affermazione si giustifica anche intuitivamente: infatti se gli eventi sono mutuamente esclusivi il verificarsi dell'uno esclude il verificarsi dell'altro e quindi saper o meno che uno dei due si è verificato influenza la nostra aspettativa sul verificarsi o meno dell'altro.

2.2.1 Probabilità condizionata

Spesso, con riferimento ad un determinato esperimento, le probabilità cambiano in relazione all'eventuale informazione disponibile in merito ai risultati dell'esperimento stesso. Intuitivamente la probabilità di ottenere "un due" lanciando un dado è diversa dalla probabilità che si verifichi tale evento se è noto che il risultato della prova è pari. Tale considerazione viene formalizzata introducendo la probabilità condizionata. A tal fine, dato un evento B , avente probabilità non nulla, si ha

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \quad P(B) \neq 0 \quad (2.6)$$

Tale corrispondenza soddisfa gli assiomi di Kolmogorov e, quindi, definisce una nuova legge di probabilità denominata *probabilità condizionata*. Invero si ha:

A1 $P(A|B) \geq 0, \quad \forall A \in \mathcal{E}$

A2 $P(\Omega|B) = \frac{P(\Omega \cap B)}{P(B)} = 1$

A3 $P(A_1 \cup A_2|B) = \frac{P((A_1 \cup A_2) \cap B)}{P(B)} = \frac{P(A_1 \cap B) + P(A_2 \cap B)}{P(B)}$

$$= P(A_1|B) + P(A_2|B), \quad \forall A_1, A_2 \in \mathcal{E} : A_1 \cap A_2 = \emptyset$$

La definizione data di probabilità condizionata soddisfa le condizioni che intuitivamente sono richieste alla probabilità condizionata. Infatti, dalla (2.6) segue che:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = 1 \quad \forall A \supseteq B$$

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A)}{P(B)} \geq P(A) \quad \forall A \subseteq B$$

$$P(A|B) = 0 \quad \forall A, B : A \cap B = \emptyset$$

La prima di tali relazioni afferma che, se il verificarsi di B è condizione sufficiente per il verificarsi di A , allora la probabilità di A condizionata a B deve essere uno; la seconda asserisce che, se il verificarsi di B è condizione necessaria per il verificarsi di A , allora la probabilità di A condizionata a B non deve diminuire; infine la terza relazione afferma che se A e B sono mutuamente esclusivi, allora il verificarsi dell'uno esclude il verificarsi dell'altro e, quindi la probabilità dell'uno condizionata all'altro deve essere nulla. Tali proprietà sono illustrate in figura 2.10. Si noti inoltre che, se A e B sono statisticamente indipendenti ed a

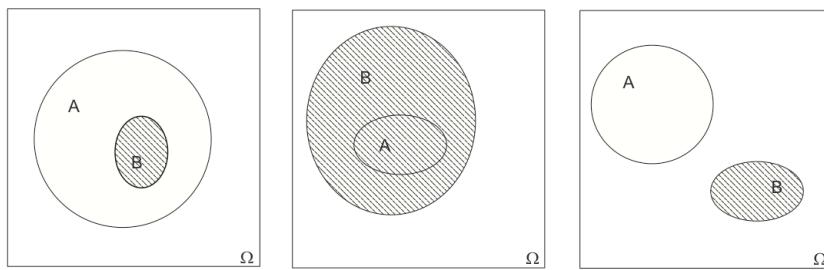


Figura 2.10: Diagrammi di Venn connessi alle proprietà della probabilità condizionata

probabilità non nulla allora risulta:

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

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

In altri termini, per eventi statisticamente indipendenti, il verificarsi o meno dell'uno non muta la probabilità di verificarsi dell'altro. Viceversa per eventi correlati probabilità a priori e probabilità condizionate differiscono. In tal caso, se $P(A|B) > P(A)$ si dice che A è positivamente correlato a B , altrimenti si dice che A è negativamente correlato a B . È possibile dimostrare che se A è correlato positivamente a B allora anche B è positivamente correlato ad A . Si noti che $P(A|B)$ oltre che probabilità di A condizionata a B è anche numericamente uguale alla *verosimiglianza* di B per A . Le due nozioni però sono diverse: invero $P(\cdot|B)$, come già detto è una legge di probabilità, mentre la verosimiglianza, definita da $P(A|\cdot)$, come è facile verificare, non lo è. In altri termini nel primo caso l'evento condizionante è fisso, mentre varia quello condizionato; nel caso della verosimiglianza invece, è l'evento condizionato a non variare, mentre quello condizionante varia.

Il significato della probabilità condizionata è ulteriormente chiarito dalla sua interpretazione frequentistica. Infatti, secondo tale interpretazione risulta:

$$P(A|B) \approx \frac{\frac{N_{AB}}{N}}{\frac{N_B}{N}} = \frac{N_{AB}}{N_B}$$

ove N_{AB} e N_B denotano rispettivamente il numero di volte che si verificano in N prove gli eventi $A \cap B$ e, rispettivamente, B . Pertanto la probabilità condizionata è la frazione di volte che si verifica l'evento A non più in tutte le prove, ma limitatamente a quelle in cui si verifica B . In altri termini le prove in cui B non si verifica vanno scartate e non contribuiscono più al calcolo della frequenza relativa. La definizione di probabilità condizionata e le sue proprietà sono presentate sinteticamente nella tabella 2.5.

Tabella 2.5

Probabilità condizionata e sue proprietà.

Probabilità condizionata	$P(A B) \triangleq \frac{P(A \cap B)}{P(B)}$	$P(B) \neq 0$
	$P(A B) = 1$	$\forall A \supseteq B$
Proprietà	$P(A B) = \frac{P(A)}{P(B)} \geq P(A), \quad \forall A \subseteq B$	
	$P(A B) = 0$	$\forall A, B : A \cap B = \emptyset$
Eventi indipendenti	$P(A B) = P(A)$	
	$P(B A) = P(B)$	

2.2.2 Leggi Fondamentali

Legge della probabilità composta

Dalla definizione di probabilità condizionata segue che:

$$P(A \cap B) = P(A)P(B|A) = P(B)P(A|B)$$

Tale relazione esprime la cosiddetta legge della probabilità composta (o congiunta): la probabilità di una congiunzione di eventi è data dal prodotto della probabilità incondizionata dell'uno per la probabilità condizionata dell'altro.

Legge di Bayes

Dalla legge della probabilità congiunta segue che:

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

Tale relazione è nota come legge o formula di Bayes e consente di scambiare i ruoli di evento condizionato e condizionante.

Legge della probabilità totale

Si consideri una partizione $\{A_n\}$ finita o numerabile di Ω , costituita da eventi a probabilità non nulla, come illustrato in figura 2.11.

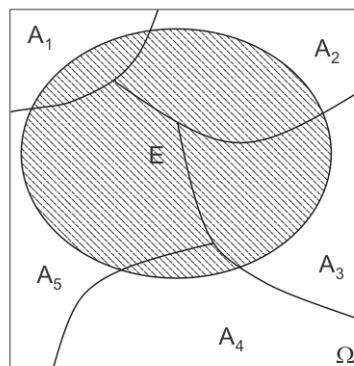


Figura 2.11: Partizione dell'evento certo ricoprente E

Si noti che gli eventi A_n , $n \in I$, sono a due a due mutuamente esclusivi e la loro unione è Ω . Osservando che

$$E = E \cap \Omega = E \cap \bigcup_{n \in I} A_n = \bigcup_{n \in I} E \cap A_n$$

ed essendo le congiunzioni $E \cap A_n$, $n \in I$ a due a due mutuamente esclusive, per l'additività della probabilità si ottiene

$$P(E) = \sum_{n \in I} P(E \cap A_n) = \sum_{n \in I} P(A_n)P(E|A_n)$$

relazione nota come legge della probabilità totale.

Legge di Bayes (seconda formulazione)

Esprimendo nella formula di Bayes la probabilità dell'evento condizionante a mezzo della legge della probabilità totale si ottiene la seguente formulazione alternativa della legge di Bayes

$$P(A_n|E) = \frac{P(A_n)P(E|A_n)}{\sum_{i \in I} P(A_i)P(E|A_i)}, \quad \forall n \in I$$

Per comodità del lettore la definizione di probabilità condizionata, le sue proprietà e leggi fondamentali della probabilità sono riassunte nella tabella 2.6.

Tabella 2.6
Leggi fondamentali.

Legge della probabilità composta	$P(A \cap B) = P(A)P(B A) = P(B)P(A B)$
Legge di Bayes	$P(A B) = \frac{P(A)}{P(B)}P(B A)$
Legge della probabilità totale	$P(E) = \sum_{n \in I} P(A_n)P(E A_n)$
Legge di Bayes seconda formulazione	$P(A_n E) = \frac{P(A_n)P(E A_n)}{\sum_{n \in I} P(E A_n)}$

2.3 VARIABILI ALEATORIE

Si consideri il lancio di una moneta: i possibili risultati di tale esperimento sono

$$\Omega = \{\text{testa, croce}\}$$

e possibili eventi sono tutti i sottoinsiemi di Ω ; per completare la descrizione probabilistica di tale esperimento è sufficiente assegnare la probabilità p dell'evento $\{\text{testa}\}$. Tale esperimento

è del tutto equivalente all'emissione di un bit da parte di una sorgente binaria nel senso che i due esperimenti differiscono solo per la denominazione dei due possibili risultati, 0 e 1 per la sorgente binaria, *testa* e *croce* per il lancio della moneta.

Si consideri ora l'esperimento del lancio di un dado: i possibili risultati di tale esperimento sono

$$\Omega = \{f_1, f_2, f_3, f_4, f_5, f_6\}$$

e si supponga che si sia interessati non al particolare risultato del lancio, ma solo a stabilire se il risultato sia pari o dispari, cioè se si verifichi o meno l'evento:

$$E = \{f_2, f_4, f_6\}$$

In tal caso per caratterizzare probabilisticamente l'esperimento è sufficiente considerare come possibili eventi

$$\mathcal{E} = \{\emptyset, E, \overline{E}, \Omega\}$$

ed assegnare la probabilità p dell'evento E ; anche tale esperimento è equivalente ai precedenti se ci si limita a considerare come possibili risultati la *parità* o meno del lancio del dado. Le considerazioni esposte trovano un inquadramento formale nel concetto di variabile aleatoria (v.a.). A tale scopo si consideri un esperimento aleatorio e sia Ω lo spazio dei campioni: una variabile aleatoria X è una funzione reale² definita su Ω , cioè:

$$X : \omega \in \Omega \longrightarrow x = X(\omega) \in \mathbb{R} \quad (2.7)$$

Il codominio della v.a. X è detto alfabeto di X . In altri termini una variabile aleatoria associa un valore numerico ai risultati di un esperimento. La figura 2.12 chiarisce, da un punto di vista grafico, la definizione di variabile aleatoria.

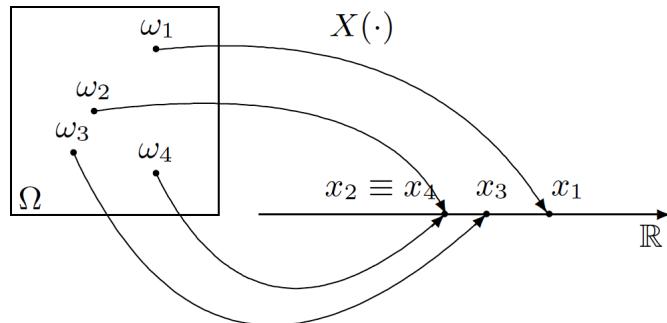
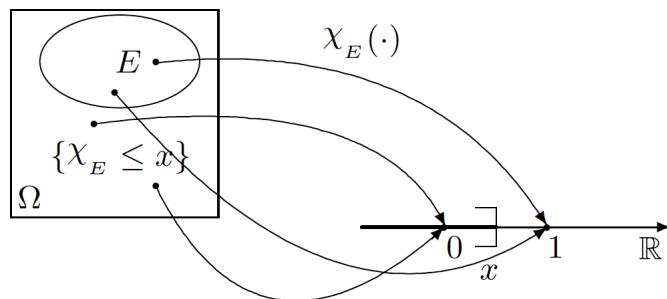
Indicatore di un evento

Si definisce indicatore dell'evento E , e la si denota con χ_E , la variabile aleatoria:

$$\chi_E : \omega \in \Omega \longrightarrow \begin{cases} 1 & \text{se } \omega \in E \\ 0 & \text{se } \omega \notin E \end{cases}$$

il cui alfabeto è l'insieme discreto $\{0, 1\} \subseteq \mathbb{R}$. La corrispondenza tra lo spazio dei campioni e l'insieme dei numeri reali definita dalla v.a. χ_E è illustrata in figura 2.13. Si noti che tale variabile aleatoria elimina gli aspetti che non interessano dell'esperimento in esame, nel senso che non distingue i valori dei singoli risultati, ma solo se appartengono o meno ad E .

² talvolta si considerano v.a. a valori in $\overline{\mathbb{R}} = [-\infty, +\infty]$ in tal caso si impone però la condizione $P(\{X = \pm\infty\}) = 0$

Figura 2.12: Rappresentazione della variabile aleatoria X .Figura 2.13: La variabile aleatoria χ_E , indicatrice dell'evento E .

Istante d'arrivo del viaggiatore

Si riprenda in esame l'esempio relativo ad una coppia di arrivi nell'intervallo $[0, T]$ introdotto nel capitolo precedente; la funzione:

$$I_V : (t_V, t_T) \in [0, T]^2 \longrightarrow t_V$$

definisce la variabile aleatoria "istante d'arrivo del viaggiatore" il cui alfabeto è l'intervallo continuo $[0, T]$. Si noti che, fissato il tempo t_V tutte le uscite sperimentali sul segmento (t_V, t_T) vengono mappate nell'unico valore t_V della v.a. I_V .

2.4 FUNZIONE DI DISTRIBUZIONE CUMULATIVA

La caratterizzazione probabilistica di una v.a. X è fornita dalla sua distribuzione di probabilità $P_X(\cdot)$; ma la $P_X(\cdot)$ è una funzione d'insieme e quindi non è uno strumento analitico con cui si è in genere molto familiari. In alternativa tale caratterizzazione può essere effettuata tramite

una funzione reale di variabile reale, la Funzione di Distribuzione Cumulativa (CDF), definita da:

$$F_X : x \in \mathbb{R} \longrightarrow F_X(x) = P(\{X \leq x\}) \quad (2.8)$$

ed il suo calcolo va condotto valutando la probabilità dell'evento $\{X \leq x\}$.

CDF di una variabile aleatoria Bernoulliana

Una v.a. Bernoulliana è una v.a. X binaria avente alfabeto $\mathcal{A}_X = \{0, 1\}$. Una v.a. Bernoulliana di parametro p , ove $p = P(\{X = 1\})$, si indica sinteticamente mediante la notazione

$$X \sim \mathcal{B}(1, p). \quad (2.9)$$

Per calcolare la CDF di una v.a. Bernoulliana, consideriamo l'evento $\{X \leq x\}$, costituito dai seguenti elementi

$$\{X \leq x\} = \begin{cases} \emptyset & \text{se } x < 0 \\ \{1\} & \text{se } 0 \leq x < 1 \\ \Omega & \text{se } x \geq 1 \end{cases}$$

per cui si ottiene

$$F_X(x) = \begin{cases} 0 & \text{se } x < 0 \\ q = 1 - p & \text{se } 0 \leq x < 1 \\ 1 & \text{se } x \geq 1 \end{cases} \quad (2.10)$$

riportata in figura 2.14.

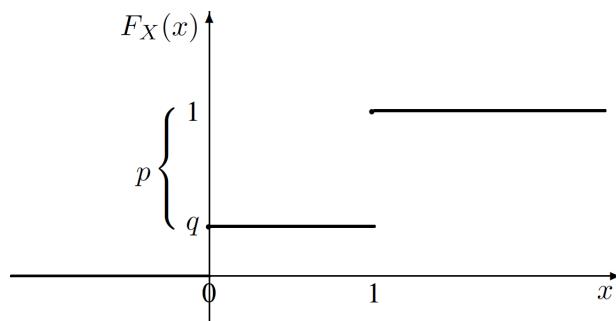


Figura 2.14: CDF di una v.a. Bernoulliana

CDF dell'indicatore di un evento

Si consideri la variabile I_E indicatrice dell'evento E . Come è facile verificare, risulta

$$\{I_E \leq x\} = \begin{cases} \emptyset & \text{se } x < 0 \\ \bar{E} & \text{se } 0 \leq x < 1 \\ \Omega & \text{se } x \geq 1 \end{cases}$$

Conseguentemente, la CDF di una variabile indicatrice è Bernoulliana con parametro $q = P(\bar{E})$.

CDF dell'istante d'arrivo del viaggiatore I_V

Come secondo esempio, calcoliamo la CDF della v.a. "istante d'arrivo del viaggiatore" I_V introdotta precedentemente. Avendosi

$$\{I_V \leq x\} = \begin{cases} \emptyset & \text{se } x < 0 \\ [0, x] \times [0, T] & \text{se } 0 \leq x \leq T \\ \Omega & \text{se } x \geq T \end{cases}$$

la CDF di I_V vale:

$$F_{I_V}(x) = \begin{cases} 0 & \text{se } x < 0 \\ \frac{x}{T} & \text{se } 0 \leq x \leq T \\ 1 & \text{se } x \geq T \end{cases}$$

Una v.a. avente tale CDF dicesi uniforme nell'intervallo $[0, T]$ e sinteticamente si scrive

$$I_V \sim U(0, T).$$

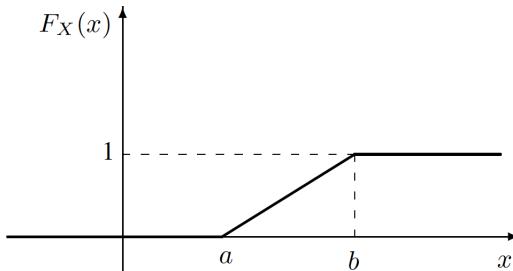
Più in generale, una qualsiasi v.a. avente la CDF:

$$F_X(x) = \begin{cases} 0 & \text{se } x < a \\ \frac{x-a}{b-a} & \text{se } a \leq x \leq b \\ 1 & \text{se } x \geq b \end{cases} \quad (2.11)$$

riportata in figura 2.15, sinteticamente denominata

$$X \sim U(a, b). \quad (2.12)$$

si dice uniforme nell'intervallo $[a, b]$. ◀

Figura 2.15: Funzione di distribuzione cumulativa della variabile $X \sim U(a, b)$

2.4.1 CDF empirica

È utile dare una interpretazione frequentistica della CDF; a tale scopo, supposto di avere n osservazioni della v.a. X , diciamole:

$$x(1), x(2), \dots, x(n)$$

si definisce Funzione di Distribuzione Cumulativa Empirica, o sinteticamente ECDF, la seguente funzione reale di variabile reale

$$\hat{F}_X : x \in \mathbb{R} \longrightarrow \hat{F}_X(x) = \frac{\{(x(i) \leq x)\}}{n} \in [0, 1] \subseteq \mathbb{R} \quad (2.13)$$

ove $\{(x(i) \leq x)\}$ denota il numero di osservazioni $x(i)$, $i = 1, 2, \dots, n$, minori o uguali ad x . Se il numero n delle osservazioni è sufficientemente elevato l'ECDF rappresenta una buona stima della CDF, si ha cioè

$$F_X(x) \approx \hat{F}_X(x)$$

Osserviamo esplicitamente che un solo insieme di osservazioni è sufficiente per stimare l'intera CDF; inoltre il calcolo è più efficiente se si ordinano le osservazioni. Infatti dette

$$x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$$

le osservazioni ordinate, come è immediato verificare, si ha:

$$\hat{F}_X(x) = \begin{cases} 0 & \text{se } x < x_{(1)} \\ \frac{k}{n} & \text{se } x_{(k)} \leq x < x_{(k+1)} \\ 1 & \text{se } x \geq x_{(n)} \end{cases}$$

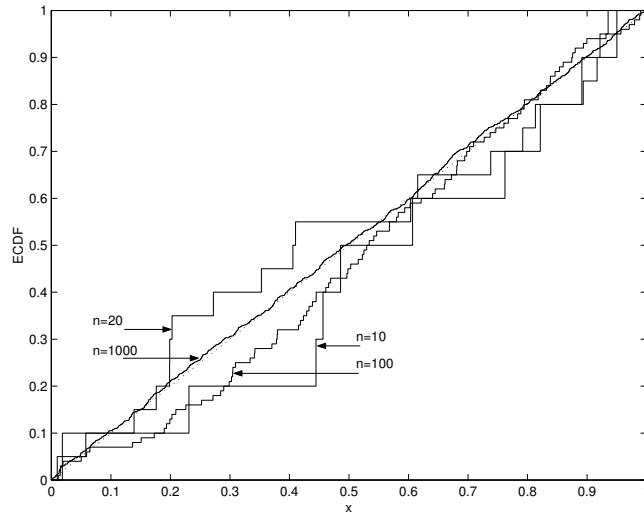


Figura 2.16: ECDF di una v.a. $U \sim U(0, 1)$ valutata su n osservazioni

Nella figura 2.16 è riportato l’andamento della ECDF di una v.a. $U \sim U(0, 1)$ calcolata su n osservazioni, con $n \in \{10, 20, 100, 1000\}$; nella stessa figura, con linea tratteggiata, è anche riportato il grafico della CDF. Le curve riportate mostrano come, al crescere del numero delle osservazioni, la misura della CDF diventi sempre più accurata.

2.4.2 Proprietà della CDF

Come detto nei paragrafi precedenti la CDF caratterizza, in termini probabilistici, una variabile aleatoria. Le principali proprietà della CDF sono riassunte nella tabella 2.7; tali proprietà sono dette anche *costitutive* in quanto ogni funzione che soddisfa tali proprietà può essere riguardata come la CDF di una variabile aleatoria. Tale osservazione giustifica perché spesso in pratica si introducano i vari modelli di v.a. senza precisare lo spazio di probabilità su cui sono definite. Sulla scorta del tipo di CDF, le variabili aleatorie possono essere di tipo discreto, continuo e misto. Precisamente una v.a. si dice di tipo discreto se la sua CDF è costante a tratti ovvero è una funzione a scalini; si dice di tipo continuo se la CDF è continua su tutto \mathbb{R} , mentre si dice di tipo misto in tutti gli altri casi, se cioè la CDF è discontinua su un insieme discreto di punti, ma non è costante a tratti.

2.5 FUNZIONE DI DISTRIBUZIONE DI PROBABILITÀ

Una v.a. X si dice discreta se tale è il suo alfabeto, per cui, se lo spazio dei campioni Ω è discreto ogni variabile aleatoria definita su di esso è di tipo discreto. Le v.a. discrete sono

Tabella 2.7
Proprietà costitutive della CDF

P0	Definizione	$F_X : x \in \mathbb{R} \longrightarrow F_X(x) \triangleq P(\{X \leq x\})$
P1	Valori in $\pm\infty$	$F_X(-\infty) = 0 \quad \& \quad F_X(+\infty) = 1$
P2	La CDF è non decrescente	$x_1 < x_2 \implies F_X(x_1) \leq F_X(x_2)$
P3	Continuità da destra	$\lim_{h \rightarrow 0^+} F_X(x + h) = F_X(x)$

caratterizzate dall'avere una CDF costante a tratti, cioè del tipo riportato in figura (2.17). Per

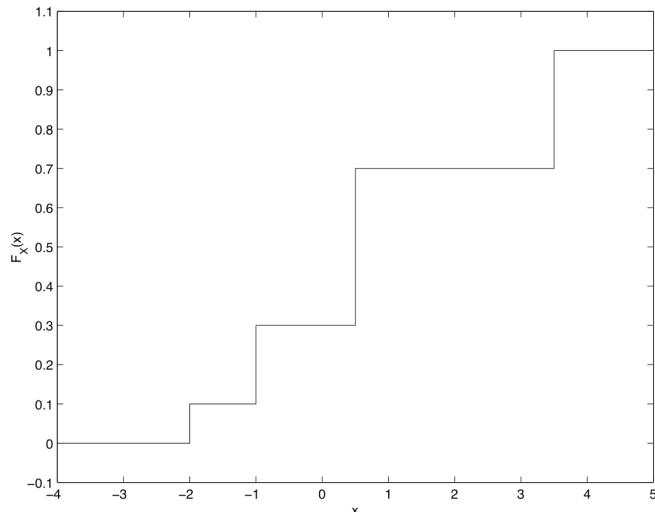


Figura 2.17: Andamento tipico della CDF di una v.a. discreta.

le v.a. discrete una caratterizzazione probabilistica alternativa alla CDF è quella in termini di funzione di distribuzione di probabilità, detta anche funzione masse di probabilità (PMF). Precisamente la PMF $P_x(\cdot)$ di una v.a. discreta X è la funzione:

$$P_x : x \in \mathcal{A}_X \longrightarrow P_x(x) = P(\{X = x\}) \quad (2.14)$$

Dalla definizione di PMF segue che tale funzione gode delle seguenti due proprietà, per altro di semplice verifica:

Si noti che tali proprietà sono caratterizzanti nel senso che non solo una qualunque PMF le

P1 $P_x(x) \geq 0 \quad \forall x \in \mathcal{A}_X$ (Non negatività)

P2 $\sum_{x \in \mathcal{A}_X} P_x(x) = 1$ (Normalizzazione)

soddisfa, ma anche una qualsiasi funzione definita su un insieme discreto \mathcal{A}_X che goda di tali proprietà è la PMF di una qualche variabile aleatoria.

Nota la PMF di una v.a. discreta, è immediato valutarne la CDF come somma di tutte le masse di probabilità corrispondenti a valori che non eccedono il valore x in cui si vuole valutare la PMF, cioè

$$F_X(x) = \sum_{\{v \in \mathcal{A}_X : v \leq x\}} P_X(v) \quad (2.15)$$

Le proprietà della PMF sono riassunte nella tabella 2.8.

Tabella 2.8
Principali proprietà della PMF

P0	Definizione	$P_x : x \in \mathcal{A}_X \longrightarrow P(\{X = x\})$
P1	Non negatività	$P_x(x) \geq 0$
P2	Normalizzazione	$\sum_{x \in \mathcal{A}_X} P_x(x) = 1$
P3	Probabilità di un evento	$P(\{X \in A\}) = \sum_{x \in A \cap \mathcal{A}_X} P_x(x)$
P4	CDF	$F_X(x) = \sum_{\{v \in \mathcal{A}_X : v \leq x\}} P_X(v)$

Distribuzione di Poisson

La distribuzione di Poisson è definita dalla PMF

$$P_X(k) = e^{-a} \frac{a^k}{k!} \quad a > 0, \quad k = 0, 1, \dots \quad (2.16)$$

dipendente dal parametro a . Essa è non negativa; inoltre, avendosi

$$\sum_{k=0}^{+\infty} P_X(k) = e^{-a} \sum_{k=0}^{+\infty} \frac{a^k}{k!} = e^{-a} e^a = 1$$

essa è una PMF.

Distribuzione Binomiale

Consideriamo l'esperimento descritto dalla successione di n prove, in cui il risultato della singola prova è di tipo *successo* oppure *insuccesso*. L'esperimento può essere descritto mediante una successione di variabili aleatorie Bernoulliane X_i ognuna delle quali assume il valore "1" nel caso di successo oppure il valore "0" nel caso di insuccesso. I due risultati si ottengono con probabilità p e $q = 1 - p$, rispettivamente. La variabile aleatoria

$$X = \sum_{i=1}^n X_i \quad (2.17)$$

è una v.a. discreta avente alfabeto $\mathcal{A}_X = \{0, 1, \dots, n\}$ e la sua PMF sinteticamente denominata con

$$X \sim \mathcal{B}(n, p). \quad (2.18)$$

prende il nome di Bernoulliana di parametri n e p . La probabilità che la variabile aleatoria X assuma il valore k è data da

$$P_X(k) = P(\{X = k\}) = \binom{n}{k} p^k q^{n-k}, \quad k = 0, 1, \dots, n. \quad (2.19)$$

Si noti che $P_X(k)$ corrisponde alla probabilità di avere esattamente k successi ed $n - k$ insuccessi in n prove indipendenti (tale probabilità è uguale a $p^k q^{n-k}$) presa tante volte quante sono le possibili configurazioni con cui k successi ed $n - k$ insuccessi possono avvenire in n prove. Il numero di tali configurazioni è dato dal coefficiente binomiale

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}.$$

2.6 FUNZIONE DI DENSITÀ DI PROBABILITÀ

Per le variabili aleatorie continue, una caratterizzazione probabilistica alternativa rispetto alla CDF è quella in termini di funzione densità di probabilità (pdf): precisamente la pdf $f_X(\cdot)$ di una v.a. continua è la derivata della CDF, cioè si ha:

$$f_X : x \in \mathbb{R} \longrightarrow f_X(x) = \frac{dF_X(x)}{dx} \quad (2.20)$$

Conseguentemente la CDF è una primitiva della pdf e precisamente quella che vale 0 in $-\infty$; in altri termini risulta:

$$F_X(x) = \int_{-\infty}^x f_X(u) du \quad (2.21)$$

Si noti che dalle relazioni (2.20) e (2.21) segue che, contrariamente alla CDF e alla PMF, la PDF non è una probabilità, ma le sue dimensioni fisiche sono date dal reciproco di quelle

della v.a.; ad esempio se la X è un tensione aleatoria allora le dimensioni della sua PDF sono Volt $^{-1}$, se X è una durata allora le dimensioni della sua PDF sono sec $^{-1}$, e così via.

Dalla definizione di derivata segue che la probabilità che la v.a. assuma valori in $A \subseteq \mathbb{R}$, cioè la probabilità dell'evento $\{X \in A\}$, è calcolabile come l'integrale esteso ad A della PDF, infatti

$$P(\{X \in A \subseteq \mathbb{R}\}) = \int_A f_X(x) dx$$

La PDF gode delle seguenti due proprietà:

P1 $f_X(x) \geq 0, \forall x \in \mathbb{R}$ (la PDF è non negativa)

P2 $\int_{-\infty}^{+\infty} f_X(x) dx = 1$ (normalizzazione)

La prima deriva dalla monotonicità della CDF, mentre la seconda esprime la probabilità dell'evento certo in termini di PDF. Tali proprietà caratterizzano una PDF nel senso che una qualsiasi funzione reale che le soddisfi è la PDF di una v.a. (Teorema di esistenza per v.a. continue). Definizione e proprietà della PDF sono riassunte nella tabella 2.9.

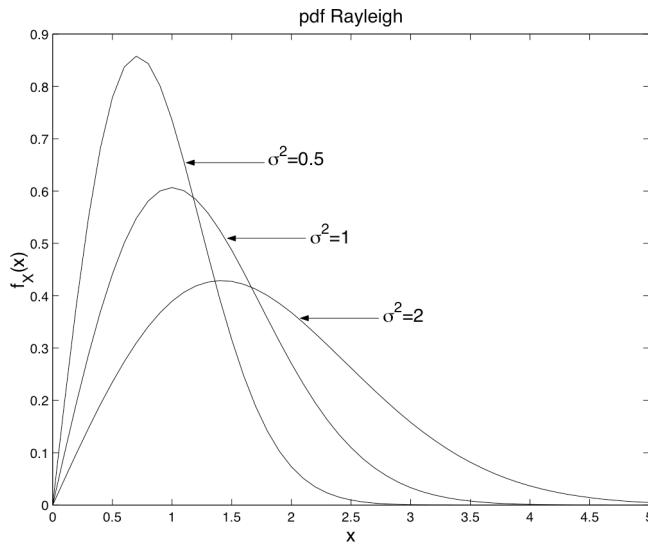
Tabella 2.9

		Proprietà della PDF
P0	Definizione	$f_X : x \in \mathbb{R} \longrightarrow f_X(x) = \frac{d}{dx} F_X(x)$ $f_X(x) = \lim_{\Delta x \rightarrow 0} \frac{P(\{x - \Delta x < X \leq x\})}{\Delta x}$
P1	Non negatività	$f_X(x) \geq 0$
P2	Normalizzazione	$\int_{-\infty}^{+\infty} f_X(x) dx = 1$
P3	Probabilità di un evento	$P(\{X \in A\}) = \int_A f_X(x) dx$
P4	CDF	$F_X(x) = \int_{-\infty}^x f_X(\xi) d\xi$

Variabile di Rayleigh

La pdf di una v.a. X di tipo Rayleigh di parametro σ^2 è definita come

$$f_X(x) = \frac{x}{\sigma^2} e^{-\frac{x^2}{2\sigma^2}} u(x) \quad (2.22)$$

Figura 2.18: PDF di variabili Rayleigh con $\sigma^2 = 0.5, 1, 2$

dove $u(x)$ è la funzione gradino unitario, uguale ad 1 per $x \in (0, \infty)$ e 0 altrimenti. Il cui grafico della pdf è riportato in figura 2.18, per alcuni valori del parametro σ^2 .

Come evidenziato in figura 2.18, e come è immediato verificare, il parametro σ è il valore modale, cioè il valore in corrispondenza del quale la PDF è massima.

Variabile uniforme

Si consideri una v.a. $X \sim U(a, b)$; derivando la CDF (2.11) si ha:

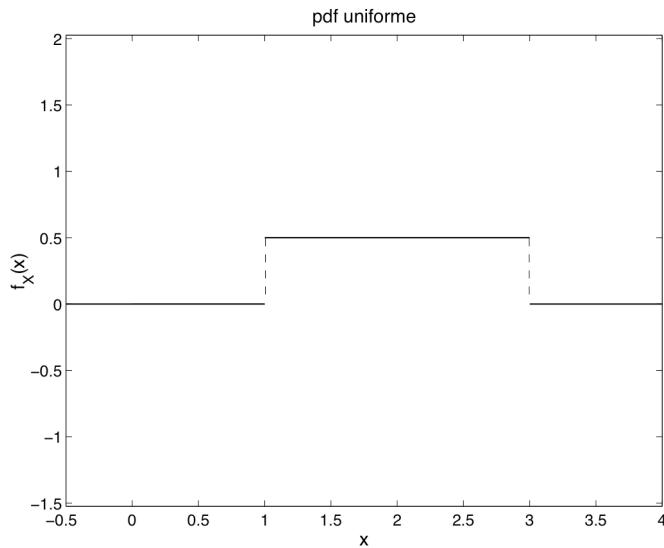
$$f_X(x) = \frac{1}{b-a} \text{rect}_{[a,b]}(x) \quad (2.23)$$

ove $\text{rect}_{[a,b]}(x)$ è la funzione:

$$\text{rect}_{[a,b]}(x) = \begin{cases} 1 & \text{se } x \in (a, b) \\ 0 & \text{se } x \notin [a, b] \end{cases}$$

In figura 2.19, a scopo esemplificativo, è riportato il grafico della pdf di una variabile aleatoria uniforme in $[1, 3]$.

Dunque una v.a. uniforme in $[a, b]$ è caratterizzata dall'avere una PDF costante all'interno di tale intervallo e nulla all'esterno.

Figura 2.19: PDF di un v.a. $X \sim U(1, 3)$ *Variabile esponenziale*

Si consideri la funzione

$$f(x) = \lambda e^{-\lambda x} u(x) \quad \lambda > 0 \quad (2.24)$$

il cui grafico è riportato in figura 2.20 per alcuni valori del parametro λ . Essa è non negativa ed integrata su \mathbb{R} vale 1: dunque è la PDF di una v.a.. Una v.a. avente la PDF definita dalla (2.24) è sinteticamente denotata:

$$X \sim Ex(\lambda)$$

e dicesi di tipo esponenziale di parametro λ .

Variabile Gaussiana o Normale

Una v.a. X_0 si dice Gaussiana standard se ha la seguente PDF:

$$f_{X_0}(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \quad (2.25)$$

il cui grafico è riportato in figura 2.21. La CDF della v.a. gaussiana standard si ricava dall'integrale

$$F_{X_0}(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-u^2/2} du$$

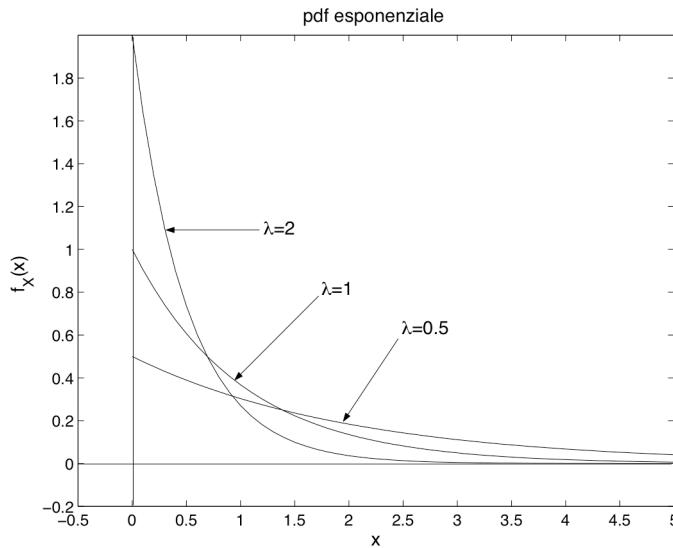


Figura 2.20: PDF di variabili esponenziali con $\lambda = 0.5, 1, 2$

che, tuttavia, non è esprimibile in forma chiusa. Il problema può essere risolto ricorrendo alla funzione $Q(\cdot)$ definita dall'integrale:

$$Q(x) \triangleq \int_x^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\xi^2/2} d\xi$$

per cui, applicando la proprietà di normalizzazione, risulta, $F_{X_0}(x) = 1 - Q(x)$. Definizione e proprietà della funzione $Q(\cdot)$ sono riportate nel paragrafo seguente.

Le v.a. gaussiane (non standard) sono la famiglia di variabili aleatorie generate dalla gaussiana standard mediante la trasformazione

$$X = \sigma X_0 + \mu, \quad \sigma > 0, \mu \in \mathbb{R} \quad (2.26)$$

e sinteticamente si scrive

$$X \sim \mathcal{N}(\mu, \sigma^2)$$

La pdf di una v.a. $\mathcal{N}(\mu, \sigma^2)$ si ricava da quella della gaussiana standard osservando che, per

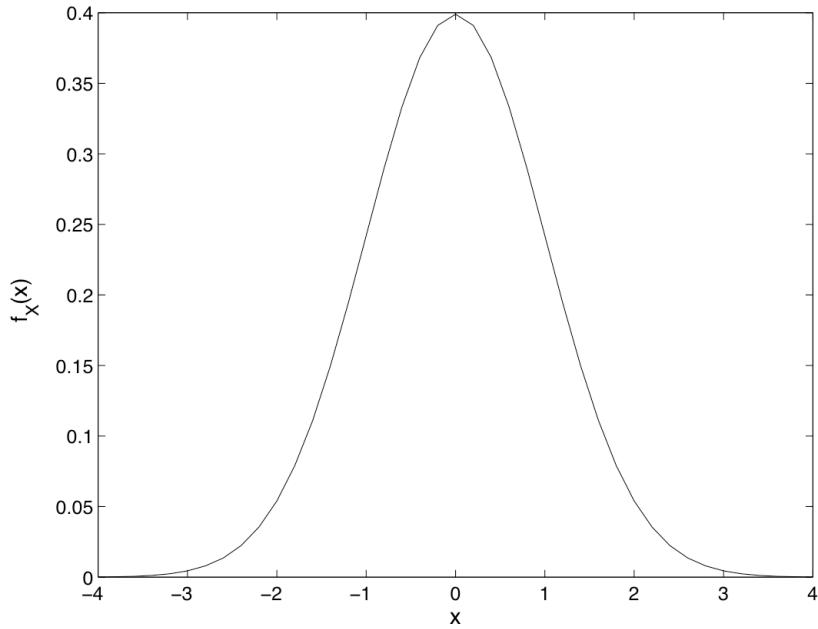


Figura 2.21: PDF della variabile gaussiana standard

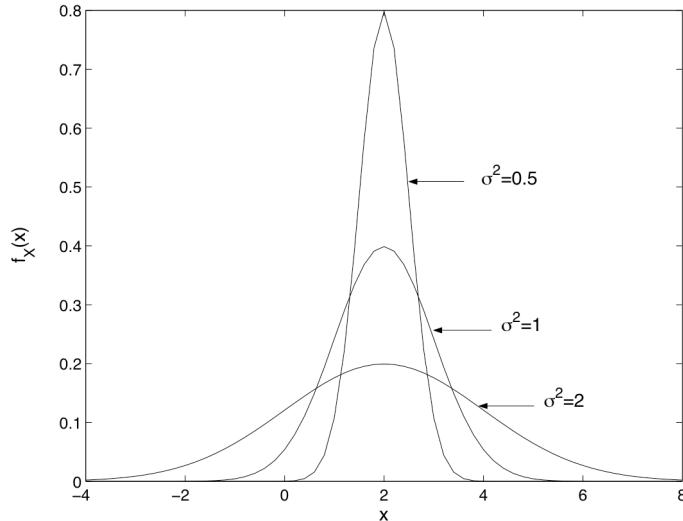
la CDF, risulta

$$\begin{aligned}
 F_X(x) &= P(\{X \leq x\}) = P(\{\sigma X_0 + \mu \leq x\}) \\
 &= P\left(\left\{X_0 \leq \frac{x-\mu}{\sigma}\right\}\right) = F_{X_0}\left(\frac{x-\mu}{\sigma}\right) \\
 &= 1 - Q\left(\frac{x-\mu}{\sigma}\right)
 \end{aligned} \tag{2.27}$$

da cui, derivando rispetto alla variabile x si ottiene

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right] \tag{2.28}$$

In figura 2.22 sono riportati i grafici della PDF gaussiana per alcuni valori dei parametri μ e σ^2 .

Figura 2.22: PDF di variabili gaussiane con $\mu = 2$ e $\sigma^2 = 0.5, 1, 2$

2.6.1 La funzione $Q(\cdot)$

La funzione $Q(\cdot)$ è, per definizione, la CDF complementare di una v.a. gaussiana standard

$$Q(x) \triangleq P(\{X_0 > x\}) = P(\{X_0 \geq x\}), \quad X_0 \sim \mathcal{N}(0, 1) \quad (2.29)$$

Si noti esplicitamente che, essendo la gaussiana una v.a. di tipo continuo, le probabilità non cambiano se le diseguaglianze sono strette ($<$ o $>$) oppure no (\leq o \geq). Poiché la funzione $Q(\cdot)$ è una CDF complementare, gode delle seguenti proprietà, di immediata dimostrazione:

P1 $Q(+\infty) = 0 \leq Q(x) \leq Q(-\infty) = 1$ (la $Q(\cdot)$ è limitata)

P2 $x_1 < x_2 \implies Q(x_1) > Q(x_2)$ (la $Q(\cdot)$ è strettamente decrescente)

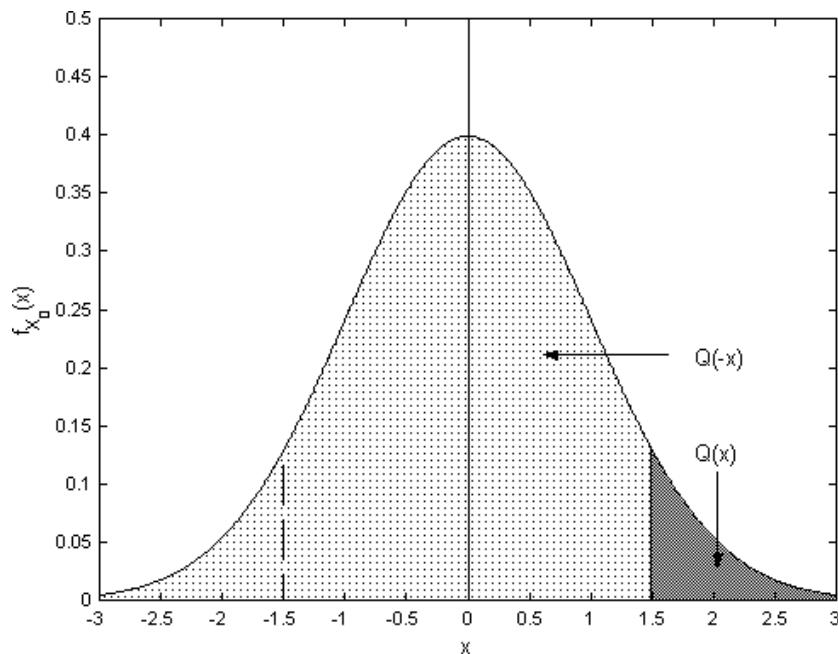
P3 $Q(-x) = 1 - Q(x)$ (simmetria)

La figura 2.23 illustra la definizione e la proprietà di simmetria della funzione $Q(\cdot)$. La funzione $Q(\cdot)$ non è una funzione elementare; pertanto va valutata per via numerica o a mezzo di tabelle: nel seguito ci limiteremo a considerare valori positivi dell'argomento, in quanto per valori negativi si può utilizzare la proprietà di simmetria. In tabella 2.10 sono riportati i valori della $Q(\cdot)$ per valori di $0 \leq x < 6$ e con passo 0.01; le prime due cifre della x sono riportate nella prima colonna, mentre la terza (centesimi) è riportata nella prima riga: ad esempio il valore di $Q(2.37)$ lo si ottiene in corrispondenza della riga etichettata 2.3 e della colonna etichettata 7.

x	0	1	2	3	4	5	6	7	8	9
0	0.5	0.49601	0.49202	0.48803	0.48405	0.48006	0.47608	0.4721	0.46812	0.46414
0.1	0.46017	0.4562	0.45224	0.44828	0.44433	0.44038	0.43644	0.43251	0.42858	0.42465
0.2	0.42074	0.41683	0.41294	0.40905	0.40517	0.40129	0.39743	0.39358	0.38974	0.38591
0.3	0.38209	0.37828	0.37448	0.3707	0.36693	0.36317	0.35942	0.35569	0.35197	0.34827
0.4	0.34458	0.3409	0.33724	0.33356	0.32997	0.32636	0.32276	0.31918	0.31561	0.31207
0.5	0.30854	0.30503	0.30153	0.29806	0.2946	0.29116	0.28774	0.28434	0.28096	0.27776
0.6	0.27425	0.27093	0.26763	0.26435	0.26109	0.25785	0.25463	0.25143	0.24825	0.2451
0.7	0.24196	0.23885	0.23576	0.2327	0.22965	0.22663	0.22363	0.22065	0.2177	0.21476
0.8	0.21186	0.20897	0.20611	0.20327	0.20045	0.19766	0.19489	0.19215	0.18943	0.18673
0.9	0.18406	0.18141	0.17879	0.17619	0.17361	0.17106	0.16853	0.16602	0.16354	0.16109
1	0.15866	0.15625	0.15386	0.15151	0.14917	0.14686	0.14457	0.14231	0.14007	0.13786
1.1	0.13567	0.1335	0.13136	0.12924	0.12714	0.12507	0.12302	0.121	0.119	0.11702
1.2	0.11507	0.11314	0.11123	0.10935	0.10749	0.10565	0.10383	0.10204	0.10027	0.098525
1.3	0.0968	0.095098	0.093418	0.091759	0.090123	0.088508	0.086915	0.085343	0.083793	0.082264
1.4	0.080757	0.07927	0.077804	0.076359	0.074934	0.073529	0.072145	0.070781	0.069437	0.068112
1.5	0.066807	0.065522	0.064255	0.063008	0.06178	0.060571	0.05938	0.058208	0.057053	0.055917
1.6	0.054799	0.053699	0.052616	0.051551	0.050503	0.049471	0.048457	0.04746	0.046479	0.045514
1.7	0.04465	0.043633	0.042716	0.041815	0.04093	0.040059	0.039204	0.038364	0.037538	0.036727
1.8	0.03593	0.035148	0.03438	0.033625	0.032884	0.032157	0.031443	0.030742	0.030054	0.029379
1.9	0.028717	0.028067	0.027429	0.026803	0.02619	0.025588	0.024998	0.024419	0.023852	0.023295
2	0.02275	0.022216	0.021692	0.021178	0.020675	0.020182	0.019699	0.019226	0.018763	0.018309
2.1	0.017864	0.017429	0.017003	0.016586	0.016177	0.015778	0.015386	0.015003	0.014629	0.014262
2.2	0.013903	0.013553	0.013209	0.012874	0.012545	0.012224	0.011911	0.011604	0.011304	0.011011
2.3	0.010724	0.010444	0.01017	0.0099031	0.0096419	0.0093867	0.0091375	0.008894	0.0086563	0.0084242
2.4	0.0081975	0.0079763	0.0077603	0.0075494	0.0073436	0.0071428	0.0069469	0.0067557	0.0065691	0.0063872
2.5	0.0062097	0.0060366	0.0058677	0.0057031	0.00553861	0.00535426	0.0052336	0.0050849	0.00494	0.0047988
2.6	0.0046612	0.0045271	0.0043965	0.0042692	0.0041453	0.0040246	0.003907	0.0037926	0.0036811	0.0035726
2.7	0.003467	0.0033642	0.0032641	0.0031667	0.003072	0.0029798	0.0028901	0.0028028	0.0027179	0.0026354
2.8	0.0025551	0.0024771	0.0024012	0.0023274	0.0022557	0.002186	0.0021182	0.0020524	0.0019884	0.0019262
2.9	0.0018658	0.0018071	0.0017502	0.0016948	0.0016411	0.0015889	0.0015382	0.001489	0.0014412	0.0013949
3	0.0013499	0.0013062	0.0012639	0.0012228	0.0011829	0.0011442	0.0011067	0.0010703	0.001035	0.0010008

x	0	1	2	3	4	5	6	7	8	9
3.1	0.0009676	0.00093544	0.00090426	0.00087403	0.00084474	0.00081635	0.00078885	0.00075219	0.00073638	0.00071136
3.2	0.00065714	0.00066367	0.00064095	0.00061895	0.00059765	0.00057703	0.00055706	0.00053774	0.00051904	0.00050094
3.3	0.00048342	0.00046648	0.00045009	0.00043423	0.00041889	0.00040406	0.00038971	0.00037584	0.00036243	0.00034946
3.4	0.00035693	0.00032481	0.00031311	0.00030179	0.00029086	0.00028029	0.00027009	0.00026023	0.00025071	0.00024151
3.5	0.00023263	0.00022405	0.00021577	0.00020778	0.00020006	0.00019262	0.00018543	0.00017849	0.00017118	0.00016534
3.6	0.00015911	0.0001531	0.0001473	0.00014171	0.00013632	0.00013112	0.00012128	0.00011662	0.00011213	
3.7	0.0001078	0.00010363	9.9611e-05	9.574e-05	9.201e-05	8.8417e-05	8.4957e-05	8.1624e-05	7.8414e-05	7.5324e-05
3.8	7.2348e-05	6.9483e-05	6.6726e-05	6.4072e-05	6.1517e-05	5.9059e-05	5.6694e-05	5.4418e-05	5.2228e-05	5.0122e-05
3.9	4.8096e-05	4.6148e-05	4.4274e-05	4.2473e-05	4.0741e-05	3.9076e-05	3.7475e-05	3.5936e-05	3.4458e-05	3.3037e-05
4	3.1671e-05	3.0359e-05	2.9099e-05	2.7888e-05	2.6776e-05	2.5609e-05	2.4536e-05	2.3507e-05	2.2518e-05	2.1569e-05
4.1	2.0658e-05	1.9783e-05	1.8944e-05	1.8138e-05	1.7365e-05	1.6624e-05	1.5912e-05	1.523e-05	1.4575e-05	1.3948e-05
4.2	1.3346e-05	1.2769e-05	1.2215e-05	1.1685e-05	1.1176e-05	1.0689e-05	1.0221e-05	9.7736e-06	9.3447e-06	8.9337e-06
4.3	8.5399e-06	8.1627e-06	7.8015e-06	7.4552e-06	7.1241e-06	6.8069e-06	6.5031e-06	6.2123e-06	5.934e-06	5.6675e-06
4.4	5.4125e-06	5.1683e-06	4.935e-06	4.7117e-06	4.4979e-06	4.2935e-06	4.098e-06	3.9116e-06	3.7322e-06	3.5612e-06
4.5	3.3977e-06	3.2414e-06	3.092e-06	2.9492e-06	2.8127e-06	2.6823e-06	2.5577e-06	2.4386e-06	2.3249e-06	2.2162e-06
4.6	2.1125e-06	2.0133e-06	1.9187e-06	1.8283e-06	1.742e-06	1.6597e-06	1.581e-06	1.506e-06	1.4344e-06	1.366e-06
4.7	1.3008e-06	1.2389e-06	1.1792e-06	1.1226e-06	1.0656e-06	1.0171e-06	9.6796e-07	9.2113e-07	8.7648e-07	8.3391e-07
4.8	7.9333e-07	7.5465e-07	7.1779e-07	6.8367e-07	6.492e-07	6.1731e-07	5.8693e-07	5.5794e-07	5.3043e-07	5.0418e-07
4.9	4.7918e-07	4.5538e-07	4.3272e-07	4.1115e-07	3.9061e-07	3.7107e-07	3.5247e-07	3.3476e-07	3.1793e-07	3.0196e-07
5	2.8665e-07	2.7215e-07	2.5836e-07	2.4524e-07	2.3277e-07	2.2091e-07	2.0963e-07	1.9891e-07	1.8872e-07	1.7903e-07
5.1	1.6983e-07	1.6108e-07	1.5277e-07	1.4487e-07	1.3737e-07	1.3024e-07	1.2347e-07	1.1705e-07	1.1094e-07	1.0515e-07
5.2	9.6644e-08	9.4422e-08	8.9462e-08	8.4755e-08	8.0288e-08	7.6058e-08	7.2028e-08	6.8212e-08	6.4592e-08	6.1158e-08
5.3	5.7901e-08	5.4813e-08	5.1884e-08	4.9168e-08	4.6473e-08	4.3977e-08	4.1611e-08	3.9368e-08	3.7243e-08	3.5229e-08
5.4	3.332e-08	3.1512e-08	2.988e-08	2.8177e-08	2.6648e-08	2.5185e-08	2.3807e-08	2.2502e-08	2.1266e-08	2.0097e-08
5.5	1.899e-08	1.7942e-08	1.695e-08	1.6012e-08	1.5124e-08	1.4283e-08	1.3489e-08	1.2737e-08	1.2026e-08	1.1353e-08
5.6	1.0718e-08	1.0110e-08	9.5479e-09	9.0105e-09	8.5025e-09	8.0224e-09	7.5686e-09	7.1399e-09	6.7347e-09	6.3532e-09
5.7	5.9904e-09	5.3262e-09	5.0215e-09	4.7338e-09	4.4262e-09	4.2028e-09	3.9636e-09	3.7355e-09	3.5193e-09	
5.8	3.3157e-09	3.1236e-09	2.9424e-09	2.7714e-09	2.616e-09	2.4579e-09	2.3143e-09	2.179e-09	2.0513e-09	1.931e-09
5.9	1.8175e-09	1.7105e-09	1.6097e-09	1.5147e-09	1.4251e-09	1.3407e-09	1.2612e-09	1.1863e-09	1.1157e-09	1.0492e-09

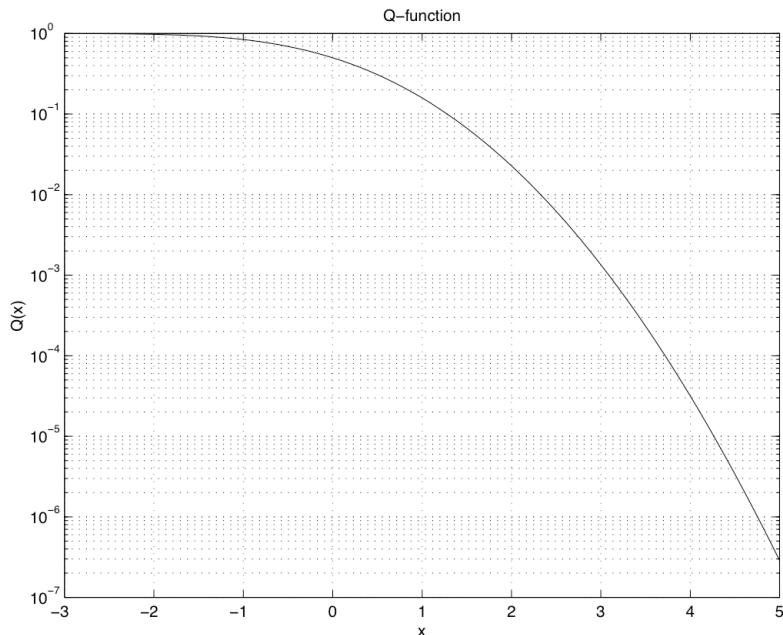
Tabella 2.10: Valori della funzione $Q(\cdot)$

Figura 2.23: Definizione e proprietà di simmetria della $Q(\cdot)$.

Per valori sufficientemente elevati dell'argomento ($x \geq 3$) si possono utilizzare le seguenti approssimazioni per difetto e per eccesso della $Q(\cdot)$

$$\frac{1}{\sqrt{2\pi}x} \left(1 - \frac{1}{x^2}\right) e^{-x^2/2} \leq Q(x) \leq \begin{cases} \frac{1}{2} e^{-x^2/2} \\ \frac{1}{\sqrt{2\pi}x} e^{-x^2/2} \end{cases}$$

Infine in figura 2.24 è riportato il grafico della funzione $Q(\cdot)$ in scala semilogaritmica.

Figura 2.24: Grafico della funzione $Q(\cdot)$.**Tabella 2.11**

Probabilità che $X \sim \mathcal{N}(\mu, \sigma^2)$ assuma valori in un intervallo.

$$P(\{X \leq x\}) = 1 - Q\left(\frac{x - \mu}{\sigma}\right)$$

$$P(\{X > x\}) = Q\left(\frac{x - \mu}{\sigma}\right)$$

$$P(\{x_1 < X \leq x_2\}) = Q\left(\frac{x_1 - \mu}{\sigma}\right) - Q\left(\frac{x_2 - \mu}{\sigma}\right)$$

$$P(\{-x < X \leq x\}) = 1 - Q\left(\frac{x + \mu}{\sigma}\right) - Q\left(\frac{x - \mu}{\sigma}\right)$$

2.7 MEDIA E MOMENTI DI UNA VARIABILE ALEATORIA

La caratterizzazione di una variabile aleatoria in termini di CDF o PDF/PMF non sempre è disponibile; ma, quand'anche lo fosse, spesso è una descrizione troppo dettagliata. Pertanto è utile avere una caratterizzazione più sintetica che descriva il comportamento globale della variabile aleatoria. Tale caratterizzazione consiste nel definire il comportamento medio delle grandezze oggetto dell'analisi. Si definisce valor medio della variabile aleatoria X il numero reale

$$\mathbb{E}[X] = \begin{cases} \sum_{x \in \mathcal{A}_X} x P_X(x) & \text{se } X \text{ è una v.a. discreta} \\ \int_{-\infty}^{+\infty} x f_X(x) dx & \text{se } X \text{ è una v.a. continua} \end{cases} \quad (2.30)$$

e viene anche denotato con μ_X , omettendo il pedice se è chiaro a quale variabile ci si riferisce.

Variabile aleatoria Bernoulliana

Per una v.a. Bernoulliana (2.9), risulta

$$\mathbb{E}[X] = 1 \times p + 0 \times q = p$$

Dunque la media di una v.a. Bernoulliana coincide con la probabilità che la variabile assuma il valore 1.

Variabile aleatoria esponenziale

Sia $X \sim \mathcal{E}x(\lambda)$; tenendo presente la (2.24) si ha

$$\mathbb{E}[X] = \int_{-\infty}^{+\infty} x \lambda e^{-\lambda x} u(x) dx = \frac{1}{\lambda}$$

Dunque la media di una variabile aleatoria esponenziale coincide col reciproco del parametro.

Variabile aleatoria gaussiana

Sia $X \sim \mathcal{N}(\mu, \sigma^2)$; tenendo presente la rappresentazione (2.26) per le v.a. gaussiane si ha:

$$\mathbb{E}[X] = \mathbb{E}[\sigma X_0 + \mu] = \sigma \mathbb{E}[X_0] + \mu$$

ove X_0 è una v.a. gaussiana standard. Inoltre, essendo:

$$\mathbb{E}[X_0] = \int_{-\infty}^{+\infty} x \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{x^2}{2}\right\} dx = 0$$

risulta

$$\mathbb{E}[X] = \mu$$

Dunque il parametro di posizione μ di una v.a. gaussiana coincide con la sua media.

Variabile aleatoria Binomiale

Sia $X \sim \mathcal{B}(n, p)$ una v.a. Binomiale di parametro p , ricordando che una v.a. binomiale può riguardarsi come la somma di n v.a. Bernoulliane indipendenti con lo stesso parametro, si ha:

$$\mathbb{E}[X] = \mathbb{E}\left[\sum_{i=1}^n X_i\right] = \sum_{i=1}^n \mathbb{E}[X_i] = np . \quad (2.31)$$

Teorema fondamentale per il calcolo della media

La relazione seguente risulta particolarmente utile per il calcolo della media di una funzione di variabili aleatorie. Considerata una v.a. X , la media di una sua funzione è calcolabile come

$$\mathbb{E}\{g(X)\} = \int_{-\infty}^{+\infty} g(x)f_X(x) dx \quad (2.32)$$

se la n variabile è di tipo continuo, ovvero come

$$\mathbb{E}\{g(X)\} = \sum_{x \in \mathcal{A}} g(x)P_X(x) \quad (2.33)$$

se la variabile è discreta, con alfabeto \mathcal{A} .

Pertanto per calcolare la media di una funzione di variabili aleatorie non è necessario il preventivo calcolo della PDF (o della PMF) della v.a. $g(X)$, ma è sufficiente la PDF (o la PMF) della v.a. X .

Il momento del secondo ordine:

$$\overline{X^2} = \mathbb{E}[X^2] \quad (2.34)$$

è anche denominato valore quadratico medio della v.a. X , mentre la sua radice quadrata

$$X_{rms} \triangleq \sqrt{\mathbb{E}[X^2]} \quad (2.35)$$

è detta valore *rms* (root mean square) o valore efficace. I momenti intorno alla media μ_X di una v.a. X , cioè le quantità:

$$\mathbb{E}[(X - \mu_X)^k]$$

sono detti momenti centrali. Il momento centrale

$$\sigma_X^2 \equiv \text{Var}[X] \triangleq \mathbb{E}[(X - \mu_X)^2] \quad (2.36)$$

è la varianza della v.a. X e la sua radice quadrata è la deviazione standard

$$\sigma_X \triangleq \sqrt{\mathbb{E}[(X - \mu_X)^2]} \quad (2.37)$$

Nel seguito, se dal contesto è chiara la v.a. a cui ci si riferisce, il pedice X sarà omesso. Si consideri ora l'effetto di una trasformazione affine sulla varianza; risulta

$$\begin{aligned} \text{Var}[aX + b] &= \mathbb{E}[(aX + b) - \mathbb{E}[aX + b]]^2 \\ &= \mathbb{E}[(aX - a\mathbb{E}[X])^2] \\ &= a^2 \sigma_X^2 \end{aligned}$$

Pertanto una traslazione non ha effetto sulla varianza, ma solo sulla media, mentre un cambiamento di scala con fattore a dà luogo ad una variazione secondo il fattore a^2 della varianza.

Momenti di una variabile aleatoria Bernoulliana.

Per una v.a. Bernoulliana, il momento di ordine k può essere calcolato come segue:

$$\mathbb{E}[X^k] = 1^k \times p + 0^k \times q = p$$

Dunque tutti i momenti di una v.a. Bernoulliana sono uguali a p . La varianza vale

$$\sigma_X^2 = p - p^2 = pq \quad (2.38)$$

come è facile verificare.

Momenti di una variabile aleatoria esponenziale.

Consideriamo una v.a. esponenziale di parametro unitario, $X_0 \sim \mathcal{E}(1)$. I suoi momenti sono dati da

$$\mathbb{E}[X_0^k] = \int_0^{+\infty} x^k e^{-x} dx$$

e si possono valutare ricorsivamente. Infatti,

$$\mathbb{E}[X_0^k] = \int_0^{+\infty} x^k e^{-x} dx = - \int_0^{+\infty} x^k \frac{d}{dx} [e^{-x}] dx$$

da cui, integrando per parti si ricava il legame di tipo ricorsivo

$$\mathbb{E}[X_0^k] = k \mathbb{E}[X_0^{k-1}]$$

Iterando tale relazione, tenendo presente che $\mathbb{E}[X_0^0] = 1$, si ricava l'espressione esplicita dei momenti della v.a. esponenziale standard

$$\mathbb{E}[X_0^k] = k! \quad (2.39)$$

Momenti centrali di una v.a. gaussiana

Ricordiamo che una v.a. gaussiana $X \sim \mathcal{N}(\mu, \sigma^2)$ è generata da una v.a. gaussiana standard mediante una trasformazione affine, si ha

$$X = \sigma X_0 + \mu \quad X_0 \sim \mathcal{N}(0, 1)$$

Conseguentemente i momenti centrali di una gaussiana sono relazionabili a quelli di una gaussiana standard come segue

$$\mathbb{E}[(X - \mu)^k] = \mathbb{E}[(\sigma X_0)^k] = \sigma^k \mathbb{E}[X_0^k]$$

Pertanto è sufficiente calcolare i momenti $\mathbb{E}[X_0^k]$:

$$\mathbb{E}[X_0^k] = \int_{-\infty}^{+\infty} x^k \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx$$

Per k dispari tali momenti sono tutti nulli in quanto l'integrando è dispari e l'intervallo di integrazione è simmetrico. Per k pari, si ottiene

$$\mathbb{E}[X_0^k] = 1 \cdot 3 \cdot 5 \cdots (k-1)!!$$

dove la notazione $(k-1)!!$ indica il prodotto degli interi dispari fino a $k-1$.

2.8 CARATTERIZZAZIONE CONGIUNTA

2.8.1 CDF congiunta di due variabili aleatorie

Per caratterizzare da un punto di vista probabilistico la coppia di v.a. X, Y non è sufficiente la conoscenza delle CDF $F_X(x)$ ed $F_Y(y)$ delle singole v.a.; invero, note tali funzioni, è possibile valutare la probabilità di eventi del tipo $\{X \leq x\}$ o del tipo $\{Y \leq y\}$. Se però siamo interessati a calcolare la probabilità dell'evento congiunto

$$\{X \leq x\} \cap \{Y \leq y\}$$

ci accorgiamo che ciò non è possibile mediante le funzioni precedenti se non nel caso in cui i due eventi siano indipendenti, nel qual caso risulta

$$P(\{X \leq x\} \cap \{Y \leq y\}) = P(\{X \leq x\}) P(\{Y \leq y\})$$

È dunque necessaria una conoscenza più approfondita dell'esperimento, esprimibile mediante la funzione di distribuzione cumulativa bidimensionale.

In analogia a quanto esposto per v.a. singole, si definisce funzione di distribuzione cumulativa congiunta delle variabili X e Y la funzione

$$F_{XY} : (x, y) \in \mathbb{R}^2 \longrightarrow F_{XY}(x, y) = P(\{X \leq x\} \cap \{Y \leq y\}) \quad (2.40)$$

Le CDF delle singole v.a. $F_X(x)$ ed $F_Y(y)$ sono dette CDF marginali. Per la CDF congiunta di una coppia di variabili aleatorie valgono proprietà simili a quelle della CDF di una variabile aleatoria.

2.8.2 PDF e PMF congiunta di due variabili aleatorie

Nel caso di v.a. discrete, in alternativa alla CDF congiunta, la caratterizzazione probabilistica può essere effettuata in termini di PMF congiunta, cioè mediante la funzione:

$$P_{XY} : (x, y) \in \mathcal{A}_X \times \mathcal{A}_Y \longrightarrow P_{XY}(x, y) = P(\{X = x\} \cap \{Y = y\}) \quad (2.41)$$

ove \mathcal{A}_X e \mathcal{A}_Y rappresentano l'alfabeto delle v.a. X e Y , rispettivamente.

Nel caso di v.a. continue, la caratterizzazione probabilistica può essere effettuata in termini di PDF congiunta delle due v.a. X e Y definita come derivata seconda mista della loro CDF congiunta, cioè

$$f_{XY}(x, y) = \frac{\partial^2 F_{XY}(x, y)}{\partial x \partial y} \quad (2.42)$$

Per la CDF, PMF e PDF congiunta di una coppia di variabili aleatorie valgono proprietà simili a quelle di una variabile aleatoria.

2.9 DISTRIBUZIONI E DENSITÀ CONDIZIONATE

Nel capitolo precedente abbiamo visto come l'utilizzo della probabilità condizionata renda spesso più semplice la formalizzazione di problemi in cui sono presenti due o più cause di aleatorietà. Nel caso di variabili aleatorie il ruolo della probabilità condizionata è svolto da CDF, PMF e PDF condizionate.

2.9.1 CDF condizionata ad un evento

La CDF di una v.a. X condizionata ad un evento B avente probabilità non nulla, congruentemente con le definizioni di probabilità condizionata e di CDF già date, è la funzione:

$$F_X(x|B) = P(\{X \leq x\}|B) = \frac{P(\{X \leq x\} \cap B)}{P(B)} ; \quad P(B) \neq 0 \quad (2.43)$$

Analogamente, nel caso che X sia una v.a. discreta, la PMF condizionata a B è la funzione

$$P_x(\cdot|B) : x \in \mathcal{A}_X \longrightarrow P_x(x|B) = \frac{P(\{X = x\} \cap B)}{P(B)} \quad (2.44)$$

mentre, nel caso che X sia una v.a. continua, la PDF condizionata è la derivata della CDF condizionata, cioè:

$$f_X(x|B) = \frac{dF_X(x|B)}{dx} \quad (2.45)$$

Esempio: Radar ad impulsi

In un sistema radar gli impulsi riflessi hanno un'ampiezza $R \sim Ray(\sigma^2)$, tuttavia sullo schermo sono visualizzati solo gli impulsi la cui ampiezza supera una certa soglia x_0 . Siamo interessati a determinare la CDF e la PDF degli impulsi visualizzati.

Iniziamo con l'osservare che la CDF e la PDF degli impulsi visualizzati sono semplicemente la CDF e la PDF dell'ampiezza R condizionate all'evento $\{R > x_0\}$. Si ha quindi:

$$\begin{aligned} F_R(x|\{R > x_0\}) &= \frac{P(\{R \leq x\} \cap \{R > x_0\})}{P(\{R > x_0\})} \\ &= \frac{F_R(x) - F_R(x_0)}{1 - F_R(x_0)} u(x - x_0). \end{aligned}$$

In particolare, se $R \sim Ray(\sigma^2)$, la sua CDF è pari a

$$F_R(x) = 1 - \exp\left[-\frac{x^2}{2\sigma^2}\right]$$

Di conseguenza, sostituendo quest'ultima espressione nella precedente si ottiene:

$$F_R(x|R > x_0) = \left(1 - \exp\left[-\frac{x^2 - x_0^2}{2\sigma^2}\right]\right) u(x - x_0)$$

Infine, per quanto concerne la PDF degli impulsi visualizzati è sufficiente effettuare la derivata della CDF condizionata; si ha pertanto:

$$f_R(x|\{R > x_0\}) = \frac{x}{\sigma^2} \exp\left[-\frac{x^2 - x_0^2}{2\sigma^2}\right] u(x - x_0)$$

2.9.2 CDF di X condizionata all'evento $\{Y \leq y\}$

Particularizzando la definizione (2.43) al caso in esame, si ha che la CDF della v.a. X condizionata all'evento $\{Y \leq y\}$ è la funzione:

$$F_X(x|\{Y \leq y\}) = \frac{P(\{X \leq x\} \cap \{Y \leq y\})}{P(\{Y \leq y\})} = \frac{F_{XY}(x,y)}{F_Y(y)}$$

e la corrispondente densità di probabilità condizionata, supposta esistente, si ottiene derivando la CDF rispetto alla variabile x

$$f_X(x|\{Y \leq y\}) = \frac{1}{F_Y(y)} \frac{\partial F_{XY}(x,y)}{\partial x}$$

2.9.3 PMF condizionata

Nel caso che X ed Y siano v.a. discrete, la PMF della v.a. X condizionata all'evento $\{Y = y\}$, supposto a probabilità non nulla si ottiene particularizzando la definizione generale al caso in esame

$$\begin{aligned} P_x(x|\{Y = y\}) &= P(\{X = x\}|\{Y = y\}) \\ &= \frac{P(\{X = x\} \cap \{Y = y\})}{P(\{Y = y\})} = \frac{P_{XY}(x,y)}{P_Y(y)} \end{aligned}$$

Si noti che nel caso in esame la notazione più comune è $P_{x|y}(x|y)$ invece di $P_x(x|\{Y = y\})$; inoltre si parla semplicemente di PMF condizionata, senza precisare che l'evento condizionante è $\{Y = y\}$. In definitiva, la PMF condizionata della v.a. X è la funzione:

$$P_{x|y}(x|y) : x \in \mathcal{A}_X \longrightarrow \frac{P_{xy}(x,y)}{P_Y(y)} \quad P_Y(y) \neq 0 \quad (2.46)$$

La relazione precedente può anche essere assunta come definizione di PMF condizionata in quanto, come è facile verificare la funzione definita dalla (2.46) gode delle proprietà caratterizzanti una PMF.

2.9.4 PDF condizionata

Nel caso di v.a. continue, in analogia a quanto esposto per le v.a. discrete, si definisce la PDF condizionata della v.a. X data la v.a. Y la funzione

$$f_{X|Y}(x|y) : x \in \mathbb{R} \longrightarrow \frac{f_{XY}(x,y)}{f_Y(y)} \quad f_Y(y) \neq 0 \quad (2.47)$$

2.9.5 Leggi della probabilità per la PDF

Dalla definizione di PDF condizionata segue la *legge della probabilità composta* per le PDF, cioè

$$f_{XY}(x, y) = f_Y(y)f_{X|Y}(x|y) = f_X(x)f_{Y|X}(y|x) \quad (2.48)$$

conseguentemente risulta:

$$f_{X|Y}(x|y) = f_{Y|X}(y|x) \frac{f_X(x)}{f_Y(y)} \quad (2.49)$$

che esprime la **legge di Bayes** per v.a. Infine, integrando in dy entrambi i membri della (2.48) si ottiene:

$$f_X(x) = \int_{-\infty}^{\infty} f_{X|Y}(x|y)f_Y(y)dy \quad (2.50)$$

che esprime la **legge della probabilità totale** per le PDF. Definizione e proprietà della PDF congiunta di due v.a. sono riassunte nella tabella 2.12.

Definizione	$f_{X Y}(x y) : \quad x \in \mathbb{R} \quad \longrightarrow \frac{f_{XY}(x, y)}{f_Y(y)} \quad f_Y(y) \neq 0$
Legge probabilità composta	$f_{XY}(x, y) = f_Y(y)f_{X Y}(x y) = f_X(x)f_{Y X}(y x)$
Legge di Bayes	$f_{X Y}(x y) = f_{Y X}(y x) \frac{f_X(x)}{f_Y(y)}$
Legge probabilità totale	$f_X(x) = \int_{-\infty}^{\infty} f_{X Y}(x y)f_Y(y)dy$

Tabella 2.12

Proprietà della PDF condizionata di due variabili aleatorie

2.10 VARIABILI ALEATORIE INDEPENDENTI

L'indipendenza tra n v.a. si ha quando la loro CDF congiunta è il prodotto delle singole CDF marginali, se cioè:

$$F_{X_1 \dots X_n}(x_1, x_2, \dots, x_n) = F_{X_1}(x_1)F_{X_2}(x_2) \cdots F_{X_n}(x_n)$$

Equivalentemente, supposto le v.a. continue, derivando, si ricava che per v.a. indipendenti la PDF congiunta è il prodotto delle PDF marginali, cioè:

$$f_{X_1 \dots X_n}(x_1, x_2, \dots, x_n) = f_{X_1}(x_1)f_{X_2}(x_2) \cdots f_{X_n}(x_n) \quad (2.51)$$

Si noti che la relazione precedente è valida anche per v.a. discrete, a patto di sostituire le PMF alle PDF.

2.11 MOMENTI CONGIUNTI DI DUE VARIABILI ALEATORIE

I momenti, in particolare media e deviazione standard, forniscono una caratterizzazione globale di una variabile aleatoria. Nel caso in cui si considerano due o più variabili aleatorie è utile definire altri momenti che indicano il grado di influenza reciproca tra le variabili aleatorie. Considerate due v.a. X ed Y , si definisce momento congiunto di ordine $k = m + r$ la media

$$\mathbb{E}[X^m Y^r]$$

Particolare importanza rivestono i momenti centrali e non centrali del primo e del secondo ordine, corrispondenti alle coppie di indici $\{m, r\} = (10, 01, 20, 02, 11)$. I momenti del primo ordine: $\mu_X = \mathbb{E}[X]$ e $\mu_Y = \mathbb{E}[Y]$ sono le medie delle singole variabili aleatorie mentre i momenti del secondo ordine sono i valori quadratici medi delle singole variabili aleatorie e la correlazione tra le due variabili

$$r_{XY} \equiv \text{corr}[X, Y] \triangleq \mathbb{E}[XY]$$

Analogamente, se si considerano i momenti centrali del secondo ordine, si riottengono le varianze della singole v.a. e la loro covarianza

$$c_{XY} \equiv \text{cov}[X, Y] \triangleq \mathbb{E}[(X - \mu_X)(Y - \mu_Y)]$$

Sviluppando il prodotto che compare nella definizione di covarianza si ha:

$$\begin{aligned} \mathbb{E}[(X - \mu_X)(Y - \mu_Y)] &= \mathbb{E}[XY - X\mu_Y - Y\mu_X + \mu_X\mu_Y] \\ &= \mathbb{E}[XY] - \mathbb{E}[X]\mu_Y - \mathbb{E}[Y]\mu_X + \mu_X\mu_Y \\ &= \mathbb{E}[XY] - \mu_X\mu_Y, \end{aligned}$$

pertanto correlazione e covarianza sono legate tra loro dalla relazione:

$$r_{XY} = c_{XY} + \mu_X\mu_Y \quad (2.52)$$

2.11.1 Incorrelazione e indipendenza

Due variabili aleatorie si dicono incorrelate se la loro covarianza è nulla; ma, dalla (2.52) segue che, equivalentemente, due v.a. sono incorrelate se la correlazione è pari al prodotto delle loro medie. Nel caso in cui la correlazione sia nulla, le variabili si dicono ortogonali. Conseguentemente, se almeno una delle due v.a. incorrelate ha media nulla la correlazione è zero e le variabili, oltre che incorrelate, sono anche ortogonali, il che costituisce il caso

più comune di v.a. ortogonali. L'incorrelazione è una condizione nettamente più debole dell'indipendenza statistica e non va confusa con quest'ultima, nel senso che sussiste la seguente implicazione.

Proposizione

Date due v.a. X ed Y , statisticamente indipendenti esse sono anche incorrelate.

Prova: supponiamo, per fissare le idee che le due v.a. siano di tipo continuo. In tal caso, essendo statisticamente indipendenti, la loro PDF congiunta è il prodotto delle due PDF marginali; conseguentemente risulta:

$$\begin{aligned}\mathbb{E}[XY] &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} xy f_{XY}(x,y) dx dy = \\ &= \int_{-\infty}^{+\infty} xf_X(x) dx \int_{-\infty}^{+\infty} yf_Y(y) dy = \mathbb{E}[X]\mathbb{E}[Y]\end{aligned}$$

il che prova l'asserto.

Esempio: variabili aleatorie incorrelate, ma dipendenti

Si consideri la coppia di v.a.

$$\begin{cases} X = \cos \Theta \\ Y = \sin \Theta \end{cases} \quad \Theta \sim U(-\pi, \pi)$$

avendosi:

$$\begin{aligned}\mathbb{E}[X] &= \mathbb{E}[\cos \Theta] = \int_{-\pi}^{+\pi} \frac{1}{2\pi} \cos \alpha d\alpha = 0 \\ \mathbb{E}[Y] &= \mathbb{E}[\sin \Theta] = \int_{-\pi}^{+\pi} \frac{1}{2\pi} \sin \alpha d\alpha = 0\end{aligned}$$

le v.a. sono entrambe a media nulla; inoltre, poiché

$$\mathbb{E}[XY] = \mathbb{E}[\cos \Theta \sin \Theta] = \int_{-\pi}^{+\pi} \frac{1}{2\pi} \cos \alpha \sin \alpha d\alpha = 0$$

le variabili sono anche ortogonali e incorrelate. Tuttavia esse non sono statisticamente indipendenti in quanto i valori della X e della Y sono, a meno del segno, funzionalmente legate, con legame funzionale implicitamente definito dall'equazione:

$$X^2 + Y^2 = 1$$

In altri termini le due v.a. assumono, con probabilità 1, valori sulla circonferenza del cerchio di raggio unitario e centro l'origine.

Esempio: coppie di v.a. congiuntamente Gaussiane

Una coppia di v.a. gaussiane si ottiene da una trasformazione affine non singolare di v.a. gaussiane standard indipendenti, cioè

$$\begin{aligned} X_1 &= a_{11}X_{01} + a_{12}X_{02} + \mu_1 \\ X_2 &= a_{21}X_{01} + a_{22}X_{02} + \mu_2 \end{aligned}$$

con X_{01}, X_{02} v.a. gaussiane standard indipendenti. La PDF congiunta ha la forma generale

$$f_{\mathbf{x}}(\mathbf{x}) = \frac{1}{(2\pi)^{N/2}|\mathbf{C}|^{1/2}} \exp \left[-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T \mathbf{C}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right] \quad (2.53)$$

dove la matrice \mathbf{C} è detta matrice di covarianza e $\boldsymbol{\mu} = (\mu_1, \mu_2)^T$ è il vettore delle medie. Il generico termine a_{ij} della matrice di covarianza è la covarianza mutua

$$a_{ij} = \mathbb{E}[(X_i - \mu_i)(X_j - \mu_j)]$$

per cui si ha

$$\mathbf{C} = \begin{bmatrix} \mathbb{E}[(X_1 - \mu_1)^2] & \mathbb{E}[(X_1 - \mu_1)(X_2 - \mu_2)] \\ \mathbb{E}[(X_1 - \mu_1)(X_2 - \mu_2)] & \mathbb{E}[(X_2 - \mu_2)^2] \end{bmatrix}$$

il che spiega la denominazione matrice di covarianza data a \mathbf{C} . I termini sulla diagonale principale coincidono con le varianze delle singole v.a. mentre i termini sulla diagonale secondaria possono essere espressi attraverso il coefficiente di correlazione, definito come

$$\rho = \frac{\mathbb{E}[(X_1 - \mu_1)(X_2 - \mu_2)]}{\sigma_1 \sigma_2} \quad (2.54)$$

Calcolando il determinante e l'inversa della matrice \mathbf{C} , che, come è facile verificare, valgono

$$\det(\mathbf{C}) = \sigma_1 \sigma_2 (1 - \rho^2) \quad \mathbf{C}^{-1} = \frac{1}{1 - \rho^2} \begin{bmatrix} \frac{1}{\sigma_1^2} & \frac{\rho}{\sigma_1 \sigma_2} \\ \frac{\rho}{\sigma_1 \sigma_2} & \frac{1}{\sigma_2^2} \end{bmatrix}$$

è possibile esplicitare la forma generale (2.53) della PDF congiunta in termini di medie, varianze e coefficiente di correlazione. Infatti sostituendo l'espressione determinata della \mathbf{C}^{-1}

nella (2.53) si ottiene:

$$f_{x_1 x_2}(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2(1-\rho^2)^{1/2}} \exp \left[-\frac{1}{2(1-\rho^2)} \times \right. \\ \left. \frac{(x_1 - \mu_1)^2}{\sigma_1^2} - 2\rho \frac{(x_1 - \mu_1)(x_2 - \mu_2)}{\sigma_1\sigma_2} + \frac{(x_2 - \mu_2)^2}{\sigma_2^2} \right] \quad (2.55)$$

Si noti infine che, ponendo $\rho = 0$ nell'espressione della PDF congiunta di due v.a. gaussiane, si ottiene

$$f_{x_1 x_2}(x_1, x_2) = \frac{1}{\sqrt{2\pi\sigma_1^2}} \exp \left[-\frac{1}{2} \frac{(x_1 - \mu_1)^2}{\sigma_1^2} \right] \times \\ \frac{1}{\sqrt{2\pi\sigma_2^2}} \exp \left[-\frac{1}{2} \frac{(x_2 - \mu_2)^2}{\sigma_2^2} \right] \quad (2.56)$$

pertanto, se due v.a. congiuntamente gaussiane sono incorrelate sono anche indipendenti. In figura 2.25 è mostrata la PDF congiunta di due v.a. gaussiane con varianze unitarie e coefficiente di correlazione $\rho = 0.5$ e $\rho = 0.95$. È immediato notare che quando ρ si tende ad 1 la superficie si concentra intorno alla retta $y = x$; in tal caso, infatti, come evidenziato dalla (2.54), deve risultare $\mathbb{E}[(X_1 - \mu_1)(X_2 - \mu_2)] = \sigma_1\sigma_2$ che si ottiene se le variabili $(X_1 - \mu_1)$ e $(X_2 - \mu_2)$ sono proporzionali. Infatti, ponendo $(X_1 - \mu_1) = \alpha(X_2 - \mu_2)$ si ricava

$$\rho = \frac{\mathbb{E}[\alpha(X_2 - \mu_2)(X_2 - \mu_2)]}{\alpha\sigma_2^2} = 1$$

La figura 2.26 mostra le curve di livello della pdf bidimensionale, nello stesso caso di fig. 2.25. Le curve rappresentano delle ellissi che si ottengono ponendo

$$(\mathbf{x} - \boldsymbol{\mu})^T \mathbf{C}^{-1} (\mathbf{x} - \boldsymbol{\mu}) = cost.$$

2.12 STIME BASATE SU SEQUENZE DI OSSERVAZIONI

Siamo abituati ad osservare che l'accuratezza con cui si misurano grandezze non deterministiche dipende dal numero di osservazioni disponibili. La teoria della stima, di cui si presenteranno nel seguito alcuni elementi, si occupa di studiare quantitativamente tale affermazione.

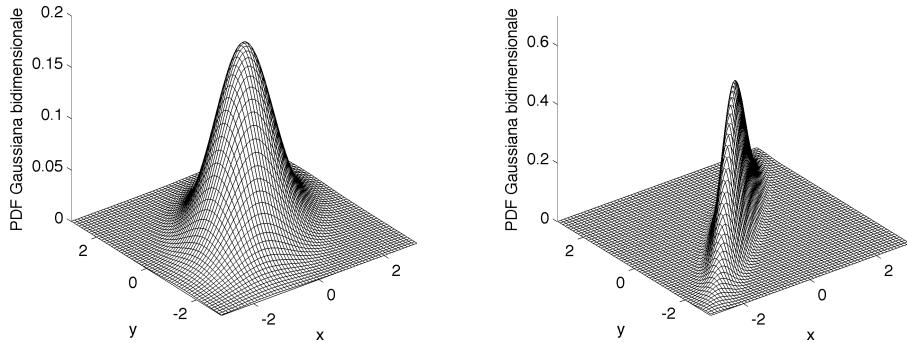


Figura 2.25: Funzione di densità di probabilità gaussiana bidimensionale, $\sigma_X = \sigma_Y = 1$, $\rho = 0.5$, $\rho = 0.95$.

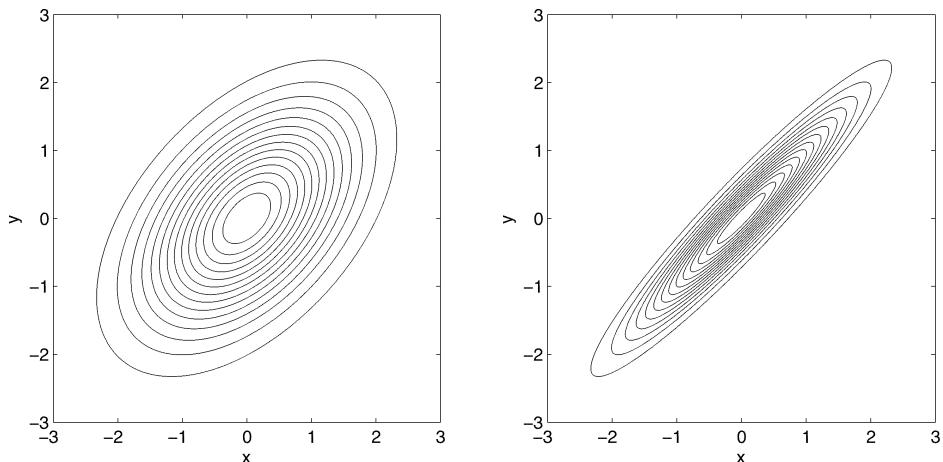


Figura 2.26: Curve di livello della funzione di densità di probabilità gaussiana bidimensionale, $\sigma_X = \sigma_Y = 1$, $\rho = 0.5$, $\rho = 0.95$.

2.12.1 Media campionaria

Iniziamo con lo studiare il comportamento asintotico della media aritmetica di n v.a., detta anche media campionaria, definita come

$$S_n \triangleq \frac{1}{n} \sum_{i=1}^n X_i \quad (2.57)$$

La media campionaria è una variabile aleatoria la cui media statistica si ricava semplicemente utilizzando la proprietà di linearità; si ha infatti:

$$\mathbb{E}[S_n] = \mathbb{E}\left[\frac{1}{n} \sum_{i=1}^n X_i\right] = \frac{1}{n} \sum_{i=1}^n \mathbb{E}[X_i] = \mu$$

dove μ è la media statistica delle variabili aleatorie X_i . Sinteticamente, si dice che S_n è uno stimatore della media statistica μ . Inoltre, risultando $\mathbb{E}[S_n] = \mu$ la media dello stimatore coincide con la grandezza da stimare, per cui lo stimatore è *non polarizzato*. Considerando la varianza dello stimatore, nel caso di stimatore non-polarizzato e variabili aleatorie indipendenti ed identicamente distribuite (i.i.d.), si ha

$$\begin{aligned} \text{Var}[S_n] &= E\left[\left(\frac{1}{n} \sum_{i=1}^n X_i - \frac{1}{n} \sum_{i=1}^n \mathbb{E}[X_i]\right)^2\right] = \\ &= \mathbb{E}\left[\left(\frac{1}{n} \sum_{i=1}^n (X_i - \mathbb{E}[X_i])\right)^2\right] = \\ &= \frac{1}{n^2} \mathbb{E}\left[\sum_{i=1}^n \sum_{j=1}^n (X_i - \mathbb{E}[X_i])(X_j - \mathbb{E}[X_j])\right] = \\ &= \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \mathbb{E}[(X_i - \mathbb{E}[X_i])(X_j - \mathbb{E}[X_j])] \end{aligned}$$

dove si può riconoscere che il generico addendo dell'ultima sommatoria è la covarianza delle variabili aleatorie X_i ed X_j . In definitiva si ottiene

$$\text{Var}[S_n] = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \text{Cov}[X_i, X_j] \quad (2.58)$$

e, nel caso in cui le variabili aleatorie siano indipendenti, risulta

$$\text{Cov}[X_i, X_j] = \begin{cases} \text{Var}[X_i] = \sigma_{X_i}^2 & \text{se } i = j \\ 0 & \text{se } i \neq j \end{cases}$$

per cui la varianza della somma di variabili aleatorie incorrelate è data dalla somma delle varianze delle singole variabili aleatorie,

$$\text{Var}[S_n] = \frac{1}{n^2} \sum_{i=1}^n \sigma_{X_i}^2$$

Se infine le variabili, oltre che incorrelate, hanno tutte la stessa varianza, se cioè

$$\sigma_{X_i}^2 = \sigma_X^2, \quad \forall i \in \{1, 2, \dots, n\}$$

si ottiene

$$\text{Var}[S_n] = \frac{\sigma^2}{n}$$

Quanto stabilito per la media campionaria è riassunto dal seguente asserto:

Proposizione

Date n variabili aleatorie i.i.d., la media campionaria è uno stimatore non polarizzato della media statistica delle variabili e la sua varianza è uguale alla varianza delle variabili scalata per un fattore $1/n$, cioè

$$\mathbb{E}[S_n] = \mu \quad \text{Var}[S_n] = \frac{\sigma^2}{n} \quad (2.59)$$

2.12.2 Legge dei grandi numeri

Le espressioni ricavate per media e varianza della media campionaria di n variabili aleatorie i.i.d. evidenziano che l'errore m.s. che si commette nell'approssimare la S_n con la sua media μ , decresce all'aumentare del numero degli addendi. Tale affermazione costituisce la *Legge dei grandi numeri*, che stabilisce che la media campionaria di una successione di variabili aleatorie i.i.d. converge asintoticamente alla loro media statistica. In particolare, poichè la varianza dello stimatore tende asintoticamente a zero, la convergenza va intesa in *media quadratica*,

$$\lim_n \mathbb{E}[|S_n - \mu|^2] = 0 \quad (2.60)$$

Si noti esplicitamente che la convergenza in media quadratica non comporta che vadano a zero le possibili determinazioni dello scarto assoluto $|S_n - \mu|$ ma ciò che tende a zero è solo lo scarto quadratico medio. Alla luce di tale definizione, e sulla scorta di quanto stabilito, resta dimostrato il seguente asserto.

Legge dei grandi numeri

data una successione di variabili aleatorie iid X_n , $n \in \mathbb{N}$ la sua media campionaria S_n converge, in media quadratica, al valore comune μ della media statistica delle singole variabili aleatorie

$$\lim_n \frac{1}{n} \sum_{i=1}^n X(i) = \mu \quad \text{m.s.}$$

Notiamo esplicitamente che tale formulazione è valida anche se le *v.a.* non sono indipendenti, ma semplicemente incorrelate, avendo però tutte la stessa media e la stessa varianza. La legge dei grandi numeri sostanzia, da un punto di vista analitico, l'affermazione che la media statistica coincide con la media aritmetica di un numero molto grande di osservazioni. In termini più corretti si può affermare che pur essendo la media campionaria a sua volta una variabile aleatoria, al crescere del numero delle osservazioni essa tende a diventare una quantità deterministica. Inoltre, l'errore quadratico medio che si commette nell'approssimare la media statistica μ con la media campionaria è inversamente proporzionale al numero n di osservazioni.

La legge dei grandi numeri è lo strumento base per la misura di grandezze probabilistiche: probabilità, medie statistiche, CDF, *pmf*, etc., va però riformulata per adattarla alla grandezza d'interesse: precisamente va individuata una successione di variabili aleatorie la cui media statistica coincida con la grandezza che si intende stimare.

Esempio: stima della probabilità di un evento

Definendo l'indicatore dell'evento

$$I_A = \begin{cases} 1 & \text{se l'evento } A \text{ risulta verificato} \\ 0 & \text{altrimenti} \end{cases}$$

si ottiene $\mathbb{E}[I_A] = 1 \times P(A) + 0 \times (1 - P(A)) = P(A)$ per cui, dalla legge dei grandi numeri, risulta che la media campionaria dell'indicatore tende alla probabilità dell'evento al divergere del numero di osservazioni. D'altra parte, la media campionaria dell'indicatore coincide con il numero di volte N_A in cui l'evento A si verifica; da cui si ottiene il seguente asserto.

Convergenza della frequenza relativa

La frequenza relativa di un evento A in n prove indipendenti, effettuate tutte nelle medesime condizioni, converge, in media quadratica, alla probabilità $P(A)$ dell'evento A :

$$\lim_n \frac{N_A(n)}{n} = P(A) \quad \text{m.s.}$$

In pratica non si opera al limite, ma con un numero finito, ancorché grande di osservazioni; tuttavia è possibile valutare il numero di osservazioni necessarie per misurare la $P(A)$ con un prefissato errore *rms* ovvero di calcolare l'errore *rms* che si commette effettuando la misura con un prefissato numero di osservazioni, come illustrato dal seguente esempio:

Esempio: misura della Probabilità di un evento

L'utilizzo della variabile aleatoria I_A – indicatrice dell'evento A – ha consentito di tradurre il problema della stima di $P(A)$ nel calcolo di una media campionaria ed, in ultima analisi, in

un problema di conteggio. La varianza dello stimatore è semplicemente data da

$$\text{Var}(I_A) = \frac{1}{n} \text{Var}(X_A)$$

dove la variabile aleatoria X_A è bernoulliana, con probabilità di successo uguale a $P(A)$. La varianza di X_A è data da $P(A)(1 - P(A))$ da cui si ottiene che la deviazione standard della stima è data da

$$\sigma = \sqrt{\frac{P(A)[1 - P(A)]}{n}} \leq \frac{1}{2\sqrt{n}}$$

ove la maggiorazione si giustifica immediatamente se si tiene presente che

$$\max_{0 \leq P(A) \leq 1} P(A)[1 - P(A)] = \frac{1}{4} \quad (2.61)$$

Si noti che, a differenza di σ , la sua maggiorazione è indipendente dal valore della $P(A)$ e coincide con σ per $P(A) = 0.5$. Valutiamo ora l'errore medio che si commette nella misura della $P(A)$ a partire da $n = 1000$ osservazioni i.i.d. ed il numero di osservazioni necessarie per avere un errore medio inferiore a 10^{-2} . Effettuando la stima su 1000 dati, si ha

$$\sigma \leq \frac{1}{2\sqrt{n}} = \frac{1}{2\sqrt{1000}} = 0.016$$

Volendo calcolare il numero di dati necessari al fine di ottendere un errore medio minore di 10^{-2} , si ha

$$\frac{1}{2\sqrt{n}} \approx 10^{-2} \implies n \approx 2500$$

In realtà è sufficiente un numero inferiore di dati in quanto si è utilizzata una maggiorazione dell'errore medio, ma il calcolo preciso richiede la conoscenza della probabilità che si vuole stimare.

2.13 ESERCIZI PROPOSTI

Ex. 2.1 Si determini lo spazio dei campioni relativo al doppio lancio di un dado. Si individuino inoltre i seguenti eventi:

$$A \equiv \{"la somma dei due lanci è 5"\}$$

$$B \equiv \{"primo lancio pari"\}$$

$$C \equiv \{"primo lancio pari e la somma dei due lanci è 5"\}$$

$$D \equiv \{"primo lancio pari oppure la somma dei due lanci è 5"\}$$

Ex. 2.2 Si consideri un canale binario, cioè un mezzo di trasmissione che accetta in ingresso simboli “0” e “1” e fornisce in uscita simboli “0” e “1”. A causa dei disturbi presenti sul canale, però, non sempre il simbolo in uscita è uguale a quello che viene effettivamente trasmesso, è cioè possibile che si verifichi un errore di trasmissione. Per tale motivo, il funzionamento del canale binario può essere correttamente descritto come un esperimento aleatorio. Si individuino lo spazio dei campioni relativo alla trasmissione di un solo simbolo e gli eventi:

$$A \equiv \{ \text{"si è verificato un errore nella trasmissione del simbolo 1"} \}$$

$$B \equiv \{ \text{"si è verificato un errore di trasmissione"} \}$$

$$C \equiv \{ \text{"si è ricevuto il simbolo 0"} \}$$

$$D \equiv \{ \text{"si è trasmesso il simbolo 1"} \}$$

Ex. 2.3 (Paradosso di De Mèrè) All'inizio del 1600 De Mèrè, si trovò di fronte al seguente quesito: supponiamo di disporre di dadi non truccati: è più facile realizzare un 6 su quattro lanci consecutivi oppure un doppio 6 in 24 lanci?

De Mèrè risolse il problema nel seguente modo: se il dado non è truccato, possiamo dire che le possibilità di avere un 6 sono 1 su 6. Se lanciamo 4 volte 1 dado, queste possibilità dovrebbero essere $4 * 1/6 = 2/3$. Se lanciamo due dadi possiamo avere 36 risultati possibili, cioè tutti gli accoppiamenti tra i valori di una faccia di un dado (che sono 6) e i valori della faccia dell'altro dado (che sono sempre 6), quindi la possibilità di avere un doppio 6 ad ogni lancio la possiamo porre uguale a 1/36. Se effettuiamo 24 lanci, la possibilità diventa $24 * 1/36 = 24/36 = 2/3$. Così ragionando la risposta è che i due giochi di dadi ci danno le stesse probabilità di vittoria. De Mèrè, accanito giocatore, si accorse che uno dei due giochi era meno favorevole dell'altro. Il sospetto di aver fatto degli errori lo indusse a scrivere a Pierre De Fermat il quale, in una serie di scambi epistolari con Blaise Pascal, giunse alla soluzione: il doppio 6 su 24 lanci è un evento più difficile a realizzarsi di un singolo 6 su 4 lanci. Giustificare tale affermazione.

Ex. 2.4 Un esperimento aleatorio consiste nell'estrarre tre carte da un mazzo di carte napoletane. Considerato l'evento $E = \{\text{le carte estratte sono tutte di coppe}\}$, stabilire in quali delle seguenti estrazioni si verifica l'evento \bar{E} :

$$\omega_1 \equiv (\text{2 di bastoni, 5 di coppe, asso di denari})$$

$$\omega_2 \equiv (\text{asso di spade, 7 di denari, 10 di spade})$$

$$\omega_3 \equiv (\text{asso di spade, 6 di coppe, 10 di spade})$$

$$\omega_4 \equiv (\text{asso di coppe, 2 di coppe, 3 di coppe})$$

Ex. 2.5 Si considerino i quattro gruppi sanguigni 0, A, B e AB. Una qualsiasi persona può ricevere il sangue da un donatore del suo stesso gruppo ed anche da un donatore del gruppo 0, mentre una persona del gruppo AB può ricevere il sangue da uno qualsiasi dei quattro gruppi. Un esperimento aleatorio consiste nel determinare il gruppo sanguigno per ogni coppia di donatori che arriva in una banca del sangue.

1. Si determini lo spazio dei campioni relativo a quest'esperimento.
2. Si individuino i seguenti eventi:

$$C \equiv \{ \text{"il secondo donatore può ricevere il sangue dal primo"} \}$$

$$D \equiv \{ \text{"ciascuno dei due donatori può ricevere il sangue dall'altro"} \}$$

$$E \equiv \{ \text{"nessuno dei due donatori può ricevere il sangue dall'altro"} \}$$

3. Nell'ipotesi che gli eventi elementari siano equiprobabili, si calcolino le probabilità dei tre eventi individuati.

Ex. 2.6 Un esperimento aleatorio consiste nel provare il funzionamento di una lampadina osservando la sua durata. Definire il corrispondente spazio dei campioni, specificando se è di tipo continuo o di tipo discreto. Si individuino inoltre i seguenti eventi:

$$A \equiv \{ \text{"la durata della lampada è superiore a 10"} \}$$

$$B \equiv \{ \text{"la durata della lampada è non inferiore a 5"} \}$$

$$C \equiv \{ \text{"la durata della lampada è compresa tra 5 e 30"} \}$$

Ex. 2.7 Una moneta non truccata viene lanciata quattro volte. Si determini lo spazio dei campioni e si calcoli la probabilità dei seguenti eventi:

$$A \equiv \{ \text{"si ottengono esattamente tre teste"} \}$$

$$B \equiv \{ \text{"si ottiene almeno una testa"} \}$$

$$C \equiv \{ \text{"il numero di teste è uguale al numero di croci"} \}$$

$$D \equiv \{ \text{"il numero di teste è maggiore del numero di croci"} \}$$

Ex. 2.8 Dati due eventi A e B si mostri che:

$$P(A \cup B) = 1 - P(\overline{A} \cap \overline{B}).$$

Ex. 2.9 Calcolare le seguenti probabilità:

a. $P(\overline{B})$;

b. $P(A \cup \overline{B})$;

c. $P(B - A)$;

d. $P(B \cap \overline{A})$;

e. $P(\overline{A} \cup \overline{B})$.

sapendo che: $P(A) = P(B) = \frac{1}{3}$ $P(A \cap B) = \frac{1}{10}$.

Ex. 2.10 Si dimostrino le proprietà **P1-P5** della probabilità.

Ex. 2.11 Siano A e B due eventi a probabilità non nulla. Dimostrare che se sono indipendenti non sono mutuamente esclusivi e viceversa.

Ex. 2.12 Relativamente all'esperimento del lancio di una coppia di dadi non truccati, si considerino gli eventi

$$\begin{aligned} A &\equiv \{ \text{"facce uguali"} \} \\ B &\equiv \{ \text{"la somma è compresa fra 7 e 10 (inclusi)"} \} \\ C &\equiv \{ \text{"la somma è 2, 7 o 8"} \} \end{aligned}$$

Si verifichi se i tre eventi sono indipendenti.

Ex. 2.13 Due giocatori, A e B, lanciano alternativamente una moneta e vince chi per primo ottiene testa. Si assuma che A inizi il gioco. Determinare la probabilità che vinca A supposta la moneta ben bilanciata. Ripetere il calcolo per una moneta arbitraria.

Ex. 2.14 Siano A e B due eventi. Dimostrare che se $P(A|B) = P(A|\bar{B})$ allora A e B sono statisticamente indipendenti.

Ex. 2.15 Uno studente può sostenere l'esame di Economia con uguale probabilità con i professori A, B e C, i quali bocciano con probabilità 0.1, 0.3 e 0.2 rispettivamente. Sapendo che uno studente è stato bocciato, qual è la probabilità che abbia sostenuto l'esame con A?

Ex. 2.16 Un giocatore disonesto trucca un dado in modo da ottenere il numero 6 in un lancio con probabilità $2/3$ e un qualsiasi altro risultato con probabilità $1/15$. Sfortunatamente (per lui) al momento del lancio il dado truccato si trova mescolato con altri due dadi non truccati. Il giocatore sceglie un dado a caso, lo lancia, e ottiene 6. Valutare la probabilità che sia stato lanciato il dado truccato. Ripetere il calcolo sapendo che, lanciato una seconda volta lo stesso dado, si è ottenuto ancora 6.

Ex. 2.17 Sia X una v.a. tipo Laplace, cioè con PDF

$$f_X(x) = Ae^{-\lambda|x|}$$

Allo scopo di limitarne la dinamica, tale variabile viene cimata al valore $\pm V$ con la trasformazione

$$Y = \begin{cases} +V & \text{se } X \geq V \\ X & \text{se } X \in (-V, V) \\ -V & \text{se } X \leq -V \end{cases}$$

Determinare il livello di cimatura V in modo che il valore rms dell'errore $\epsilon = X - Y$ sia un centesimo del valore rms di X . Stabilire inoltre la frazione di tempo durante la quale si ha cimatura.

Ex. 2.18 Siano X e Y due v.a. a media nulla, identicamente distribuite, ma correlate con coefficiente (indice) di correlazione ρ . Calcolare valor medio, valore quadratico medio e varianza delle due v.a.

$$\begin{aligned} Z &= X \\ V &= Y - \rho X \end{aligned}$$

Dimostrare inoltre che tali v.a. sono incorrelate.

Ex. 2.19 Si consideri un canale di trasmissione con ingresso X ed uscita $Y = X + D_1 + D_2$. Nell'ipotesi che le v.a. X, D_1, D_2 siano gaussiane indipendenti e precisamente si abbia:

$$X \sim \mathcal{N}(\mu_X, \sigma_X^2), \quad D_1 \sim \mathcal{N}(0, \sigma_D^2) \quad D_2 \sim \mathcal{N}(0, \sigma_D^2)$$

calcolare media, valore ms e varianza dell'ingresso e dell'uscita del canale; determinare inoltre il coefficiente di correlazione tra l'ingresso e l'uscita. Stabilire infine che cambia se, a parità di medie e varianze, le v.a. non sono gaussiane, ma hanno PDF arbitraria.

Ex. 2.20 Si consideri un canale di trasmissione con ingresso $X \sim \mathcal{B}(1, 1/2)$ ed uscita Y , variabile aleatoria discreta, con alfabeto $\mathcal{A}_Y = \{-1, 0, 1\}$ e PMF condizionali riportate in tabella.

y	$f_{Y X}(y 0)$	$f_{Y X}(y 1)$
1	0	$1 - \epsilon$
0	ϵ	ϵ
-1	$1 - \epsilon$	0

Calcolare $E[Y]$, $E[Y^2]$, e σ_Y^2 ; determinare inoltre il coefficiente di correlazione tra l'ingresso e l'uscita del canale.

Ex. 2.21 Siano X_1 e X_2 due variabili aleatorie gaussiane indipendenti, $X_1 \sim \mathcal{N}(1, 4)$ e $X_2 \sim \mathcal{N}(0.5, 0.36)$. Date le due variabili aleatorie $Y_1 = X_1$ e $Y_2 = X_1 + X_2$ si determinino medie e varianze di Y_1 e Y_2 ed il loro coefficiente di correlazione .

Ex. 2.22 Date le variabili aleatorie $X \sim \mathcal{N}(1, 0.16)$ e $Y \sim \mathcal{N}(2, 0.25)$, congiuntamente gaussiane con coefficiente di correlazione $\rho_{XY} = 0.8$, calcolare la $P(\{Z > 4\})$, ove $Z = X + Y$.

N.B. Se X e Y sono congiuntamente Gaussiane la loro somma è gaussiana.

Ex. 2.23 In un negozio di materiale elettrico è possibile trovare cavi la cui lunghezza può essere schematizzata come una variabile aleatoria uniforme nell'intervallo $[0.9, 1.1]$ metri, $[0.95, 1.05]$ metri e $[0.98, 1.02]$ metri, a seconda che essi siano stati prodotti negli stabilimenti A,B e C rispettivamente. Sapendo che un cavo proviene da A con probabilità 0.5, da B con probabilità 0.2 e da C con probabilità 0.3, calcolare:

1. la lunghezza media di un cavo;
2. la probabilità che la lunghezza di un cavo prodotto nello stabilimento A si discosti dalla media per più di 1 cm in valore assoluto;
3. la probabilità che un cavo sia stato prodotto nello stabilimento A, sapendo che la sua lunghezza si discosta dalla media per meno di 1 cm in valore assoluto.

Determinare inoltre come varia la lunghezza media se, a causa di un malfunzionamento, lo stabilimento C comincia a produrre cavi con lunghezza uniforme nell'intervallo $[1.48, 1.52]$.

Ex. 2.24 Il signor Caio vuole regalare un braccialetto alla sua fidanzata, ed entra quindi in una gioielleria, dove il commesso gli mostra un gran numero di braccialetti il cui prezzo in euro è una

v.a. X con PDF

$$f_X(x) = \frac{1}{100} \Lambda \left[\frac{x - 200}{408} \right].$$

ove $\Lambda(x)$ è la funzione:

$$\Lambda(x) = \begin{cases} 1 - |x| & se \quad |x| \leq 1 \\ 0 & se \quad |x| > 1 \end{cases}$$

Il signor Caio spiega al commesso che intende spendere tra 150 e 500 euro, per cui quest'ultimo seleziona il sottoinsieme desiderato di braccialetti tra i quali il signor Caio effettua la sua scelta.

Valutare quanto spende in media il signor Caio.

Ex. 2.25 Una variabile aleatoria X ha la seguente CDF

$$F_X(x) = \begin{cases} 0 & x \leq 0 \\ \frac{x}{2} & 0 < x \leq 1 \\ x - \frac{1}{2} & 1 < x \leq \frac{3}{2} \\ 1 & x > \frac{3}{2} \end{cases}$$

Schizzare il grafico di $F_X(x)$. Valutare inoltre le seguenti probabilità

$$P(\{X \leq 1/2\}), \quad P(\{X \geq 1/2\}), \quad P(\{X \leq 1.25\}) \quad P(\{X = 1.25\})$$

Ex. 2.26 L'esperimento in esame consiste nel lanciare un dado fino a quando non si ottiene un tre. Nell'ipotesi che i lanci siano indipendenti valutare la PMF della variabile aleatoria X che conta il numero di lanci.

Ex. 2.27 Verificare che la funzione

$$P_X : \quad k \in \mathbb{N} \longrightarrow q^{k-1} p \quad p \geq 0, \quad p + q = 1$$

è una PMF (distribuzione geometrica di parametro p).

Ex. 2.28 Una variabile aleatoria X ha la seguente PDF

$$f_X(x) = \begin{cases} \frac{2x}{9} & 0 < x < 3 \\ 0 & altrimenti \end{cases}$$

Verificare che la funzione assegnata soddisfi le proprietà caratterizzanti la PDF. Valutare inoltre le seguenti probabilità

$$P(\{X \leq 2\}), \quad P(\{X < 2\}), \quad P(\{-1 < X < 1.5\}).$$

Ex. 2.29 Per valutare l'efficacia di un farmaco, lo si somministra ad n ammalati di caratteristiche simili e senza relazioni genetiche, e si prende poi in esame il numero X di individui che guarisce dalla malattia. Determinare la PMF della v.a. X .

Ex. 2.30 Di una v.a X sono disponibili i seguenti due insiemi di 10 osservazioni:

$$\begin{array}{cccccccccc} 0.95 & 0.23 & 0.61 & 0.49 & 0.76 & 0.46 & 0.02 & 0.82 & 0.44 & 0.62 \\ 0.79 & 0.52 & 0.74 & 0.18 & 0.41 & 0.93 & 0.92 & 0.43 & 0.89 & 0.33 \end{array}$$

Determinare la ECDF di tale v.a a partire da ognuno dei due insiemi di osservazioni, nonché quella basata su tutte e 20 le osservazioni disponibili. Tracciare il grafico di tali ECDF.

Ex. 2.31 La durata di una lampada è modellabile come una v.a. $\mathcal{E}x(\lambda)$, con $\lambda^{-1} = 1000$ ore. In tale ipotesi valutare la frazione di lampade ancora in funzione dopo 2000 ore; si ripeta il calcolo limitatamente alle lampade che durano almeno 1500 ore. Valutare inoltre dopo quante ore di funzionamento il 50% delle lampade si è fulminato.

Ex. 2.32 Le lampade del lotto A e del lotto B hanno una durata che è una v.a. $\mathcal{E}x(\lambda_A)$ e rispettivamente $\mathcal{E}x(\lambda_B)$, con $\lambda_A = \frac{1}{800}$ ore $^{-1}$ e $\lambda_B = \frac{1}{1000}$ ore $^{-1}$.

Determinare la probabilità che una lampada scelta a caso provenga dal lotto A sapendo che si è fulminata dopo almeno 1200 ore.

Ex. 2.33 L'informazione binaria $X \sim \mathcal{B}(1, 1/2)$ è trasmessa come $+V$ se $X = 0$ altrimenti come $-V$; la trasmissione è disturbata dall'aggiunta di un rumore additivo D , indipendente dal simbolo trasmesso, avente la seguente PDF

$$f_X(x) = \frac{\lambda}{2} e^{-\lambda|x|}$$

In ricezione il valore osservato R è confrontato con 0 e si decide per la trasmissione di uno zero se $R \geq 0$ e per un uno se $R \leq 0$. Determinare il rapporto segnale rumore

$$SNR = 20 \log_{10}(V\lambda)$$

in modo che la probabilità di corretta decisione sia del 99.999%.

Capitolo 3

Segnali e sistemi nel dominio del tempo

3.1 INTRODUZIONE

La natura fisica dei segnali può essere molto diversa, ma la loro caratteristica essenziale è comune: il segnale descrive il modo di variare di una grandezza misurabile. Pertanto un segnale può essere rappresentato matematicamente da una funzione di una o più variabili indipendenti.

Esempio: Segnale vocale

Il segnale descrive una pressione acustica variabile nel tempo, $p(t)$. Tipicamente, il suono vocalico “a”, presenta una spicata ripetitività (segna quasi periodico) mentre il segnale corrispondente al suono consonantico, “s” non esibisce alcuna periodicità.

Esempio: Segnale di immagine in toni di grigio

Una immagine in toni di grigio è descritta da un valore di grigio in funzione di due coordinate spaziali, (x, y) .

Esempio: Segnale BPSK (Binary Phase Shift Keying).

Per trasmettere un segnale binario, si associa al simbolo “1” il segnale $s_1(t)$ ed al simbolo “0” il segnale $s_2(t)$. Quindi alla sequenza di simboli 100111 corrisponderà la sequenza dei segnali $s_1(t), s_2(t), s_2(t), s_1(t), s_1(t), s_1(t)$. Se $s_2(t) = -s_1(t)$ si ottiene una coppia di segnali BPSK.

Esempio: Serie storiche

Una successione di valori omogenei prende il nome di serie storica o serie temporale. Le temperature minime e quelle massime del giorno in una data località, l’indice di borsa, il numero di copie vendute al giorno da un quotidiano sono serie con cadenza giornaliera; l’indice del costo della vita costituisce una serie con cadenza mensile, etc..

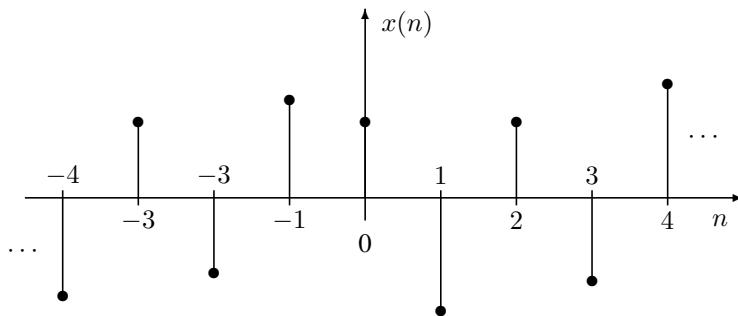


Figura 3.1: Rappresentazione grafica di una sequenza

3.2 CLASSIFICAZIONE DEI SEGNALI

Gli esempi precedenti indicano che i segnali possono essere funzioni di una sola variabile o di più variabili; nel primo caso si dicono *monodimensionali* e nel secondo *multidimensionali*: ad esempio il segnale d'immagine è bidimensionale. Nel seguito considereremo solo segnali monodimensionali, anche se la maggior parte di quanto diremo vale anche per i segnali pluridimensionali. Generalmente la variabile indipendente è di tipo temporale: adottiamo la convenzione di denominare comunque tale variabile come tempo. Con riguardo all'insieme di definizione della variabile tempo si definiscono le classi dei *segnali a tempo continuo* e dei *segnali a tempo discreto* come segue.

I segnali (a tempo) continuo, o *forme d'onda*, sono i segnali definiti su un insieme continuo; il segnale vocale e il segnale BPSK sono esempi di segnali a tempo continuo. Per convenzione un segnale a tempo continuo è individuato dall'uso della variabile indipendente t , o τ ; ad esempio $x(t)$, $y(\tau)$ denotano convenzionalmente segnali a tempo continuo.

I segnali (a tempo) discreto, o *sequenze*, sono i segnali definiti in un insieme discreto di tempi: la variabile indipendente viene di norma indicata con una delle lettere da i a n e si intende, salvo avviso contrario, che descriva l'insieme dei numeri relativi, o un suo sottinsieme; così per esempio, la notazione $x(n)$ individua un segnale discreto. Il modo di rappresentare graficamente una sequenza è illustrato nella Fig. 3.1.

Si noti che il segnale BPSK è continuo nel tempo, ma i valori assunti sono discreti, in particolare binari, mentre i valori del segnale vocale sono continui; analogamente i possibili valori della temperatura minima giornaliera descrivono un insieme continuo, mentre il numero delle copie vendute ogni giorno di un quotidiano sono necessariamente un numero intero. Corrispondentemente a questa distinzione si può anche introdurre una classificazione sulla base della gamma dei valori assunti (dinamica) in *segnali ad ampiezza continua* o *analogici* e *segnali ad ampiezza discreta*; i segnali a tempo ed ampiezza discreta, che assumono solo un numero finito di valori, si chiamano anche *numerici* o *digitali*.

Una sequenza $x(n)$ può rappresentare un fenomeno per cui la variabile indipendente

è intrinsecamente discreta, oppure può rappresentare successivi campioni di un fenomeno che evolve in modo continuo. Per esempio la successione delle temperature minime o massime giornaliere in una data località, od altre successioni inerenti grandezze metereologiche, demografiche od econometriche sono segnali intrinsecamente discreti; d'altra parte l'elaborazione della voce al calcolatore richiede che i segnali vocali vengano rappresentati mediante una successione di valori presi ad intervalli costanti di tempo (tipicamente gli intervalli sono di un centinaio di microsecondi): questo tipo di rappresentazione si chiama campionamento uniforme o semplicemente campionamento. Analogamente le immagini fisse possono essere riprodotte mediante una griglia regolare di elementi di immagine (*pixel*) attribuendo a ciascuno di essi la luminosità corrispondente dell'immagine originaria (campionamento spaziale). Nel seguito sequenze e forme d'onda verranno trattate in parallelo in modo da poterne, di volta in volta, evidenziare le analogie e le differenze.

In tutti i processi di comunicazione, sia sotto forma di trasmissione, sia sotto forma di immagazzinamento dell'informazione, intervengono segnali non perfettamente noti, cioè *segnali casuali o aleatori*: invero alla radice stessa di informazione c'è un certo grado di incertezza o impredicibilità; il trasferimento dell'informazione, trasportata dai segnali, avviene proprio per eliminare tale incertezza o impredicibilità, parzialmente o completamente. Ad esempio il segnale corrispondente ad uno specifico simbolo BPSK è un segnale deterministico, tuttavia quando si instaura il processo di comunicazione, viene generato un segnale costituito da una successione, non nota a priori, di segnali del tipo considerato in precedenza per cui il segnale complessivo è di tipo aleatorio.

Un segnale (discreto o continuo) si dice deterministico se esso è completamente descritto da una espressione matematica, da una tabella, da un grafico o da una regola di qualunque tipo.

Traslazione temporale La traslazione temporale consiste nella trasformazione $y(t) = x(t-T)$ con $T \in \mathbb{R}$ costante.

3.3 SEGNALI DETERMINISTICI ELEMENTARI

Introduciamo alcuni segnali deterministici di uso molto comune.

3.3.1 Impulso o finestra rettangolare

Nel caso continuo l'impulso rettangolare (Fig. 3.2a) di ampiezza e durata unitarie è definito da:

$$\Pi(t) = \begin{cases} 1 & \text{se } 0.5 \leq t \leq 0.5 \\ 0 & \text{altrimenti} \end{cases} \quad (3.1)$$

Conseguentemente, l'impulso rettangolare $x(t)$, centrato in t_0 , di ampiezza A e di durata T , è l'impulso:

$$x(t) = A\Pi\left(\frac{t-t_0}{T}\right)$$

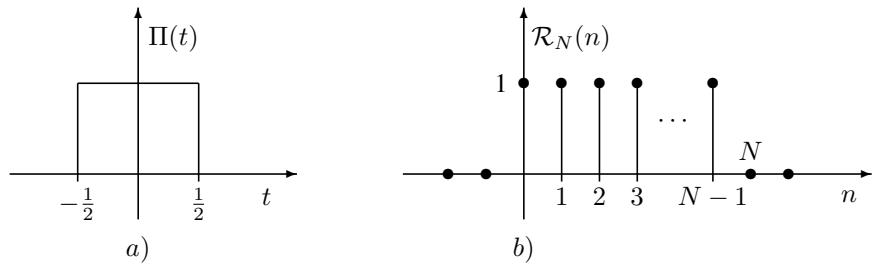


Figura 3.2: Impulso o finestra rettangolare: a) continuo, b) discreto

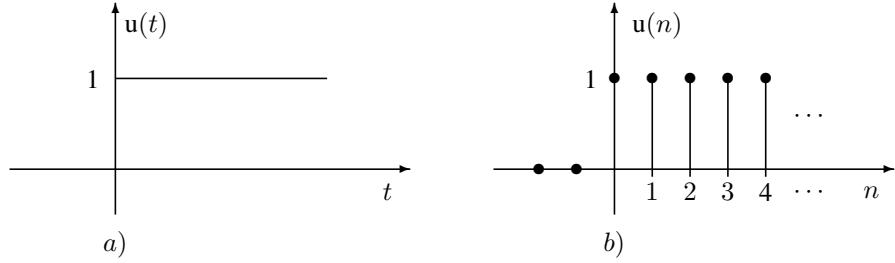


Figura 3.3: Gradino unitario: a) continuo, b) discreto

Analogamente, nel caso discreto, la finestra rettangolare $\mathcal{R}_N(n)$ di durata N (Fig. 3.2b) è definita da:

$$\mathcal{R}_N(n) = \begin{cases} 1 & \text{se } 0 \leq n \leq N-1 \\ 0 & \text{altrimenti} \end{cases} \quad (3.2)$$

3.3.2 Gradino unitario

Nel caso continuo il gradino unitario $u(t)$ (Fig. 3.3a) è definito da:

$$u(t) = \begin{cases} 1 & \text{se } t > 0 \\ 0 & \text{se } t < 0 \end{cases} \quad (3.3)$$

e nel caso discreto (3.3b) da:

$$u(n) = \begin{cases} 1 & \text{se } n \geq 0 \\ 0 & \text{se } n < 0 \end{cases} \quad (3.4)$$

Combinando due gradini di uguale ampiezza si ottengono impulsi rettangolari o finestre. Ad

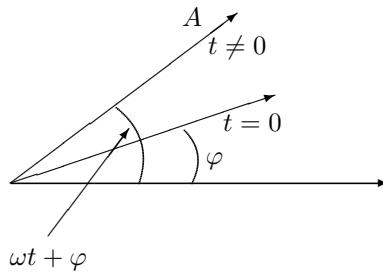


Figura 3.4: Rappresentazione grafica del fasore

esempio l'impulso rettangolare di durata T si può ottenere come segue:

$$\Pi\left(\frac{t}{T}\right) = u\left(t + \frac{T}{2}\right) - u\left(t - \frac{T}{2}\right)$$

3.3.3 Fasore (tempo continuo)

Il fasore (o vettore rotante) di pulsazione ω , frequenza f , ampiezza A e fase iniziale φ , è il segnale

$$x(t) = Ae^{j(\omega t + \varphi)} = Ae^{j(2\pi f t + \varphi)} \quad (3.5)$$

Il fasore è rappresentato, nel piano complesso, da un vettore rotante con velocità angolare ω rad./sec., ovvero velocità di rotazione di f giri al secondo o Hz, che per convenzione è positiva in senso antiorario (Fig. 3.4). Si noti che il fasore è un segnale periodico di periodo

$$T = \frac{1}{f} = \frac{2\pi}{\omega}$$

3.3.4 Segnale sinusoidale (tempo continuo)

La sinusoide di ampiezza $A > 0$, fase iniziale φ e pulsazione ω è il segnale

$$x(t) = A \cos(\omega t + \varphi). \quad (3.6)$$

Utilizzando le formule di Eulero è immediato verificare che

$$A \cos(\omega t + \varphi) = \frac{1}{2}Ae^{j(\omega t + \varphi)} + \frac{1}{2}Ae^{-j(\omega t + \varphi)}$$

cioè che un segnale sinusoidale è la somma di due fasori di ampiezza dimezzata, simmetrici rispetto all'asse reale e rotanti con la stessa velocità angolare $|\omega|$ ma in verso opposto. In alternativa è possibile rappresentare la sinusoide come la proiezione di un fasore sull'asse reale, cioè:

$$A \cos(\omega t + \varphi) = \Re \left\{ Ae^{j(\omega t + \varphi)} \right\}$$

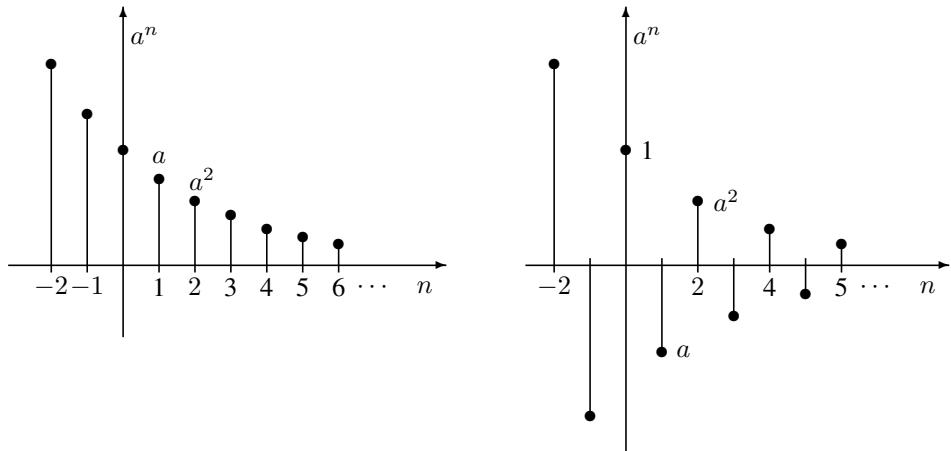


Figura 3.5: Sequenze esponenziali

3.3.5 Sequenza esponenziale (tempo discreto)

La sequenza esponenziale di ampiezza C e base z è definita come

$$x(n) = Cz^n \quad (3.7)$$

dove C ed z sono in generale complessi, se invece sono entrambi reali si ha la sequenza esponenziale reale o semplicemente sequenza esponenziale. La sequenza esponenziale reale, posto $z = a$, è a segno costante se $a > 0$ ed a segni alterni se $a < 0$ (Fig. 3.5).

3.3.6 Fasore (tempo discreto)

Il fasore a tempo discreto è definito come

$$x(n) = Ae^{j(\theta n + \varphi)} = Ae^{j(2\pi\nu n + \varphi)}$$

Come nel caso continuo esso legato alla sequenza sinusoidale dalla relazione

$$x(n) = A \cos(\theta n + \varphi) = A \cos(2\pi\nu n + \varphi)$$

Le relazioni fra fasori e sinusoidi a tempo discreto sono del tutto analoghe a quelle relative al caso continuo. Invece esistono importanti differenze fra continuo e discreto riguardo alle proprietà di periodicità dei fasori e delle sinusoidi. In particolare non è vero che $e^{j2\pi\nu n}$ fluttua sempre più velocemente al crescere di ν e non è vero che $e^{j2\pi\nu n}$ è sempre periodico rispetto

a n . Tali asserzioni discendono dalla proprietà di periodicità dell'esponenziale di argomento immaginario puro; infatti avendosi

$$e^{j2\pi(\nu+k)n} = e^{j2\pi\nu n} \quad (3.8)$$

due fasori le cui frequenze si differenzino per un numero intero sono indistinguibili. Ad esempio, posto $\nu = \nu_1 = 0.1$ e $\nu = \nu_2 = 1.1$, il fasore $e^{j0.2\pi n}$ è indistinguibile dal fasore $e^{j2.2\pi n}$. Dato che i fasori, e quindi le sinusoidi discrete, sono indistinguibili se le loro frequenze differiscono di un numero intero, basta scegliere un intervallo di ampiezza unitaria nel quale convenzionalmente definire la frequenza. Le scelte più naturali sono

$$-\frac{1}{2} \leq \nu < \frac{1}{2} \quad e \quad 0 \leq \nu < 1,$$

corrispondentemente, in termini di pulsazione, l'intervallo di definizione ha ampiezza 2π

$$-\pi \leq \theta < \pi \quad e \quad 0 \leq \theta < 2\pi$$

A causa della periodicità espressa nella (3.8) la rapidità di variazione delle sinusoidi discrete non cresce costantemente all'aumentare di ν : piuttosto, come è evidenziato in Fig. 3.6, le sequenze sinusoidali variano sempre più rapidamente al crescere di ν da 0 a $\frac{1}{2}$, ma poi, al crescere di ν da $\frac{1}{2}$ a 1 variano sempre meno rapidamente (per $\nu = 0$ e per $\nu = 1$ abbiamo un segnale costante). Quindi nel caso dei segnali discreti, analogamente al caso continuo, le basse frequenze sono quelle nell'intorno di 0, mentre le alte frequenze sono quelle nell'intorno di $\pm\frac{1}{2}$. Per quanto concerne la periodicità temporale, deve esistere un numero N tale che:

$$e^{j2\pi\nu(n+N)} = e^{j2\pi\nu n}$$

Perchè ciò avvenga è necessario e sufficiente che $\nu N = k$: quindi condizione necessaria e sufficiente affinchè un fasore sia periodico nel tempo è che la sua frequenza sia un numero razionale. Se ci mettiamo in tale ipotesi, cioè $\nu = k/N$, e assumiamo che k e N siano primi fra loro, allora il periodo è N . Così per esempio la sequenza $x_1(n) = \cos(\pi n/6)$ è periodica di periodo $N = 12$, la sequenza $x_2(n) = \cos(8\pi n/31)$ è periodica di periodo $N = 31$, mentre il segnale $x_3(n) = \cos(n/6)$ non è periodico.

3.3.7 Impulso di tipo sinc

L'impulso di tipo sinc di ampiezza unitaria è definito come

$$x(t) = \text{sinc}(t) = \frac{\sin(\pi t)}{\pi t} \quad (3.9)$$

Il segnale presenta una serie di lobi laterali di durata 1, salvo il lobo centrale che ha durata

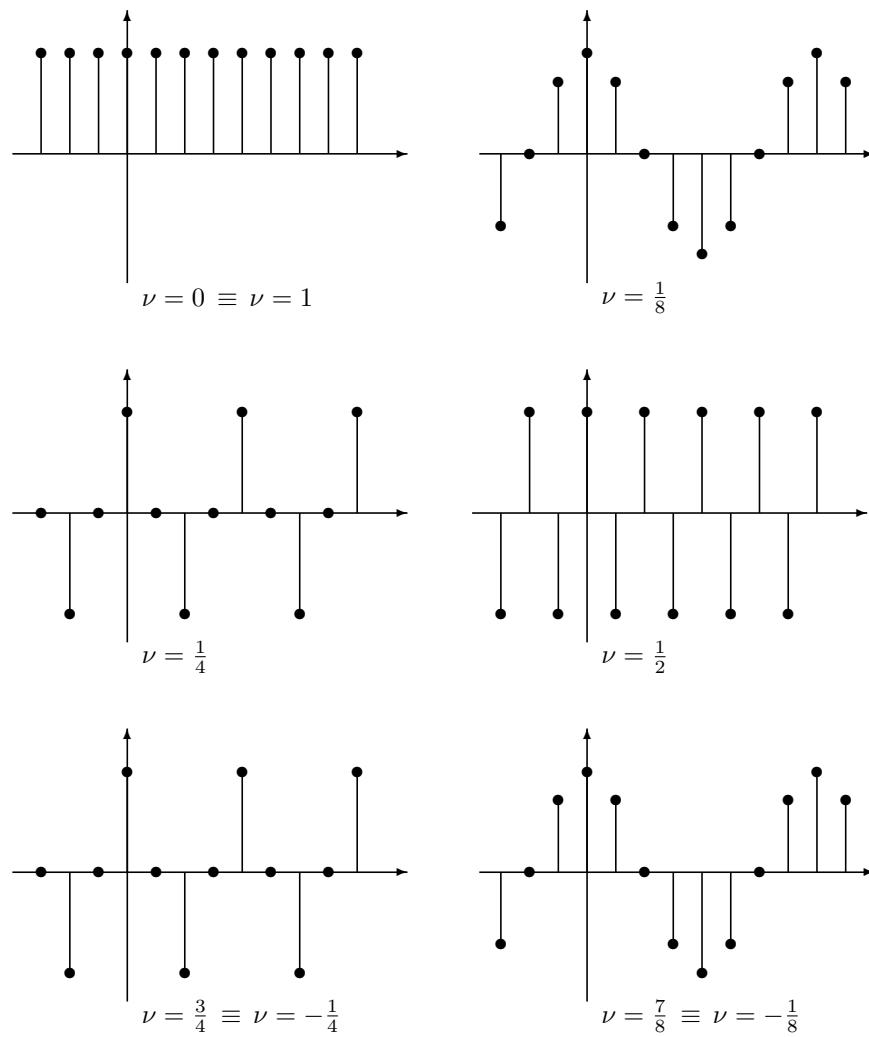


Figura 3.6: Sequenze sinusoidali per diversi valori della frequenza.

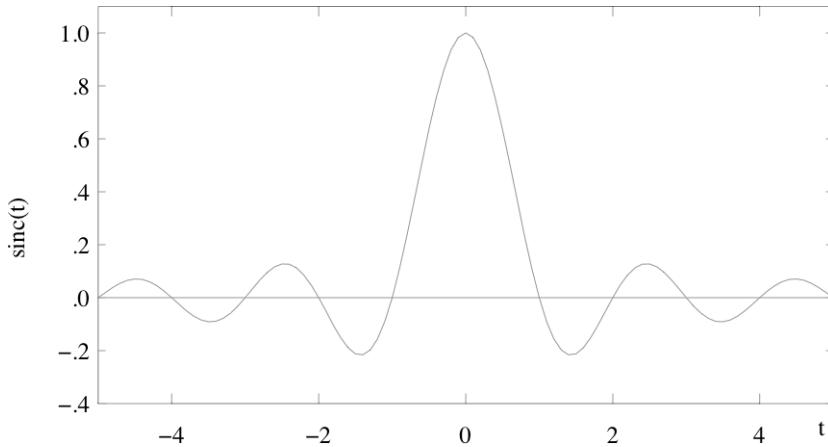


Figura 3.7: Segnale di tipo sinc

2. I lobi hanno ampiezza decrescente; in particolare il primo lobo laterale ha ampiezza 0,207 volte quella del lobo centrale, cioè si trova a $-13,26$ dB, cioè

$$\alpha_1 = 20 \log_{10} \frac{|x(0)|}{|x(t_1)|} = 13.26$$

L'ampiezza dei lobi laterali decade come $1/t$ ovvero di 6 dB/ottava = 20 dB/decade.¹

3.3.8 Impulso discreto

L'impulso unitario ideale discreto $\delta(n)$ (Fig. 3.8) è la sequenza definita da:

$$\delta(n) = \begin{cases} 1 & \text{se } n = 0 \\ 0 & \text{se } n \neq 0 \end{cases} \quad (3.10)$$

Poichè l'impulso unitario $\delta(n-k)$, locato in k , è uguale a 1 per $n = k$ e 0 altrove, è immediato verificare che, data una qualunque sequenza $x(n)$, si ha

$$x(n)\delta(n-k) = x(k)\delta(n-k) \quad (3.11)$$

cioè il prodotto di un segnale per un impulso unitario è ancora un impulso, ma di ampiezza pari al valore del segnale nel punto dove è locato l'impulso. Sommando su tutti i valori di k è

¹ si dice ottava un intervallo di frequenze (f_1, f_2) in cui $f_2 = 2f_1$, mentre se $f_2 = 10f_1$ tale intervallo si dice decade.

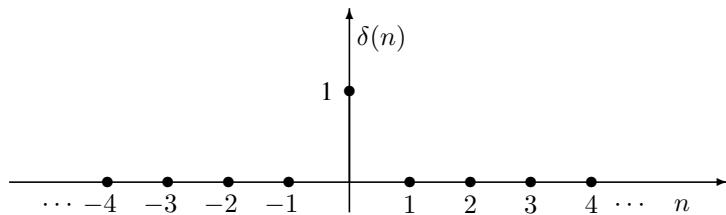


Figura 3.8: Impulso discreto

immediato verificare che

$$x(n) = \sum_{k=-\infty}^{+\infty} x(k)\delta(n-k) \quad (3.12)$$

L'eq. (3.11) mostra che operando sul segnale $x(n)$ con un singolo impulso locato in k si ottiene il campione $x(k)$ del segnale. Analogamente, secondo la (3.12), il segnale $x(n)$ dato come combinazione lineare, con coefficienti $x(k)$, di impulsi unitari locati nei vari istanti k , così come illustrato in Fig. 3.9: pertanto la (3.12) è la *proprietà di riproducibilità*.

In particolare un gradino unitario viene rappresentato come

$$u(n) = \sum_{k=-\infty}^{+\infty} u(k)\delta(n-k)$$

e quindi, tenuto conto che $u(k)$ è nullo per $k < 0$, con il cambio di indice $m = n - k$ si ha

$$u(n) = \sum_{m=-\infty}^n \delta(m)$$

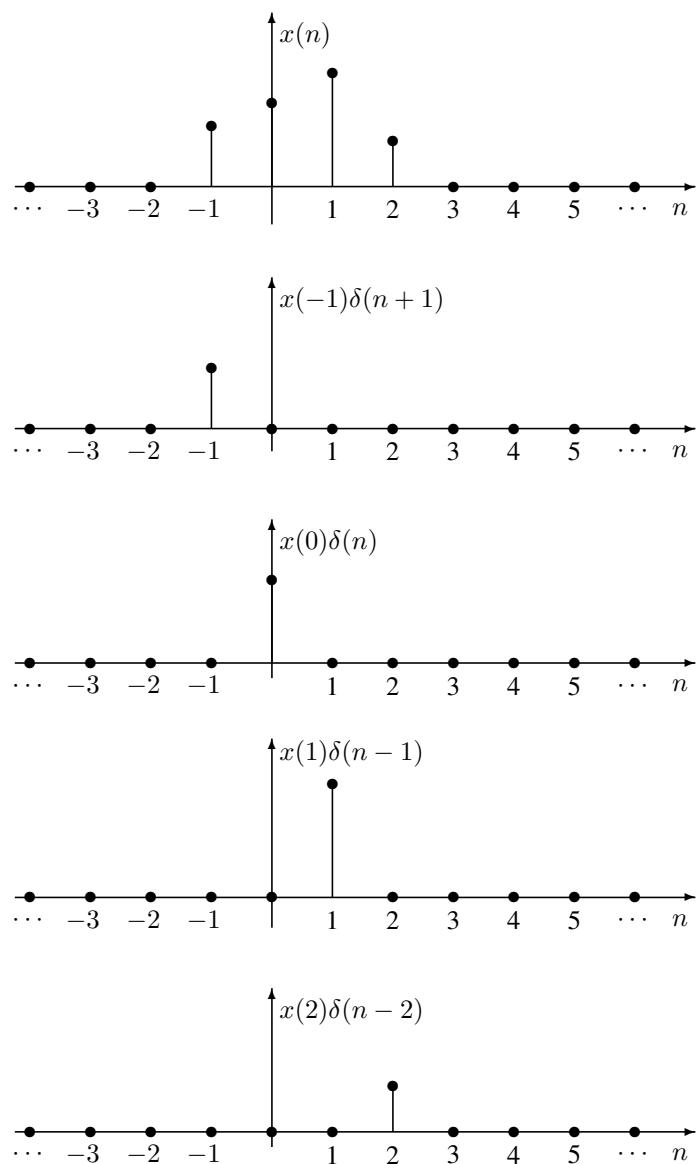
Da tale equazione segue che il gradino unitario è la somma corrente, cioè la somma tra $-\infty$ ed n , dei valori dell'impulso unitario: infatti tale somma è 0 se $n < 0$ e 1 se $n \geq 0$. Inversamente l'impulso unitario è la differenza prima $\nabla_1[\cdot]$ del gradino unitario

$$\delta(n) = u(n) - u(n-1) = \nabla_1[u(n)]$$

3.3.9 Impulso continuo (impulso di Dirac)

A differenza del caso discreto, nel quale la definizione di impulso non pone problemi di sorta, l'impulso continuo $\delta(t)$ non è una funzione in senso ordinario, bensì una funzione generalizzata. L'impulso continuo è definito dalla condizione che, per ogni funzione $x(t)$ continua in $t = 0$, si abbia

$$\int_{t_1}^{t_2} x(t)\delta(t)dt = \begin{cases} x(0) & \text{se } 0 \in (t_1, t_2) \\ 0 & \text{se } 0 \notin (t_1, t_2) \end{cases} \quad (3.13)$$

Figura 3.9: Rappresentazione di una sequenza come sovrapposizione di δ -impulsi.

Da tale definizione, utilizzando le proprietà dell'integrale, si possono ricavare formalmente le proprietà dell'impulso di Dirac. Ad esempio, posto $x(t) = 1$, $t_1 = -\infty$ e $t_2 = +\infty$, risulta:

$$\int_{-\infty}^{+\infty} \delta(t) dt = 1 \quad (3.14)$$

ciò si esprime dicendo che l'impulso $\delta(t)$ ha area unitaria; sempre dalla definizione, per $y(t)$ e $x(t)$ continui in $t = 0$, si ha:

$$\int_{-\infty}^{+\infty} y(t)x(t)\delta(t) dt = y(0)x(0)$$

ma anche

$$\int_{-\infty}^{+\infty} y(t)x(0)\delta(t) dt = y(0)x(0)$$

e quindi

$$x(t)\delta(t) = x(0)\delta(t) \quad (3.15)$$

Si ha poi, per un qualsiasi $x(t)$ continuo in 0

$$\int_{-\infty}^{+\infty} x(t)\delta(at) dt = \int_{-\infty}^{+\infty} \frac{1}{|a|} x\left(\frac{t}{a}\right) \delta(t) dt$$

e ciò è equivalente a

$$\delta(at) = \frac{1}{|a|} \delta(t)$$

che esprime la cosiddetta *proprietà di cambiamento della scala*; in particolare da tale relazione, ponendo $a = -1$, segue che $\delta(t)$ è pari. L'eq. (3.15) si generalizza considerando un impulso di Dirac applicato nell'istante di tempo τ scrivendo

$$x(t)\delta(t - \tau) = x(\tau)\delta(t - \tau) \quad (3.16)$$

e integrando ad ambo i membri dell'equazione si ottiene per ogni $x(t)$ continuo in t, si ha:

$$\int_{-\infty}^{+\infty} x(\tau)\delta(t - \tau) d\tau = x(t)$$

che esprime la *proprietà di campionamento* dell'impulso di Dirac. Dalla definizione (3.16) segue che il gradino unitario è l'integrale dell'impulso ideale, cioè

$$u(t) = \int_{-\infty}^t \delta(\tau) d\tau$$

conseguentemente risulta anche

$$\frac{du(t)}{dt} = \delta(t).$$

L'idea intuitiva che corrisponde a tutte queste proprietà è che l'impulso unitario sia zero ovunque fuorché per $t = 0$, dove è infinito, e che la sua area sia unitaria. Chiaramente nessuna funzione ordinaria soddisfa tale requisiti, ma è possibile trovare opportune famiglie di funzioni ordinarie che approssimano $\delta(t)$. Precisamente, una famiglia $\delta_T(t)$ di funzioni ordinarie converge (in senso generalizzato) a $\delta(t)$ se vale la proprietà:

$$\lim_{T \rightarrow 0} \int_{-\infty}^{+\infty} x(t) \delta_T(t) dt = x(0) \quad (3.17)$$

per ogni $x(t)$ continuo in $t = 0$; in tal caso si scrive semplicemente:

$$\lim_{T \rightarrow 0} \delta_T(t) = \delta(t)$$

Ad esempio, la proprietà (3.17) è soddisfatta per la famiglia di rettangoli

$$\delta_T(t) = \frac{1}{T} \Pi\left(\frac{t}{T}\right)$$

Più in generale una famiglia di impulsi $\delta_T(t)$ converge alla $\delta(t)$ se, al tendere a zero di T , il singolo impulso tende a concentrarsi sull'origine, la sua ampiezza diverge, mentre l'area converge ad un valore unitario. Ad esempio la famiglia di impulsi triangolari $\Lambda(\cdot)$, cioè:

$$\delta_T(t) = \frac{1}{T} \Lambda\left(\frac{t}{T}\right)$$

ove, $\Lambda(t)$ è l'impulso triangolare di ampiezza ed area unitarie, cioè è l'impulso:

$$\Lambda(t) = \begin{cases} 1 - |t| & \text{se } |t| \leq 1 \\ 0 & \text{se } |t| > 1 \end{cases} \quad (3.18)$$

converge all'impulso di Dirac al tendere a zero di T .

3.4 MEDIE TEMPORALI

Si consideri un segnale $x(t)$, la sua media temporale nell'intervallo $(t_1 \leq t \leq t_2)$ è la quantità

$$\langle x(t) \rangle_{(t_1, t_2)} = \frac{1}{t_2 - t_1} \int_{t_1}^{t_2} x(t) dt$$

analogamente la media temporale di una sequenza $x(n)$ nell'intervallo $N_1 \leq n \leq N_2$ è

$$\langle x(n) \rangle_{(N_1, N_2)} = \frac{1}{N_2 - N_1 + 1} \sum_{n=N_1}^{N_2} x(n)$$

Per le sequenze, la media temporale coincide con la media aritmetica dei campioni, mentre per le forme d'onda è l'altezza del rettangolo avente area uguale a quella sottesa dal segnale $x(t)$ nell'intervallo (t_1, t_2) e base $t_2 - t_1$. Se si fa tendere all'infinito l'ampiezza dell'intervallo si ha la media temporale (senza menzione dell'intervallo):

$$\langle x(t) \rangle = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T x(t) dt \quad (3.19)$$

nel caso continuo, ovvero

$$\langle x(n) \rangle = \lim_{N \rightarrow \infty} \frac{1}{2N + 1} \sum_{n=-N}^N x(n) \quad (3.20)$$

nel caso discreto. Alla media temporale, riguardata come segnale costante, si dà anche il nome di componente continua e la si denota con x_{dc} .

La media temporale di $x^2(\cdot)$, o di $|x(\cdot)|^2$ se si tratta di segnali complessi, si definisce valore quadratico medio o potenza media e si denota con \mathcal{P}_x :

$$\mathcal{P}_x = \langle |x(\cdot)|^2 \rangle$$

Il termine potenza è giustificato dal fatto che quest'ultima è proporzionale al valore quadratico medio: precisamente nella definizione di potenza si sottintende una costante di valore unitario, ma di opportune dimensioni; altrimenti la denominazione più appropriata è valore quadratico medio. Ad esempio se $x(t)$ è una corrente allora \mathcal{P} è la potenza media (in Watt) dissipata per effetto Joule in un resistore di 1Ω che rappresenta il valore della costante di proporzionalità. La radice quadrata del valore quadratico medio è il valore efficace, o sinteticamente valore rms, e verrà denotato con x_{rms}

$$x_{rms} = (\langle |x^2(\cdot)| \rangle)^{1/2}$$

Per una vasta classe di segnali la potenza è nulla. È questo, ad esempio, il caso dei segnali a durata limitata come l'impulso rettangolare $x(t) = \Pi(t)$ e l'impulso triangolare $x(t) = \Lambda(t)$ o non limitata come l'impulso esponenziale $x(t) = e^{-at}u(t)$ $a > 0$. Sulla scorta della potenza i segnali vengono distinti in segnali di potenza, che sono i segnali con potenza strettamente maggiore di zero, ma finita ed in segnali di energia che sono i segnali con potenza nulla ma energia finita ovvero in cui

$$\mathcal{E}_x = \int_{-\infty}^{\infty} |x(t)|^2 dt < \infty$$

per i segnali a tempo continuo, oppure

$$\mathcal{E}_x = \sum_{n=-\infty}^{\infty} |x(n)|^2 < \infty$$

per le sequenze. Notiamo che l'energia di un segnale non può essere negativa e che i segnali di energia hanno media necessariamente nulla, mentre i segnali di potenza hanno necessariamente energia infinita.

Sulla scorta della potenza, dunque, i segnali vengono distinti in segnali di potenza, che sono i segnali con potenza strettamente maggiore di zero, ma finita ed in segnali di energia che sono i segnali con potenza nulla ma energia finita. Il caso $\mathcal{P}_x = \infty$ non ha invece alcuna rilevanza pratica. Una categoria di segnali di potenza di notevole interesse è quella dei segnali periodici, per i quali la media temporale coincide con la media calcolata su di un periodo, si ha cioè

$$\langle x(t) \rangle = \frac{1}{\Delta} \int_{\Delta} x(t) dt \quad \langle x(n) \rangle = \frac{1}{L} \sum_{n=0}^{L-1} x(n)$$

ove Δ e L denotano il periodo di $x(t)$ e $x(n)$, rispettivamente.

L'operazione di media temporale, per i segnali di potenza o di energia, gode delle seguenti proprietà che si richiamano qui senza dimostrazione

P1 Linearità

Per ogni coppia (a_1, a_2) di numeri reali o complessi e per ogni coppia di segnali $(x_1(\cdot), x_2(\cdot))$ risulta:

$$\langle a_1 x_1(\cdot) + a_2 x_2(\cdot) \rangle = a_1 \langle x_1(\cdot) \rangle + a_2 \langle x_2(\cdot) \rangle$$

P2 Invarianza temporale

La media temporale di un segnale $x(\cdot)$ è invariante per traslazioni: cioè, comunque si scelga il ritardo Δ (o M se trattasi di sequenze), risulta:

$$\langle x(t - \Delta) \rangle = \langle x(t) \rangle \quad \langle x(n - M) \rangle = \langle x(n) \rangle$$

Esempio 1: Componente continua e potenza di un fasore

Si consideri un fasore di pulsazione ω , cioè il segnale

$$x(t) = e^{j\omega t}$$

Se $\omega = 0$, il segnale si riduce alla costante 1 e risulta $\langle x(t) \rangle = 1$; più in generale il valore medio di un segnale costante è uguale alla costante stessa. Per $\omega \neq 0$ il fasore è periodico di periodo $\Delta = \frac{2\pi}{\omega}$ e si ha

$$\langle e^{j\omega t} \rangle = \frac{\omega}{2\pi} \int_0^{\frac{2\pi}{\omega}} e^{j\omega t} dt = 0$$

Passando alla potenza si ha:

$$\langle |e^{j\omega t}|^2 \rangle = 1$$

e più in generale

$$\langle |Ae^{j\omega t}|^2 \rangle = A^2$$

Esempio 2: Media temporale e potenza di una sinusoide

Consideriamo il segnale

$$x(t) = A \cos(\omega t + \varphi)$$

con $\omega \neq 0$; utilizzando la rappresentazione di una sinusoide in termini di fasori e la linearità della media, si ha

$$\langle A \cos(\omega t + \varphi) \rangle = \frac{1}{2} A e^{j\varphi} \langle e^{j\omega t} \rangle + \frac{1}{2} A e^{-j\varphi} \langle e^{-j\omega t} \rangle = 0$$

Per quanto concerne la potenza, si ha

$$\langle |A \cos(\omega t + \varphi)|^2 \rangle = \frac{1}{2} A^2 + \frac{1}{2} A^2 \langle \cos[2(\omega t + \varphi)] \rangle = \frac{1}{2} A^2$$

Quindi la sinusoide è un segnale di potenza periodico a media nulla. Si osservi che l'indipendenza di media e potenza dalla fase iniziale φ è la conferma, sia pure nel caso particolare di segnali sinusoidali, dell'invarianza della media, e quindi della potenza, per traslazioni.

Esempio 3: Impulso sinc

Si consideri l'impulso sinc di ampiezza A e durata equivalente T , cioè il segnale

$$x(t) = A \operatorname{sinc}\left(\frac{t}{T}\right)$$

Si può verificare che

$$\begin{aligned} x_{dc} &= \lim_{\Delta \rightarrow \infty} \frac{1}{2\Delta} \int_{-\Delta}^{\Delta} A \operatorname{sinc}\left(\frac{t}{T}\right) dt = 0 \\ \mathcal{P}_x &= \lim_{\Delta \rightarrow \infty} \frac{1}{2\Delta} \int_{-\Delta}^{\Delta} A^2 \operatorname{sinc}^2\left(\frac{t}{T}\right) dt = 0 \\ \mathcal{E}_x &= \int_{-\infty}^{\infty} A^2 \operatorname{sinc}^2\left(\frac{t}{T}\right) dt = A^2 T \end{aligned}$$

Quindi l'impulso sinc è un segnale di energia.

La potenza e l'energia sono operazioni invarianti per traslazione, ma non sono operatori lineari. Infatti la potenza di una somma vale

$$\langle |x(\cdot) + y(\cdot)|^2 \rangle = \langle |x(\cdot)|^2 \rangle + \langle |y(\cdot)|^2 \rangle + \langle x(\cdot)y^*(\cdot) \rangle + \langle y(\cdot)x^*(\cdot) \rangle$$

ovvero, introdotta la potenza mutua \mathcal{P}_{xy} tra $x(\cdot)$ e $y(\cdot)$ e quella \mathcal{P}_{yx} tra $y(\cdot)$ e $x(\cdot)$

$$\mathcal{P}_{xy} = \langle x(\cdot)y^*(\cdot) \rangle \quad \mathcal{P}_{yx} = \langle y(\cdot)x^*(\cdot) \rangle = \mathcal{P}_{xy}^*$$

si ha

$$\mathcal{P}_{x+y} = \mathcal{P}_x + \mathcal{P}_y + \mathcal{P}_{xy} + \mathcal{P}_{yx} = \mathcal{P}_x + \mathcal{P}_y + 2\Re\{\mathcal{P}_{xy}\}$$

Le potenze mutue \mathcal{P}_{xy} e \mathcal{P}_{yx} danno conto dell'interazione in termini energetici dei segnali. Se una è nulla lo è anche l'altra: in tal caso, i segnali si dicono ortogonali e vale l'additività delle potenze.

Analogamente a quanto si verifica per le potenze, è immediato verificare, considerati due segnali di energia $x(\cdot)$ e $y(\cdot)$, l'energia \mathcal{E}_{x+y} della loro somma vale

$$\mathcal{E}_{x+y} = \mathcal{E}_x + \mathcal{E}_y + \mathcal{E}_{xy} + \mathcal{E}_{yx} = \mathcal{E}_x + \mathcal{E}_y + 2\Re\{\mathcal{E}_{xy}\}$$

ove, \mathcal{E}_{xy} e $\mathcal{E}_{yx} = \mathcal{E}_{xy}^*$ denotano rispettivamente l'energia mutua tra $x(\cdot)$ e $y(\cdot)$ e quella tra $y(\cdot)$ e $x(\cdot)$ definite da:

$$\mathcal{E}_{xy} = \int_{-\infty}^{\infty} x(t)y^*(t)dt \quad \mathcal{E}_{yx} = \int_{-\infty}^{\infty} y(t)x^*(t)dt$$

nel caso di segnali a tempo continuo e da

$$\mathcal{E}_{xy} = \sum_{n=-\infty}^{\infty} x(n)y^*(n) \quad \mathcal{E}_{yx} = \sum_{n=-\infty}^{\infty} y(n)x^*(n)$$

per i segnali a tempo discreto.

Se $\mathcal{E}_{xy} = 0$ (e quindi anche $\mathcal{E}_{yx} = 0$) allora vale l'additività per l'energia ed in tal caso, analogamente a quanto visto per i segnali di potenza, i segnali si dicono ortogonali.

Spesso la potenza è misurata in decibel (dB), cioè in unità logaritmiche rispetto ad un valore di riferimento \mathcal{P}_0 ; precisamente si ha:

$$\mathcal{P}_{dB} = 10 \log \frac{\mathcal{P}}{\mathcal{P}_0}$$

In particolare si parla di dBW e di dBm se la potenza di riferimento sono rispettivamente quella di un 1W o 1mW. Conseguentemente anche i rapporti tra ampiezze possono essere misurate in dB, ma, essendo le potenze proporzionali ai quadrati dei valori efficaci, per le ampiezze la misura in dB è data da

$$|x|_{dB} = 20 \log \left| \frac{x}{x_0} \right|$$

ove x_0 è il valore di riferimento.

3.4.1 Rappresentazione dei segnali

Per i segnali di energia e di potenza valgono numerose proprietà che possono essere utilmente interpretate alla luce di una loro rappresentazione come elementi di uno spazio lineare. È possibile dimostrare che se x e y sono due segnali di energia, allora anche ax e $x + y$, dove a è una costante arbitraria, sono segnali di energia; analogamente ogni combinazione lineare di segnali di potenza finita è un segnale di potenza finita. Inoltre, la somma di due segnali e la moltiplicazione di un segnale per uno scalare godono delle proprietà richieste dalla struttura di spazio vettoriale.

I segnali a tempo continuo e a tempo discreto di energia finita e i segnali a tempo continuo e a tempo discreto di potenza finita sono quindi, rispettivamente, elementi di quattro distinti spazi vettoriali. Gli spazi lineari in questione possono essere dotati di prodotto scalare: precisamente, nel caso dello spazio dei segnali di energia il prodotto scalare $\langle x, y \rangle$ di x e y è l'energia mutua \mathcal{E}_{xy} , mentre, nel caso dei segnali di potenza il prodotto scalare è la potenza mutua \mathcal{P}_{xy} . Si definisce, quindi

$$\langle x, y \rangle = \begin{cases} \mathcal{E}_{xy} & x, y \text{ segnali di energia} \\ \mathcal{P}_{xy} & x, y \text{ segnali di potenza} \end{cases} \quad (3.21)$$

Si osservi anche che nella relazione precedente va utilizzata l'appropriata definizione della potenza mutua a seconda del tipo di segnale di potenza, tempo continuo o tempo discreto.

Attraverso il prodotto scalare si può introdurre anche la norma di un segnale, data da

$$\|x\| = \langle x, x \rangle^{1/2}$$

così da ottenere uno *spazio normato*. In termini energetici $\|x\|^2$ è l'energia o la potenza di x , a seconda della natura dello spazio considerato. Utilizzando la norma è anche possibile introdurre una distanza $d(x, y)$ tra due segnali con la definizione

$$d(x, y) = \|x - y\|$$

con il che lo spazio diventa uno spazio metrico. Negli spazi lineari dotati di prodotto scalare vale la diseguaglianza di Schwartz

$$|\langle x, y \rangle| \leq \|x\| \|y\|$$

dove l'uguaglianza si verifica se e solo se $x = ay$, se cioè i segnali sono proporzionali. Sulla base delle precedenti considerazioni, il prodotto scalare è indicativo del grado di similitudine (proporzionalità) fra due segnali: in particolare due segnali vengono detti perfettamente simili se il loro prodotto scalare è massimo, mentre sono completamente dissimili se ortogonali.

3.4.2 Funzioni di correlazione

La funzione di correlazione $r_{xy}(\tau)$ tra due segnali è definita come prodotto scalare tra il segnale $x(t)$ ed il segnale ritardato $y(t - \tau)$ al variare del ritardo τ

$$r_{xy}(\tau) = \langle x(t), y(t - \tau) \rangle . \quad (3.22)$$

Analogamente, nel caso di sequenze, la definizione è

$$r_{xy}(m) = \langle x(n), y(n - m) \rangle . \quad (3.23)$$

Nel caso particolare in cui il secondo segnale coincida col primo, si ha la *funzione di autocorrelazione* di $x(\cdot)$, che pertanto è definita da

$$r_{xx}(\tau) = \langle x(t), x(t - \tau) \rangle \quad (3.24)$$

nel caso di forme d'onda e da

$$r_{xx}(m) = \langle x(n), x(n - m) \rangle \quad (3.25)$$

nel caso di sequenze. Si noti che per la funzione di autocorrelazione il doppio pedice è ridondante e pertanto nel seguito sarà anche utilizzata la notazione $r_x(\cdot)$. Nel caso di segnali di potenza, la (3.22) è la media temporale

$$r_{xy}(\tau) = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T x(t)y^*(t - \tau)dt$$

mentre per segnali di energia la (3.22) si esplicita nell'integrale

$$r_{xy}(\tau) = \int_{-\infty}^{\infty} x(t)y^*(t - \tau)dt.$$

Analoghe considerazioni valgono per la funzione di autocorrelazione per i segnali a tempo discreto. Si osservi che, sulla scorta dell'interpretazione data del prodotto scalare, la funzione di mutua correlazione è un indice della similitudine tra i segnali $x(t)$ ed $y(t - \tau)$ al variare del ritardo τ ; analogamente la funzione di autocorrelazione, confrontando un segnale con se stesso al variare del ritardo, è un indice della rapidità di variazione del segnale stesso, ovvero della predicitività (lineare) del valore attuale sulla scorta del valore assunto τ secondi prima.

Impulso rettangolare

Consideriamo un impulso rettangolare di durata T e di ampiezza A , cioè il segnale

$$x(t) = A \Pi\left(\frac{t}{T}\right).$$

Trattandosi di un segnale di energia ($\mathcal{E}_x = A^2T$), la sua funzione di autocorrelazione vale

$$\begin{aligned} r_x(\tau) &= \langle x(t), x(t - \tau) \rangle = \int_{-\infty}^{\infty} x(t)x^*(t - \tau)dt = \\ &= A^2 \int_{-\infty}^{\infty} \Pi\left(\frac{t}{T}\right) \Pi\left(\frac{t - \tau}{T}\right) dt = A^2 T \Lambda\left(\frac{\tau}{T}\right) = \mathcal{E}_x \Lambda\left(\frac{\tau}{T}\right) \end{aligned}$$

Segnale costante

Il segnale costante

$$x(t) = A$$

è un segnale di potenza; conseguentemente la sua funzione di autocorrelazione vale

$$r_x(\tau) = \langle x(t), x(t - \tau) \rangle = \langle x(t)x^*(t - \tau) \rangle = A^2 = \mathcal{P}_x.$$

Pertanto un segnale costante ha un'autocorrelazione costante, come è intuitivo dal momento che il segnale $x(t)$ ed il segnale ritardato $x(t - \tau)$ sono identici.

Fasori

I fasori

$$x(t) = A_1 e^{j\varphi_1} e^{j2\pi f_1 t} \quad y(t) = A_2 e^{j\varphi_2} e^{j2\pi f_2 t}$$

sono entrambi segnali di potenza; la loro funzione di mutua correlazione è data da

$$\begin{aligned} r_{xy}(\tau) &= \langle x(t), y(t - \tau) \rangle = \langle x(t)y^*(t - \tau) \rangle = \\ &= A_1 A_2 e^{j(\varphi_1 - \varphi_2)} \langle e^{j2\pi(f_1 - f_2)t} \rangle e^{j2\pi f_2 \tau} \end{aligned}$$

Da cui, ricordando che la media di un fasore è nulla a meno che la sua frequenza non sia zero, segue che la mutua correlazione di due fasori a frequenza diversa è identicamente nulla. Se invece i due fasori sono isofrequenziali, cioè $f_1 = f_2 = f_0$, allora la mutua correlazione è non nulla ed è data da

$$\begin{aligned} r_{xy}(\tau) &= A_1 A_2 e^{j(\varphi_1 - \varphi_2)} \langle e^{j2\pi(f_1 - f_2)t} \rangle e^{j2\pi f_2 \tau} = \\ &= A_1 A_2 e^{j(\varphi_1 - \varphi_2)} e^{j2\pi f_0 \tau} = \mathcal{P}_{xy} e^{j2\pi f_0 \tau}. \end{aligned}$$

Pertanto la mutua correlazione tra due fasori isofrequenziali è un fasore della stessa frequenza, ma con ampiezza pari alla loro potenza mutua. Se i due fasori isofrequenziali hanno anche la stessa ampiezza ($A_1 = A_2 = A$) e la stessa fase iniziale ($\varphi_1 = \varphi_2 = \varphi$), la relazione precedente fornisce l'autocorrelazione del fasore, precisamente risulta

$$r_x(\tau) = \langle x(t), x(t - \tau) \rangle = \langle x(t)x^*(t - \tau) \rangle = A^2 e^{j2\pi f_0 \tau} = \mathcal{P}_x e^{j2\pi f_0 \tau}$$

Pertanto l'autocorrelazione di un fasore è un fasore della stessa frequenza, con fase iniziale nulla e con ampiezza pari alla potenza del segnale.

Le proprietà delle funzioni di auto e di mutua correlazione discendono agevolmente dal fatto che esse sono dei prodotti scalari. Pertanto, la funzione di mutua correlazione gode delle seguenti proprietà:

P1 Valore nell'origine

$$r_{xy}(0) = \langle x(\cdot), y(\cdot) \rangle = \begin{cases} \mathcal{E}_{xy} \\ \mathcal{P}_{xy} \end{cases}$$

P2 Proprietà di simmetria coniugata

$$r_{xy}(\cdot) = r_{yx}^*(-\cdot)$$

P3 La funzione di mutua correlazione è limitata

$$|r_{xy}(\cdot)| \leq \|x(\cdot)\| \|y(\cdot)\|.$$

La proprietà **P1** è un'immediata conseguenza della definizione e fornisce un'interpretazione in termini energetici del valore nell'origine della mutua correlazione. La proprietà **P2** è anch'essa di immediata dimostrazione, invero, supposto per fissare le idee i segnali a tempo continuo, si ha

$$r_{xy}(\tau) = \langle x(t), y(t - \tau) \rangle = \langle y(t - \tau), x(t) \rangle^* = r_{yx}^*(-\tau)$$

Tale proprietà evidenzia che la mutua correlazione dipende dall'ordine in cui si considerano i due segnali, ma le due funzioni di mutua correlazione $r_{xy}(\cdot)$ e $r_{yx}(\cdot)$ sono tra loro legate dalla relazione di simmetria espressa dalla **P2**. La proprietà **P3** segue immediatamente dalla disuguaglianza di Schwartz e dall'invarianza della norma per traslazioni.

Passando a considerare la funzione di autocorrelazione si ha che essa gode delle seguenti proprietà:

P1 Valore nell'origine

$$r_x(0) = \|x(\cdot)\|^2 = \begin{cases} \mathcal{E}_x \\ \mathcal{P}_x \end{cases}$$

P2 Simmetria coniugata (Hermitianità)

$$r_x(\cdot) = r_x^*(-\cdot)$$

P3 La funzione di autocorrelazione è limitata ed ha un massimo nell'origine

$$|r_x(\cdot)| \leq \|x(\cdot)\|^2$$

Osserviamo che la funzione di mutua correlazione compare naturalmente quando si combinano tra loro più segnali. Ad esempio, l'autocorrelazione del segnale somma $z(t) = x(t) + y(t)$,

dove $x(t)$ e $y(t)$ sono dello stesso tipo, è data da:

$$r_z(\tau) = r_x(\tau) + r_y(\tau) + r_{xy}(\tau) + r_{yx}(\tau) = r_x(\tau) + r_y(\tau) + r_{xy}(\tau) + r_{xy}^*(-\tau)$$

Quindi la condizione $r_{xy}(\tau) = 0 \forall \tau$, è condizione sufficiente per l'additività della funzione di autocorrelazione.

Precedentemente si è visto che condizione sufficiente per l'additività dell'energia, o della potenza, è l'ortogonalità dei due segnali, cioè $r_{xy}(0) = 0$ (che, a seconda dei casi, equivale a $\mathcal{E}_{xy} = 0$ o $\mathcal{P}_{xy} = 0$). Evidentemente, la condizione $r_{xy}(\tau) = 0 \forall \tau$ è una condizione più restrittiva di $r_{xy}(0) = 0$, dal momento che essa comporta che non solo è nullo il prodotto scalare fra $x(t)$ e $y(t)$, ma anche quello fra $x(t)$ e una qualsiasi versione ritardata o anticipata di $y(t)$. Due segnali la cui mutua correlazione sia identicamente nulla, cioè $r_{xy}(\tau) = 0 \forall \tau$, si dicono *incoerenti*; per cui si può affermare che l'incoerenza è condizione sufficiente per l'additività dell'autocorrelazione oltre che per l'additività della potenza. Due fasori a frequenza diversa sono, dunque, incoerenti e l'autocorrelazione della somma di più fasori è uguale alla somma delle singole autocorrelazioni.

3.5 SISTEMI NEL DOMINIO DEL TEMPO

Un sistema è descrivibile mediante un operatore $T(\cdot)$ che converte un segnale $x(\cdot)$ detto *ingresso* in un segnale $y(\cdot)$ detto *uscita*: l'azione del sistema $T(\cdot)$ sul segnale può essere schematizzata come in figura 3.10.

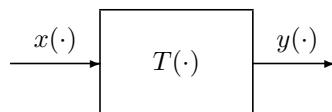


Figura 3.10: Rappresentazione di sistemi

I sistemi possono essere classificati in base ai segnali che elaborano. Così si parla di *sistema deterministico* se l'uscita corrispondente ad un qualunque ingresso deterministico è a sua volta un segnale deterministico, di *sistema numerico* se ingresso ed uscita sono segnali numerici, *discreto* se ingresso ed uscita sono sequenze, *continuo* se ingresso ed uscita sono forme d'onda. Pertanto, di norma, ingresso ed uscita di un sistema sono segnali dello stesso tipo, con le eccezioni, di notevole interesse pratico, del convertitore analogico/numerico, utilizzato per convertire un segnale ad ampiezza e tempo continui in un segnale numerico, e quello numerico/analogico impiegato per ricostruire una forma d'onda dalla sua versione numerizzata; tali sistemi sono le interfacce necessarie per poter effettuare l'elaborazione numerica di forme d'onda.

Per un sistema continuo il legame ingresso/uscita è definito dalla trasformazione

$$y(t) = T[x(t), t]$$

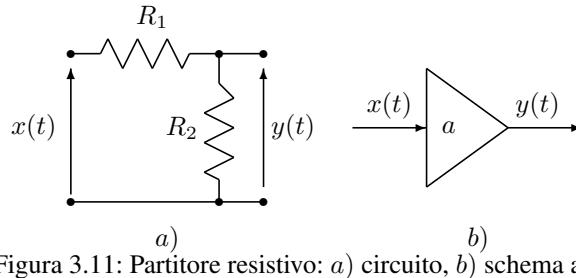


Figura 3.11: Partitore resistivo: a) circuito, b) schema a blocchi

e analogamente per un sistema discreto o numerico si ha

$$y(n) = T[x(n), n]$$

Tali notazioni sottolineano che il valore attuale dell'uscita $y(t)$ ($y(n)$) dipende sia dall'istante di tempo t (n) considerato che dal segnale d'ingresso $x(\cdot)$. I sistemi in cui l'uscita dipende, oltre che dall'ingresso, anche dal tempo si dicono tempo-varianti.

Esempio 1 Partitore resistivo

Si consideri il circuito elettrico di fig. 3.11a, denominato *partitore resistivo*: l'analisi del circuito fornisce immediatamente il legame ingresso-uscita per le tensioni:

$$y(t) = ax(t) \quad a = \frac{R_2}{R_1 + R_2}$$

dunque uscita ed ingresso risultano proporzionali con costante di proporzionalità $a \in (0, 1)$. Il partitore resistivo è un sistema continuo.

Esempio 2 Ritardo elementare

Il ritardo elementare di una posizione è il sistema discreto definito dall'equazione

$$y(n) = x(n - 1)$$

Esempio 3 Moltiplicatore (discreto) per una costante

È il sistema discreto definito dal seguente legame ingresso-uscita:

$$y(n) = ax(n)$$

Esempio 4 Filtro a media mobile (Moving Average)

Il filtro a media mobile è il sistema descritto dal legame ingresso-uscita

$$y(n) = \sum_{n=0}^N b_i x(n - i)$$

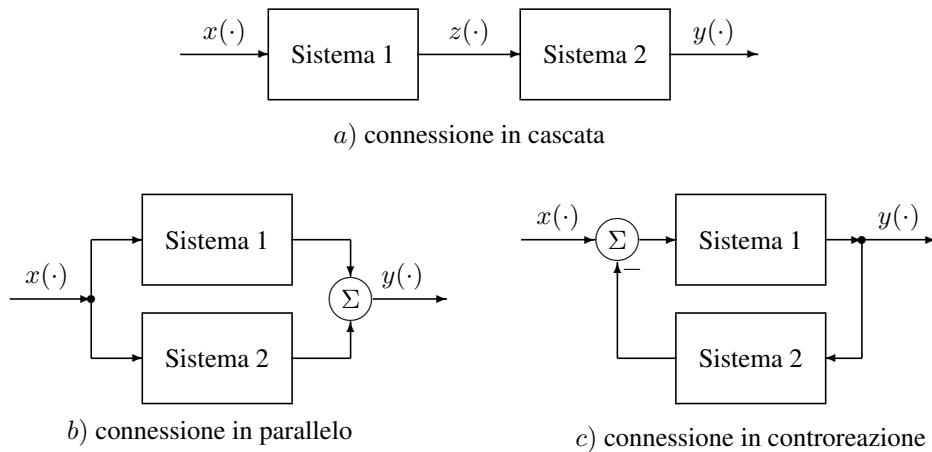


Figura 3.12: Interconnessione di sistemi.

in cui il valore attuale dell'uscita $y(n)$ è dato dalla somma pesata, con pesi b_i , del valore attuale $x(n)$ e di quelli precedenti $x(n-i)$ $i = 1, N$ dell'ingresso, il che spiega il nome di filtro a media mobile (sinteticamente filtro MA) dato a tale sistema.

È possibile costruire sistemi complessi a partire da sistemi semplici: i tipi di connessione fondamentale sono:

- connessione in cascata o serie (fig. 3.12a)
- connessione in parallelo (fig. 3.12b)
- connessione in contoreazione (feedback) (fig. 3.12c)

La figura si riferisce al caso di due sistemi, ma l'estensione delle connessioni in cascata ed in parallelo al caso di più sistemi è immediata.

Si osservi che la connessione dei sistemi oltre a indicare una strada per la realizzazione di sistemi complessi, consente anche di riguardare un dato sistema come interconnessione dei suoi componenti.

3.6 PROPRIETÀ

I sistemi vengono usualmente classificati sulla base delle loro proprietà: all'uopo è utile richiamare le principali definizioni.

3.6.1 Dispersività

Si è già osservato che l'uscita di un sistema in un determinato istante dipende in genere da tutto il segnale di ingresso: ciò si esprime dicendo che il sistema è *dispersivo o con memoria*. Viceversa un sistema si dice *non dispersivo o senza memoria* se il valore $y(t)$ (risp. $y(n)$) dell'uscita all'istante t (risp. n) dipende solo dal corrispondente valore $x(t)$ (ris. $x(n)$) dell'ingresso nello stesso istante ed eventualmente dall'istante t (risp. n). È immediato verificare che il partitore resistivo (Esempio 1) ed il moltiplicatore per una costante (Esempio 3) sono sistemi non dispersivi; invece il ritardo elementare (Esempio 2) e il filtro MA (Esempio 4) sono sistemi dispersivi.

3.6.2 Causalità

Un sistema è *causale* se il valore dell'uscita $y(t)$ all'istante t (risp. n) dipende non da tutto l'ingresso $x(\cdot)$, ma solo dai valori assunti da $x(\cdot)$ negli istanti di tempo precedenti t (risp. n) compreso; in altri termini, per un sistema causale, il valore dell'uscita non dipende dai valori futuri dell'ingresso. Conseguentemente se due ingressi ad un sistema causale sono identici fino all'istante t_0 (risp. n_0), allora anche le corrispondenti uscite sono uguali fino a tale istante. È immediato verificare che tutti i sistemi degli esempi 1-4 sono casuali, mentre il filtro MA definito dall'equazione

$$y(n) = b_0x(n) + b_1x(n+1) \quad b_1 \neq 0$$

è un sistema discreto, dispersivo e non causale.

Sebbene i sistemi casuali abbiano grande importanza, tuttavia essi non sono gli unici sistemi d'interesse pratico. Ad esempio la causalità non è una restrizione necessaria nell'elaborazione d'immagini e, più in generale, ogni qual volta la variabile indipendente non è il tempo. È parimenti superfluo imporre il vincolo della causalità per tutte le elaborazioni che non sono effettuate in tempo reale: così ad esempio i segnali geofisici, sismici, metereologici vengono spesso prima registrati e poi elaborati in tempo differito senza alcun vincolo di causalità. Infine anche nelle elaborazioni in tempo reale vengono presi in esame sistemi non casuali al fine di valutare la degradazione dovuta al vincolo di causalità nei sistemi realizzati in pratica.

3.6.3 Invertibilità

Un sistema è *invertibile* se esiste un altro sistema, detto *sistema inverso*, tale che la cascata del sistema invertibile e del suo inverso realizza la trasformazione identica (vedi fig. 3.13). Si consideri ad esempio il sistema definito dall'equazione:

$$y(n) = \sum_{k=-\infty}^n x(k)$$

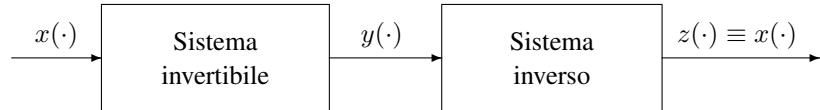


Figura 3.13: Sistema invertibile.

esso realizza la somma corrente dei valori dell'ingresso ed è denominato *accumulatore*; è immediato verificare che tale sistema è invertibile e che il suo inverso è il sistema MA definito dall'equazione:

$$z(n) = \nabla_1[y(n)] = y(n) - y(n-1) = x(n) \quad (3.26)$$

detto anche *differenza prima* e denotato con il simbolo $\nabla_1[\cdot]$.

3.6.4 Invarianza temporale

Un sistema è *temporalmente invariante* se una traslazione dell'ingresso comporta una traslazione della stessa entità anche dell'uscita: in altri termini si ha:

$$\begin{aligned} x(t) &\longrightarrow y(t) \implies x(t-T) &\longrightarrow y(t-T) \\ x(n) &\longrightarrow y(n) \implies x(n-N) &\longrightarrow y(n-N) \end{aligned}$$

È possibile verificare che i sistemi considerati negli esempi 1÷4 sono temporalmente invarianti.

Consideriamo ora il sistema definito dall'equazione

$$y(n) = nx(n)$$

La risposta $y_1(n)$ all'ingresso $x_1(n) = x(n-N)$ vale:

$$y_1(n) = nx_1(n) = nx(n-N) \neq (n-N)x(n-N) = y(n-N)$$

pertanto il sistema non è temporalmente invariante.

3.6.5 Stabilità

Un sistema è *stabile* se la risposta ad un qualunque ingresso limitato è anch'essa limitata. Tale tipo di stabilità è detta o anche *stabilità BIBO* (Bounded Input - Bounded Output). È immediato verificare che i sistemi degli esempi 1÷4 sono stabili; viceversa il sistema definito dall'equazione

$$y(n) = nx(n)$$

non è stabile; infatti la risposta ad un gradino in ingresso, cioè $x(n) = u(n)$, è una rampa, cioè $y(n) = nu(n)$, che non è limitata.

3.6.6 Linearità

Un sistema è *lineare* se è omogeneo ed additivo, cioè se esso verifica le seguenti condizioni:

Omogeneità: ad un cambiamento di scala per le ampiezze dell'ingresso corrisponde uno stesso cambiamento di scala delle ampiezza dell'uscita, cioè

$$x(\cdot) \rightarrow y(\cdot) \quad \Rightarrow \quad ax(\cdot) \rightarrow ay(\cdot)$$

qualunque sia l'ingresso $x(\cdot)$ e qualunque sia il fattore di scala a .

Additività: la risposta ad un segnale somma è la somma delle singole risposte; cioè

$$\left. \begin{array}{l} x_1(\cdot) \rightarrow y_1(\cdot) \\ x_2(\cdot) \rightarrow y_2(\cdot) \end{array} \right\} \Rightarrow x_1(\cdot) + x_2(\cdot) \rightarrow y_1(\cdot) + y_2(\cdot)$$

qualunque siano gli ingressi $x_1(\cdot)$ e $x_2(\cdot)$.

È immediato verificare che il sistema dell'esempio 2 non è lineare, mentre tutti i rimanenti lo sono.

3.7 SOMMA E INTEGRALE DI CONVOLUZIONE

Come si è visto, la proprietà caratteristica dei sistemi lineari è che se l'ingresso è una combinazione lineare di segnali

$$x(\cdot) = a_1x_1(\cdot) + a_2x_2(\cdot) + a_3x_3(\cdot) + \dots$$

allora l'uscita è data, per il principio di sovrapposizione, da:

$$y(\cdot) = a_1y_1(\cdot) + a_2y_2(\cdot) + a_3y_3(\cdot) + \dots$$

dove $y_k(\cdot)$ è l'uscita corrispondente a $x_k(\cdot)$, $k = 1, 2, 3, \dots$. È dunque conveniente rappresentare un generico ingresso $x(\cdot)$ come sovrapposizione di segnali elementari di cui sia nota la risposta, così da ottenere la risposta complessiva del sistema sovrapponendo le risposte ai singoli segnali elementari. A questo scopo una efficace decomposizione dei segnali di ingresso è quella in termini di impulsi introdotta nel capitolo precedente. Nel caso di segnali a tempo discreto, possiamo esprimere la generica sequenza $x(n)$ come

$$x(n) = \sum_{k=-\infty}^{+\infty} x(k)\delta(n-k)$$

e quindi la risposta a $x(n)$ è data dalla combinazione lineare, con coefficienti $x(k)$, delle risposte ai singoli impulsi traslati $\delta(n - k)$. Detta allora $w(n, k)$ la *risposta impulsiva in tempo-istante di applicazione* del sistema, cioè la risposta del sistema all'impulso applicato all'istante k

$$\delta(n - k) \longrightarrow w(n, k)$$

allora la risposta a $x(n)$ è data dalla cosiddetta *somma di sovrapposizione* (in tempo-istante di applicazione)

$$y(n) = \sum_{k=-\infty}^{+\infty} w(n, k)x(k) \quad (3.27)$$

In alternativa è possibile esprimere la risposta del sistema a mezzo della *risposta impulsiva in tempo-ritardo*

$$h(n, m) = w(n, n - m)$$

cioè della risposta, valutata all'istante n , ad un impulso applicato m istanti prima, in altri termini m è il ritardo, misurato rispetto all'istante di applicazione $k = n - m$, con cui si osserva l'uscita. Infatti con il cambio di variabile $k = n - m$ la (3.27) si riscrive

$$y(n) = \sum_{m=-\infty}^{+\infty} h(n, m)x(n - m) \quad (3.28)$$

ed è detta *somma di sovrapposizione in tempo-ritardo*.

Per i sistemi lineari e tempo-invarianti (LTI), ad una traslazione temporale dell'ingresso deve corrispondere una uguale traslazione temporale dell'uscita. Ciò può verificarsi solo se la risposta impulsiva in tempo-ritardo $h(n, m)$ non dipende dall'istante di osservazione n , ma solo dal ritardo m ; in tal caso, l'argomento irrilevante può essere abolito e quindi il sistema è descritto da una funzione di una sola variabile, $h(m)$. La condizione di invarianza

$$h(n, m) = h(m)$$

consente di esprimere il legame ingresso-uscita nella forma:

$$y(n) = h(n) * x(n) = \sum_{m=-\infty}^{+\infty} h(m)x(n - m) \quad (3.29)$$

Tale somma prende il nome di *convoluzione discreta* fra $h(n)$ e $x(n)$, o semplicemente di convoluzione se è chiaro che si opera su segnali discreti e, come è implicito nella (3.29), viene denotata col simbolo $h(n) * x(n)$: quindi l'uscita di un sistema LTI è la convoluzione dell'ingresso e della risposta impulsiva. Notiamo che l'ordine dei due fattori è inessenziale; infatti con cambio di variabile è immediato verificare che

$$y(n) = \sum_{m=-\infty}^{+\infty} x(m)h(n - m) \quad (3.30)$$

In altri termini la convoluzione è commutativa

$$h(n) * x(n) = x(n) * h(n)$$

Analogamente ai sistemi LTI discreti, anche i sistemi LTI continui sono caratterizzabili mediante la loro risposta impulsiva $h(\tau)$; il corrispondente legame ingresso-uscita è

$$y(t) = \int_{-\infty}^{+\infty} h(\tau)x(t - \tau) d\tau = \int_{-\infty}^{+\infty} x(\tau)h(t - \tau) d\tau = x(t) * h(t) \quad (3.31)$$

L'integrale nella (3.31) prende il nome di *convoluzione* (continua) fra $x(t)$ e $h(t)$ e converge se almeno uno dei due fattori è sommabile (in particolare converge se il sistema $h(t)$ è stabile e se l'ingresso è limitato). L'interpretazione della (3.31), e in particolare del secondo integrale è la seguente: siccome è possibile pensare a $x(t)$ come la sovrapposizione di δ -impulsi traslati $\delta(t - \tau)$, al variare del ritardo τ , di area $x(\tau)d\tau$, allora l'uscita del sistema è la sovrapposizione delle relative risposte impulsive $h(t - \tau)$ con gli stessi pesi $x(\tau)d\tau$.

Esempio 1: Convoluzione di un gradino ed una sequenza esponenziale

Ci proponiamo di valutare la risposta al gradino $x(n) = u(n)$ (fig. 3.14a) del sistema LTI avente risposta impulsiva $h(n) = a^n u(n)$, con $0 < a < 1$ (fig. 3.14e). All'uopo occorre costruire il segnale $x(n - m)$ ribaltando l'ingresso $x(m)$ (fig. 3.14b) e traslandolo di n campioni: la traslazione è in ritardo, cioè nel verso positivo dell'asse dei tempi se $n > 0$ (fig. 3.14d), o in anticipo di $|n|$ campioni se $n < 0$ (fig. 3.14c). L'uscita, per il dato valore di n , si ottiene poi sommando i valori di $h(m)$ (fig. 3.14e) pesati secondo il corrispondente valore di $x(n - m)$, cioè moltiplicando tra loro i campioni di fig. 3.14c (o di fig. 3.14d) e quelli di fig. 3.14e relativi allo stesso istante m e poi sommando i vari prodotti. Ovviamente per la commutatività della convoluzione i ruoli di segnale e risposta impulsiva possono essere scambiati. Dall'esame dei grafici di fig. 3.14a e fig. 3.14c è chiaro che, per $n < 0$, $x(n - m)$ e $h(m)$ non si sovrappongono, cioè non sono mai contemporaneamente diversi da zero e quindi $y(n) = 0$ per $n < 0$. Per $n \geq 0$, il prodotto $h(m)x(n - m)$ vale (vedi fig. 3.14a e fig. 3.14d):

$$h(m)x(n - m) = \begin{cases} a^m & 0 \leq m \leq n \\ 0 & \text{altrimenti} \end{cases}$$

Pertanto, per $n \geq 0$

$$y(n) = \sum_{m=0}^n a^m = \frac{1 - a^{n+1}}{1 - a}$$

L'andamento della risposta al gradino (*risposta indiciale*) è riportato in fig. 3.14f: si noti che tale risposta ha un asintoto pari a $1/(1 - a)$ per n tendente all'infinito.

Esempio 2: Convoluzione di due finestre rettangolari

Siano $x(n)$ e $h(n)$ due finestre rettangolari di lunghezza N e M rispettivamente, cioè:

$$x(n) = \mathcal{R}_N(n) \quad h(n) = \mathcal{R}_M(n)$$

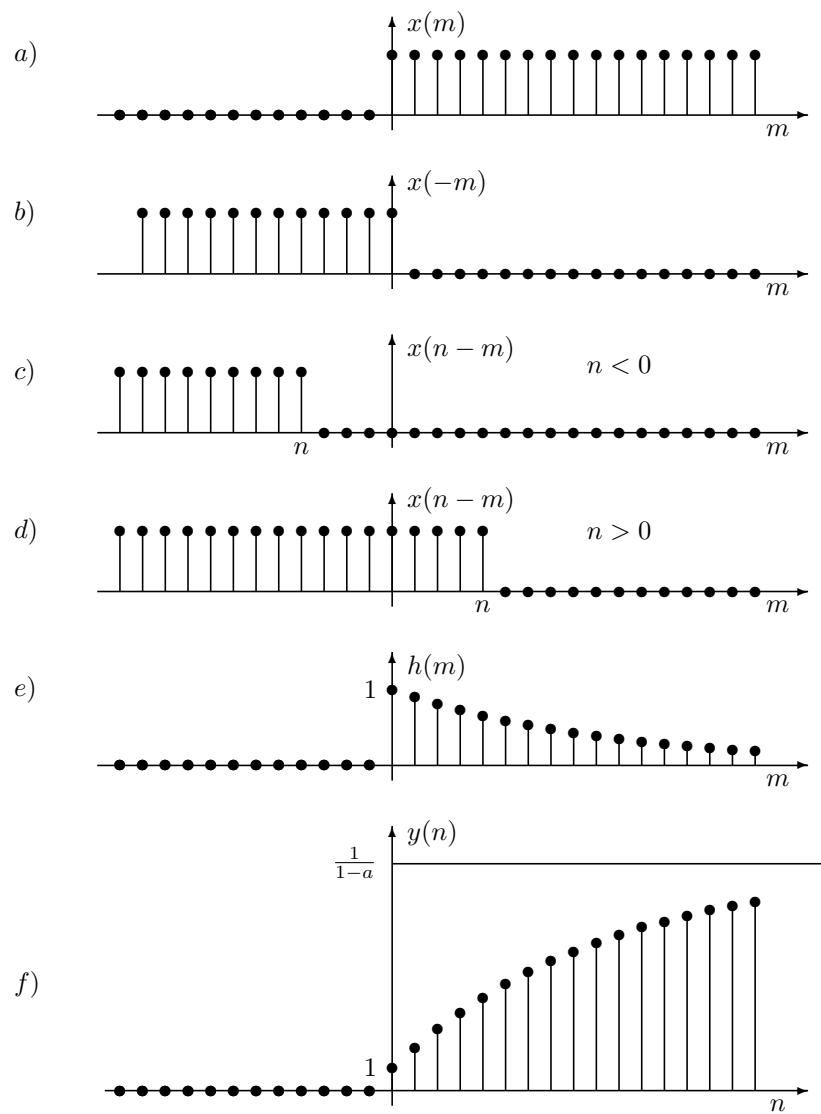


Figura 3.14: Convoluzione di un gradino e di una sequenza esponenziale.

(nelle fig. 3.15a e fig. 3.15b si è supposto $N = 4$ ed $M = 5$). Applicando la procedura precedentemente delineata, è facile verificare che $x(n) * h(n)$ ha l'andamento trapezoidale indicato in fig. 3.15c.

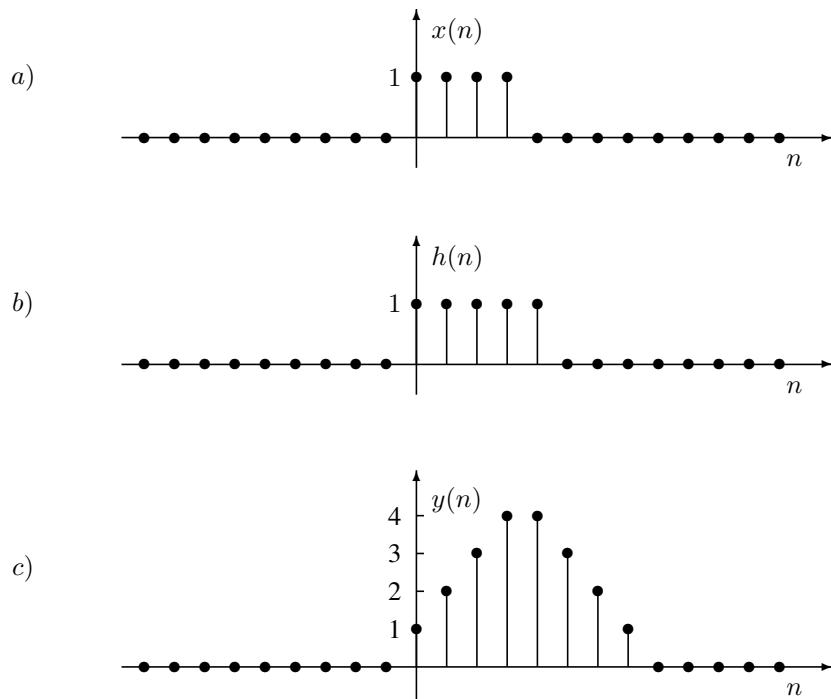


Figura 3.15: Convoluzione di due finestre rettangolari.

Dall'esempio precedente risulta che la convoluzione di due sequenze rettangolari di lunghezza N e M è una sequenza trapezoidale di lunghezza $L = N + M - 1$. Come caso particolare, ponendo $M = N$, si ricava che la convoluzione di due sequenze rettangolari della stessa lunghezza è una sequenza triangolare di lunghezza ed ampiezza $2N - 1$. Con un piccolo sforzo di generalizzazione è facile rendersi conto che la convoluzione di due qualsiasi sequenze di lunghezza finita, diciamo N e M è una sequenza di lunghezza finita $L = N + M - 1$. In altri termini se applichiamo una sequenza di lunghezza finita N in ingresso ad un sistema con memoria finita M , otteniamo in uscita una sequenza pure di lunghezza finita, ma più lunga di $M - 1$ campioni per effetto della memoria finita del sistema.

Esaminiamo ora le proprietà fondamentali della convoluzione in aggiunta alla commutatività già accennata precedentemente. Anzitutto, si può verificare che la convoluzione è *associativa*, si ha cioè:

$$x(n) * [h(n) * g(n)] = [x(n) * h(n)] * g(n) \quad (3.32)$$

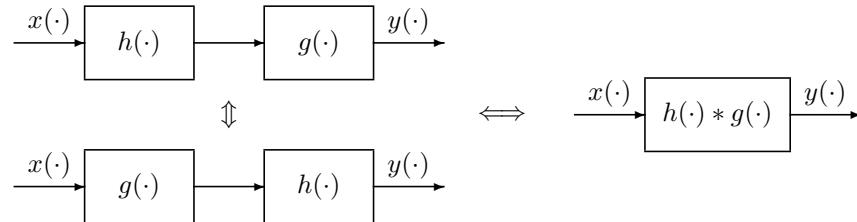


Figura 3.16: Interpretazione delle proprietà associativa e commutativa della convoluzione.

si noti che in virtù della proprietà associativa l'impiego delle parentesi quadre nella (3.32) è superfluo. Tale proprietà può essere interpretata in termini di connessione di sistemi LTI come illustrato in fig. 3.16: infatti, dalla proprietà associativa, segue che la connessione in serie di due sistemi LTI aventi risposta impulsiva $h(n)$ e $g(n)$ è equivalente a un unico sistema LTI con risposta impulsiva $h(n) * g(n)$. Inoltre, come conseguenza della proprietà associativa in congiunzione con la proprietà *commutativa*, la risposta impulsiva globale è indipendente dall'ordine con cui i sistemi sono connessi in serie. Tale proprietà può essere generalizzata ad un numero arbitrario di sistemi LTI connessi in serie.

Una terza proprietà della convoluzione è la proprietà *distributiva*

$$x(n) * [h(n) + g(n)] = x(n) * h(n) + x(n) * g(n)$$

come è facile verificare. L'interpretazione di tale proprietà è data, in termini di connessione in parallelo di sistemi LTI, in fig. 3.17: due sistemi LTI in parallelo possono essere sostituiti da un unico sistema LTI la cui risposta impulsiva è la somma delle loro singole risposte impulsive. Ovviamente questa proprietà è generalizzabile al caso di un numero arbitrario di sistemi connessi in parallelo.

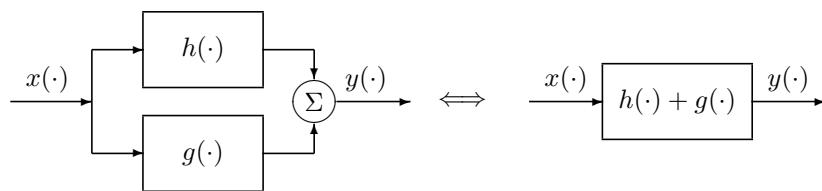


Figura 3.17: Interpretazione della proprietà distributiva della convoluzione.

È immediato verificare che l'impulso $\delta(n)$ si comporta come l'unità nei confronti della convoluzione, nel senso che si ha

$$x(n) * \delta(n) = \delta(n) * x(n) = x(n) \quad (3.33)$$

P1	commutativa	$x(\cdot) * h(\cdot) = h(\cdot) * x(\cdot)$
P2	associativa	$\begin{aligned} x(\cdot) * [h_1(\cdot) * h_2(\cdot)] &= [x(\cdot) * h_1(\cdot)] * h_2(\cdot) \\ &= x(\cdot) * h_1(\cdot) * h_2(\cdot) \end{aligned}$
P3	distributiva	$x(\cdot) * [h_1(\cdot) + h_2(\cdot)] = x(\cdot) * h_1(\cdot) + x(\cdot) * h_2(\cdot)$
P4	cambiamento di scala	$[ax(\cdot)] * h(\cdot) = x(\cdot) * [ah(\cdot)] = a[x(\cdot) * h(\cdot)]$
P5	invarianza temporale	$\begin{aligned} x(n - L) * h(n - N) &= [x * h](n - L - N) \\ x(t - T) * h(t - \Delta) &= [x * h](t - T - \Delta) \end{aligned}$
P6	esistenza dell'unità	$x(\cdot) * \delta(\cdot) = \delta(\cdot) * x(\cdot) = x(\cdot)$

Tabella 3.1

Proprietà della convoluzione.

Se si interpreta $\delta(n)$ come la risposta impulsiva del sistema identico, che è un particolare sistema LTI, la proprietà (3.33) esprime semplicemente la condizione che ingresso ed uscita di un sistema identico coincidono.

La convoluzione gode anche della *proprietà associativa mista*, cioè risulta

$$[ax(n)] * h(n) = x(n) * [ah(n)] = a[x(n) * h(n)] = ax(n) * h(n)$$

Tale proprietà traduce in termini di convoluzione l'omogeneità dei sistemi lineari, ed afferma che si può indifferentemente scalare l'ampiezza di uno dei due fattori del prodotto di convoluzione o il risultato del prodotto stesso.

Un'altra proprietà della convoluzione è l'*invarianza temporale*. Precisamente, detto $y(n)$ il risultato della convoluzione fra $x(n)$ e $h(n)$, il risultato della convoluzione fra $x(n - N)$ e $h(n)$ è $y(n - N)$, in altri termini si ha

$$x(n) * h(n) = y(n) \implies x(n - N) * h(n) = y(n - N) \quad (3.34)$$

Se interpretiamo $x(n)$ come ingresso al sistema e $h(n)$ come risposta impulsiva la (3.34) esprime semplicemente la proprietà di invarianza temporale dei sistemi LTI.

La convoluzione continua ha le stesse proprietà della convoluzione discreta; tali proprietà sono sintetizzate nella tabella 3.1 sia con riferimento al caso di segnali a tempo continuo che di sequenze. Anche nel caso di forme d'onda, tali proprietà si possono interpretare in termini di connessione di sistemi: la proprietà commutativa equivale ad affermare che i ruoli del segnale di ingresso e della risposta impulsiva possono essere scambiati; la proprietà associativa implica che la cascata di due sistemi è equivalente ad un unico sistema la cui risposta impulsiva è la convoluzione delle singole risposte impulsive; inoltre, la proprietà distributiva comporta che il parallelo di due sistemi è equivalente ad un unico sistema la cui risposta impulsiva è la

somma delle singole risposte impulsive. La procedura per calcolare la convoluzione continua è abbastanza simile a quella delineata per il caso discreto: precisamente, per valutare l'uscita per uno specifico valore di t , prima si determina il segnale $x(t-\tau)$ (considerato come funzione di τ con t fisso) mediante una riflessione intorno all'asse verticale e una traslazione (in ritardo se $t \geq 0$, in anticipo se $t < 0$); si moltiplicano poi i segnali $x(t-\tau)$ e $h(\tau)$ e si calcola l'area sottesa dal prodotto; al variare di t si ottiene l'intero segnale di uscita. Ovviamente i ruoli di $x(t)$ e $h(t)$ possono essere invertiti secondo convenienza.

Le risposta impulsiva caratterizza completamente un sistema lineare, e per tale motivo è una *risposta canonica*, conseguentemente le proprietà di un sistema lineare secondo la classificazione del paragrafo precedente, possono essere espresse in termini della risposta impulsiva. In particolare, si può effettuare la seguente classificazione:

Sistemi LTI senza memoria

Dalla (3.29) segue che un sistema lineare discreto è senza memoria, o non dispersivo, se e solo se $h(m) = 0$ per $m \neq 0$, ovvero se e solo se:

$$h(m) = a\delta(m) \quad (3.35)$$

Analogamente, per un sistema continuo la condizione è:

$$h(\tau) = a\delta(\tau) \quad (3.36)$$

Infatti se e solo se le condizioni (3.35) e (3.36) sono verificate, il valore attuale dell'uscita dipende esclusivamente dal valore dell'ingresso valutato nello stesso istante. Infatti, dalle (3.35) e (3.36) segue che in tali ipotesi, il sistema effettua semplicemente il prodotto del segnale d'ingresso per una costante a . Quando la (3.35) o (3.36) non sono verificate il sistema è con memoria (dispersivo in tempo).

Sistemi LTI causali

Il concetto di causalità si traduce, per un sistema lineare discreto, nella condizione che:

$$h(m) = 0 \quad \forall m < 0 \quad (3.37)$$

per cui la somma di convoluzione (3.28) si scrive

$$y(n) = \sum_{m=0}^{+\infty} h(m)x(n-m) \quad (3.38)$$

Infatti, se e solo se la condizione (3.37) è verificata, i valori futuri dell'ingresso, cioè i campioni $x(n-m)$ con $m < 0$, non danno contributo al valore attuale dell'uscita. Analogamente la condizione di causalità per un sistema lineare continuo è:

$$h(\tau) = 0 \quad \forall \tau < 0 \quad (3.39)$$

e il relativo integrale di convoluzione è

$$y(t) = \int_0^{+\infty} x(t-\tau)h(\tau)d\tau \quad (3.40)$$

Se le condizioni (3.37) o (3.39) non valgono il sistema è non causale. Un caso particolare di sistema non causale è il *sistema anticausale* definito dalla condizione:

$$\begin{aligned} h(m) &= 0 & \forall m > 0 \\ h(\tau) &= 0 & \forall \tau > 0 \end{aligned}$$

Sistemi lineari stabili

Anche la condizione di stabilità può essere espressa in termini della risposta impulsiva. Sia $x(n)$ un ingresso limitato, cioè $|x(n)| \leq B$, di un sistema lineare discreto, la corrispondente uscita può essere maggiorata come segue

$$|y(n)| \leq \sum_{m=-\infty}^{+\infty} |h(m)| |x(n-m)| \leq B \sum_{m=-\infty}^{+\infty} |h(m)|$$

quindi l'uscita è limitata se esiste una costante C tale che

$$\sum_{m=-\infty}^{+\infty} |h(m)| < C \quad (3.41)$$

cioè, se la risposta impulsiva è sommabile rispetto ai ritardi e la somma è limitata. Si dimostra poi che tale condizione è anche necessaria, infatti, consideriamo l'ingresso limitato definito da:

$$x(n-m) = \begin{cases} 0 & \text{se } h(m) = 0 \\ \frac{h(m)}{|h(m)|} & \text{altrimenti} \end{cases}$$

e assumiamo che la (3.41) non sia verificata; allora l'uscita è data da

$$y(n) = \sum_{m=-\infty}^{+\infty} h(m) \frac{h(m)}{|h(m)|} = \sum_{m=-\infty}^{+\infty} |h(m)|$$

e pertanto non è limitata.

Con analoghe considerazioni si verifica che condizione necessaria e sufficiente per la stabilità di un sistema continuo è che esista una costante reale C tale che:

$$\int_{-\infty}^{+\infty} |h(\tau)|d\tau < C$$

il sistema è reale se e solo se	$h(\cdot)$ è reale
il sistema è non dispersivo se e solo se	$h(m) = a\delta(m)$
il sistema è causale se e solo se	$h(m) = 0 \quad \forall m < 0$
il sistema è anticausale se e solo se	$h(m) = 0 \quad \forall m > 0$
il sistema è stabile se e solo se	$\sum_{m=-\infty}^{+\infty} h(m) < +\infty$

Tabella 3.2

Condizioni sulla risposta impulsiva per sistemi LTI.

Le condizioni cui deve soddisfare la risposta impulsiva di un sistema LTI sono riassunte nella tabella 3.2.

A proposito della connessione in serie dei sistemi, va altresì puntualizzato che l'invarianza della risposta rispetto all'ordine di connessione vale solo se tutti i sistemi sono sia lineari che tempo invarianti. Per esempio, è immediato verificare che il sistema che moltiplica per 2 (LTI) non può essere scambiato, nell'ordine di connessione, con il sistema che effettua il quadrato (tempo invariante, ma non lineare). Analogamente un sistema LTI (per esempio il filtro MA di risposta impulsiva $h(n) = u(n) - u(n-2)$) non può essere scambiato con un sistema LTV (per esempio il sistema lineare definito da $y(n) = nx(n)$) come si può verificare calcolando la risposta impulsiva delle due possibili connessioni in cascata.

Esempio 3: Accumulatore

È immediato verificare che l'accumulatore precedentemente introdotto è il sistema LTI con risposta impulsiva $h(n) = u(n)$. Tale sistema è invertibile ed il suo inverso è la *differenza prima* $\nabla_1[\cdot]$: infatti la risposta impulsiva della differenza prima è;

$$h_1(n) = \nabla_1[\delta(n)] = \delta(n) - \delta(n-1)$$

e effettuando la convoluzione con la risposta impulsiva dell'accumulatore si ottiene:

$$h(n) * h_1(n) = u(n) * [\delta(n) - \delta(n-1)] = u(n) - u(n-1) = \delta(n)$$

Esempio 4: Sistemi ARMA

Una categoria di sistemi LTI discreti di notevole importanza nelle applicazioni è costituita da quei sistemi per i quali l'ingresso $x(n)$ e l'uscita $y(n)$ soddisfano un'equazione alle differenze, lineare, a coefficienti costanti, di ordine N , cioè del tipo:

$$\sum_{k=0}^N a_k y(n-k) = \sum_{m=0}^{M-1} b_m x(n-m) \quad (3.42)$$

Iniziamo con l'osservare che, nel caso particolare $N = 0$ e $a_0 \neq 0$, tale equazione si riscrive

$$y(n) = \sum_{m=0}^{M-1} \frac{b_m}{a_0} x(n-m)$$

e pertanto definisce un sistema LTI: precisamente tale sistema è il filtro MA con risposta impulsiva

$$h(m) = \begin{cases} b_m/a_0 & m = 0, 1, \dots, M-1 \\ 0 & \text{altrimenti} \end{cases}$$

Poiché la risposta impulsiva ha durata finita, precisamente è lunga M campioni, i filtri MA sono anche detti filtri FIR (Finite Impulsive Response). I sistemi descritti dalla equazione (3.42) sono comunemente chiamati sistemi ARMA (Auto Regressive Moving Average) e comprendono come casi particolari i sistemi MA, corrispondenti, come già osservato, ad $N = 0$ e $a_0 \neq 0$, ed i sistemi AR, corrispondenti a $M = 1$ e $b_0 \neq 0$. Determiniamo la risposta impulsiva del sistema LTI, di tipo AR, definito dall'equazione

$$y(n) = ay(n-1) + x(n) \quad (3.43)$$

con la condizione iniziale $y(-1) = 0$. Essa è ricorsivamente definita da:

$$h(n) = ah(n-1) + \delta(n) \quad h(-1) = 0$$

da cui si ricava:

$$\begin{aligned} h(0) &= ah(-1) + \delta(0) = 1 \\ h(1) &= ah(0) + \delta(1) = a \\ h(2) &= ah(1) + \delta(2) = a^2 \\ &\vdots \end{aligned}$$

e quindi

$$h(n) = ah(n-1) + \delta(n) = a^n, \quad \forall n \geq 0$$

Analogamente, risolvendo l'equazione rispetto a $h(n-1)$, si ha

$$\begin{aligned} h(-2) &= a^{-1}[h(-1) - \delta(-1)] = 0 \\ h(-3) &= a^{-1}[h(-2) - \delta(-2)] = 0 \\ &\vdots \end{aligned}$$

e quindi

$$h(n-1) = a^{-1}[h(n) - \delta(n)] = 0, \quad \forall n < 0$$

Pertanto, in definitiva

$$h(n) = a^n u(n)$$

ed il sistema LTI è causale e dispersivo; inoltre è stabile se $|a| < 1$.

Viceversa se si impone la condizione $h(0) = 0$, risulta:

$$\begin{aligned} h(1) &= ah(0) + \delta(1) = 0 \\ h(2) &= ah(1) + \delta(2) = 0 \\ &\vdots \end{aligned}$$

e quindi

$$h(n) = ah(n-1) + \delta(n) = 0, \quad \forall n \geq 0$$

e, analogamente

$$\begin{aligned} h(-1) &= a^{-1}[h(0) - \delta(0)] = -a^{-1} \\ h(-2) &= a^{-1}[h(-1) - \delta(-1)] = -a^{-2} \\ &\vdots \end{aligned}$$

e, quindi

$$h(n-1) = a^{-1}[h(n) - \delta(n)] = -a^{n-1}, \quad \forall n \leq 0$$

In definitiva si ha

$$h(n) = -a^n u(-n-1)$$

e pertanto il sistema LTI è anticausale e dispersivo; inoltre esso è stabile se $|a| > 1$.

3.8 LEGAMI INGRESSO-USCITA PER LE FUNZIONI DI CORRELAZIONE

Il problema che ora affrontiamo è quello di determinare i legami tra alcuni gradezze globali, quali la componente continua e la funzione di autocorrelazione, dell'ingresso e dell'uscita di un sistema LTI. Iniziamo col valutare la componente continua dell'uscita $y_{dc} = \langle y(\cdot) \rangle$. Riferendoci, per fissare le idee, ai segnali a tempo discreto, si ha:

$$\begin{aligned} \langle y(n) \rangle &= \langle h(n) * x(n) \rangle \\ &= \left\langle \sum_{m=-\infty}^{\infty} h(m)x(n-m) \right\rangle \\ &= \langle x(n-m) \rangle \sum_{m=-\infty}^{\infty} h(m) \\ &= x_{dc} H(0) \end{aligned} \tag{3.44}$$

ove si è introdotto il *guadagno in continua* del sistema $H(0)$ pari all'area sottesa dalla risposta impulsiva, cioè:

$$H(0) = \sum_{m=-\infty}^{\infty} h(m)$$

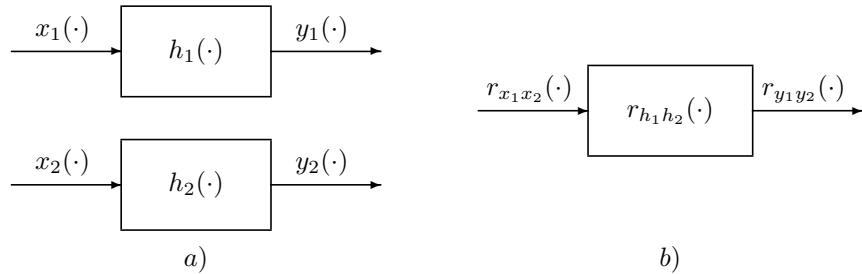


Figura 3.18: Schema per il calcolo del legame tra le mutue correlazioni degli ingressi e delle uscite.

Pertanto, la componente continua in uscita al sistema è proporzionale a quella in ingresso, con costante di proporzionalità pari al guadagno in continua.

Passando alle funzioni di autocorrelazione, consideriamo la situazione di Fig. 3.18, da cui, particolarizzando segnali e sistemi, è possibile ricavare i vari casi di interesse. I segnali in gioco possono essere di energia o di potenza, a tempo discreto o continuo: a seconda dei casi le funzioni di correlazione ammettono espressioni diverse, ma in ogni caso possono essere riguardate come prodotti scalari.

Con riferimento alla Fig. 3.18, la mutua correlazione $r_{y_1 y_2}(\cdot)$ tra le due uscite in funzione di quella $r_{x_1 x_2}(\cdot)$ tra i due ingressi e delle risposte impulsive $h_1(\cdot)$ e $h_2(\cdot)$ dei due sistemi LTI è data da

$$r_{y_1 y_2}(\cdot) = r_{x_1 x_2}(\cdot) * h_1(\cdot) * h_2^*(-\cdot) = r_{x_1 x_2}(\cdot) * r_{h_1 h_2}(\cdot) \quad (3.45)$$

$$r_{y_1 y_2}(m) = \langle y_1(n), y_2(n-m) \rangle \quad (3.46)$$

$$= \left\langle \sum_{i=-\infty}^{+\infty} h_1(i)x_1(n-i), \sum_{j=-\infty}^{+\infty} h_2(j)x_2(n-m-j) \right\rangle$$

$$r_{y_1 y_2}(m) = \sum_{i=-\infty}^{+\infty} \sum_{j=-\infty}^{+\infty} h_1(i) h_2^*(j) < x_1(n-i), x_2(n-m-j) >$$

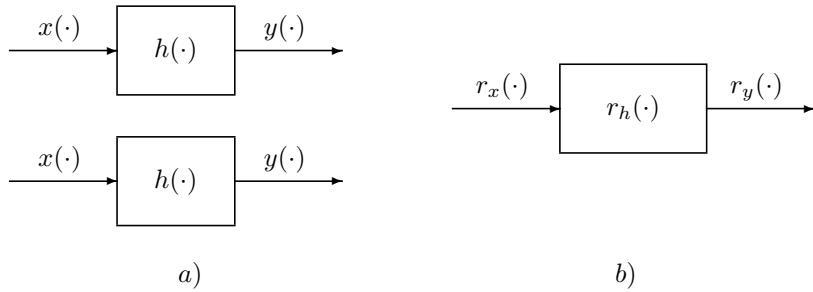


Figura 3.19: Legame ingresso uscita per l'autocorrelazione.

il prodotto scalare a secondo membro dell'equazione precedente è la mutua correlazione $r_{x_1 x_2}(m + j - i)$, conseguentemente il legame cercato è

$$r_{y_1 y_2}(m) = \sum_{j=-\infty}^{+\infty} h_2^*(j) \sum_{i=-\infty}^{+\infty} h_1(i) r_{x_1 x_2}(m+j-i)$$

Tale legame può essere anche espresso in forma più compatta in termini di convoluzione: si ha infatti

$$r_{y_1 y_2}(m) = \sum_{j=-\infty}^{+\infty} h_2^*(j)[h_1 * r_{x_1 x_2}](m+j) = \sum_{k=-\infty}^{+\infty} [h_1 * r_{x_1 x_2}](k)h_2^*(k-m)$$

L'ultimo cambio di variabili $k = m + j$ è stato introdotto per evidenziare che la residua sommatoria è ancora una convoluzione in cui il secondo fattore è $h_2^*(-\cdot)$; poiché inoltre la convoluzione è commutativa e associativa possiamo scrivere infine

$$r_{y_1 y_2}(m) = r_{x_1 x_2}(m) * \underbrace{h_1(m) * h_2^*(-m)}_{r_{h_1 h_2}(m)} = r_{x_1 x_2}(m) * r_{h_1 h_2}(m)$$

essendo $r_{h_1 h_2}(m)$ la mutua correlazione delle risposte impulsive. Una analoga derivazione vale per segnali e sistemi continui, l'unica differenza concerne la proprietà distributiva del prodotto scalare rispetto alla somma che si trasforma, con argomento limite, nella proprietà distributiva del prodotto scalare rispetto all'integrale.

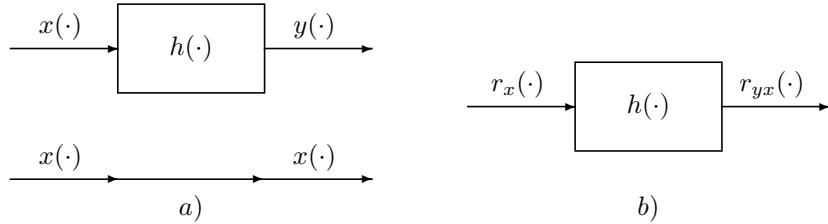


Figura 3.20: Schema di calcolo della mutua correlazione tra uscita ed ingresso.

cioè l'autocorrelazione dell'uscita è pari alla convoluzione dell'autocorrelazione dell'ingresso e di quella della risposta impulsiva: in altri termini, l'autocorrelazione della risposta di un sistema LTI è pari alla risposta all'ingresso $r_x(\cdot)$ del sistema avente risposta impulsiva $r_h(\cdot)$ (Fig. 3.19b). Dalla relazione (3.47) segue che, se l'ingresso ha un'autocorrelazione impulsiva, si ha:

$$r_y(\cdot) = a\delta(\cdot) * h(\cdot) * h^*(-(\cdot)) = a\delta(\cdot) * r_h(\cdot) = ar_h(\cdot) \quad (3.48)$$

cioè l'uscita ha un'autocorrelazione proporzionale a quella della risposta impulsiva del sistema. Pertanto è possibile generare un segnale con una preassegnata autocorrelazione filtrando, con un filtro LTI, un segnale con autocorrelazione impulsiva. Si osservi esplicitamente che il filtro LTI necessario per ottenere la correlazione assegnata risulta individuato solo tramite la funzione di autocorrelazione della sua risposta impulsiva: il problema di risalire alla risposta impulsiva a partire dalla conoscenza della sua funzione di autocorrelazione, cioè al fattore $h(\cdot)$ a partire dal prodotto di convoluzione $h(\cdot) * h^*(-(\cdot))$ (fattorizzazione dell'autocorrelazione), non è di semplice soluzione nel dominio del tempo e, inoltre, non ammette una soluzione univoca; la non unicità della soluzione può, peraltro, essere utilizzata per imporre che il filtro cercato sia causale.

L'equazione (3.45) consente anche di ricavare come casi particolari le mutue correlazioni uscita-ingresso e ingresso-uscita. Invero, sulla scorta della Fig. 3.20a, si ha:

$$r_{yx}(\cdot) = r_x(\cdot) * h(\cdot) * \delta(-(\cdot)) = r_x(\cdot) * h(\cdot) \quad (3.49)$$

che esprime il legame tra la mutua correlazione fra l'uscita e l'ingresso di un sistema LTI: dunque la mutua correlazione uscita-ingresso può calcolarsi come risposta del sistema LTI all'autocorrelazione dell'ingresso (Fig. 3.20a). In modo analogo si ottiene

$$r_{xy}(\cdot) = r_x(\cdot) * h^*(-(\cdot))$$

Analizziamo ora una interessante applicazione di tale procedura.

Esempio: Risoluzione di cammini multipli di propagazione

Si consideri la propagazione attraverso un canale non distorcente: in tal caso il segnale trasmesso $x(t)$ risulta ritardato ed attenuato per effetto della propagazione ed il mezzo non

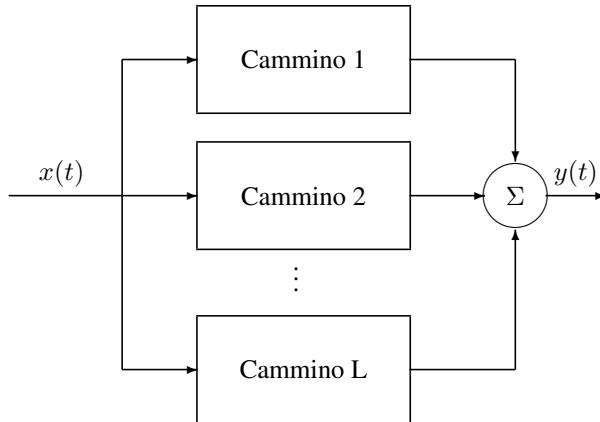


Figura 3.21: Modello di propagazione non dispersiva su cammini multipli.

dispersivo si comporta come un sistema LTI di risposta impulsiva $h(t) = a\delta(t - t_0)$. Se la propagazione avviene su cammini multipli come delineato in Fig. 3.21, supposto trasmesso $x(t)$, il segnale ricevuto vale

$$v(t) = \sum_{i=1}^N \alpha_i x(t - t_i) \quad (3.50)$$

ove α_i e t_i sono l'attenuazione e il ritardo relativi al cammino i -esimo: si pone il problema di risolvere i singoli cammini, di identificarne cioè il numero e di misurarne i vari ritardi. Il problema può essere considerato come un problema di identificazione: sollecitando il mezzo col segnale $x(t)$, la funzione di mutua correlazione fra segnale trasmesso e quello ricevuto $v(t)$ vale

$$r_{vx}(\tau) = \sum_{i=1}^L \alpha_i r_x(\tau - t_i) + r_{nx}(\tau) \quad (3.51)$$

Pertanto, ricordando che l'autocorrelazione ha un massimo nell'origine, i ritardi t_i possono essere determinati dei picchi multipli, ognuno in corrispondenza di un ritardo di propagazione, come mostrato in Fig. 3.22, ove si è supposta triangolare l'autocorrelazione del segnale trasmesso.

Si osservi che per risolvere i vari cammini, è necessario che gli impulsi presenti nella mutua correlazione non si sovrappongano e, quindi, che il segnale trasmesso abbia un'autocorrelazione di breve durata; la forma dell'autocorrelazione è inessenziale purché la durata sia piccola rispetto alla minima differenza di ritardo tra i vari cammini in modo da garantire che i picchi non si sovrappongano. Poiché $t_i = d_i/c$, essendo d_i la distanza percorsa nel mezzo e

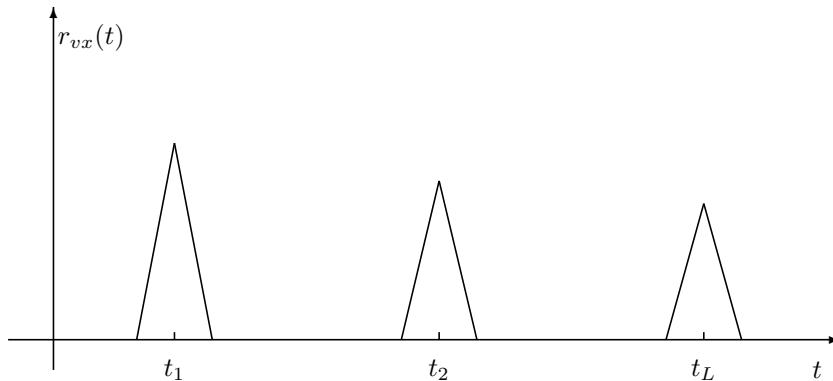


Figura 3.22: Mutua correlazione uscita ingresso in presenza di propagazione multipla.

c la velocità di propagazione, il metodo descritto si presta a risolvere problemi di valutazione della distanza (nota *c*) oppure di identificazione della velocità di propagazione nel mezzo (note le d_i).

Ex. 3.1 Sia $x(t)$ un segnale tale che $x(t) = 0 \quad \forall t < 3$. Per ognuno dei segnali elencati, determinare i valori di t per cui essi risultano nulli.

- (a) $x(1-t)$
- (b) $x(1-t) + x(2-t)$
- (c) $x(1-t)x(2-t)$
- (d) $x(3t)$
- (e) $x(t/3)$

Ex. 3.2 Disegnare il grafico dei seguenti segnali:

$$\begin{aligned} x(t) &= \Pi(t) + \Pi(-t) \\ x(t) &= \Lambda(t)\Pi(t) \\ x(t) &= \sum_{n=-\infty}^{\infty} \Lambda(t - 2n) \end{aligned}$$

Ex. 3.3 Calcolare il risultato delle seguenti espressioni:

$$\begin{aligned} &\text{sinc}(t)\delta(t) \\ &\cos(3t) * \delta(t - 1) \\ &\cos(t)\delta(3t) \end{aligned}$$

Capitolo 4

Segnali e sistemi nel dominio della frequenza

Il punto di partenza dell'analisi nel dominio del tempo è la rappresentazione di segnali (contini o discreti) mediante δ -impulsi, definita dalla formula di riproducibilità: conseguentemente la risposta di un sistema LTI ad un ingresso arbitrario si può costruire sovrapponendo le risposte ai segnali elementari costituenti l'ingresso, cioè mediante convoluzione con la risposta impulsiva. Pertanto la somma o l'integrale di convoluzione non solo rappresentano un mezzo di calcolo della risposta di un sistema LTI ad un ingresso arbitrario, ma anche indicano che le caratteristiche di un sistema LTI sono completamente determinate dalla sua risposta impulsiva e consentono quindi di studiarne le proprietà attraverso la risposta impulsiva. In questo capitolo sarà sviluppato un metodo alternativo per l'analisi dei sistemi LTI, basata sulla rappresentazione dei segnali come combinazione lineare di esponenziali complessi.

Sia $x(t) = e^{st}$, con s numero complesso, l'ingresso di un sistema LTI continuo avente risposta impulsiva $h(t)$; la corrispondente uscita è data da

$$y(t) = h(t) * e^{st} = \int_{-\infty}^{+\infty} h(\tau) e^{s(t-\tau)} d\tau = e^{st} \underbrace{\int_{-\infty}^{+\infty} h(\tau) e^{-s\tau} d\tau}_{H(s)} = H(s) e^{st}$$

ove si è posto:

$$H(s) = \left. \frac{y(t)}{x(t)} \right|_{x(t)=e^{st}} \quad (4.1)$$

Tale relazione definisce al variare di s la *funzione di trasferimento* del sistema LTI che, quindi, è il rapporto uscita/ingresso nel dominio del tempo in regime esponenziale (fig. 4.1).



Figura 4.1: Funzione di trasferimento di un sistema LTI.

Analogamente, nel caso discreto, se il sistema ha risposta impulsiva $h(n)$, per un ingresso esponenziale $x(n) = z^n$, con z numero complesso, si ha

$$y(n) = h(n) * z^n = \sum_{m=-\infty}^{+\infty} h(m)z^{n-m} = z^n \underbrace{\sum_{m=-\infty}^{+\infty} h(m)z^{-m}}_{H(z)} = H(z)z^n$$

ove

$$H(z) = \left. \frac{y(n)}{x(n)} \right|_{x(n)=z^n} \quad (4.2)$$

è la funzione di trasferimento del sistema LTI discreto (fig. 4.1). Quindi, nel caso dei sistemi continui la funzione di trasferimento è legata alla risposta impulsiva dalla relazione:

$$H(s) = \int_{-\infty}^{+\infty} h(t)e^{-st} dt \quad (4.3)$$

in altri termini essa è la trasformata (bilatera) di Laplace della risposta impulsiva. Analogamente per i sistemi discreti si ha:

$$H(z) = \sum_{m=-\infty}^{+\infty} h(m)z^{-m} \quad (4.4)$$

e la $H(z)$ è la zeta trasformata della $h(n)$. Lo studio di tali trasformate esula dallo scopo di queste note, a riguardo ci limitiamo a precisare che $H(s)$ o $H(z)$ è definita per i valori di s (risp. z) per cui l'integrale (4.3) (risp. la serie (4.4)) converge: tale insieme di valori viene usualmente denominato *regione di convergenza* (ROC).

Le due interpretazioni delle funzioni di trasferimento, cioè come rapporto uscita/ingresso nel dominio del tempo in regime esponenziale e come trasformata della risposta impulsiva, corrispondono a due metodi operativamente validi di calcolo. Nel primo dei due metodi, basato sulle relazioni (4.1) o (4.2), il sistema LTI viene sottoposto a una serie di sollecitazioni di tipo esponenziale, in modo da ricavare per punti la funzione $H(s)$ o $H(z)$. Nel secondo metodo, basato sulle (4.3) o (4.4), si ricava innanzitutto la risposta impulsiva e poi si trasforma con metodi numerici o analitici. Non è raro peraltro che i legami (4.3) e (4.4) si usino all'inverso, cioè per ricavare le risposte impulsive, note le funzioni di trasferimento, mediante le corrispondenti formule di antitrasformazione.

Esempio 1: Funzione di trasferimento di un elemento di ritardo

L'elemento di ritardo unitario discreto è definito dal legame ingresso-uscita

$$y(n) = x(n - 1)$$

L'uscita, corrispondente all'ingresso z^n , è:

$$z^n \longrightarrow z^{n-1} = z^{-1}z^n.$$

Quindi la funzione di trasferimento vale:

$$H(z) = z^{-1}$$

Tale risultato giustifica l'uso del simbolo z^{-1} per indicare gli elementi di ritardo e, analogamente, del simbolo z per indicare gli elementi di anticipo.

Nello stesso modo si vede che una linea di ritardo, cioè un sistema continuo definito dal legame ingresso-uscita

$$y(t) = x(t - T)$$

ha una funzione di trasferimento

$$H(s) = e^{-sT}$$

Esempio 2: Funzione di trasferimento di un sistema ARMA

Si consideri il sistema ARMA definito dall'equazione alle differenze

$$y(n) = \sum_{k=1}^N a_k y(n-k) + \sum_{m=0}^{M-1} b_m x(n-m)$$

con condizioni iniziali omogenee. Posto $x(n) = z^n$ e quindi $y(n) = H(z)z^n$ nell'equazione alle differenze si ottiene

$$H(z) = \frac{\sum_{m=0}^{M-1} b_m z^{-m}}{1 - \sum_{k=1}^N a_k z^{-k}}$$

Pertanto i sistemi ARMA sono caratterizzati dall'avere una funzione di trasferimento razionale in z^{-1} : in particolare per i filtri FIR (MA) $H(z)$ è un polinomio e per i sistemi AR è il reciproco di un polinomio.

4.1 RISPOSTA IN FREQUENZA DI UN SISTEMA LTI

Sebbene in generale le variabili s e z siano complesse, in questo capitolo limiteremo lo studio al caso

$$s = j\omega = j2\pi f \quad (4.5)$$

cioè s varia sull'asse immaginario e, rispettivamente, al caso

$$z = e^{j\theta} = e^{j2\pi\nu} \quad (4.6)$$

cioè z varia sulla circonferenza unitaria. Le rispettive restrizioni delle funzioni di trasferimento, se esistono, si chiamano *risposte in frequenza* o *risposte armoniche*. Quindi la risposta in frequenza per un sistema continuo è:

$$H(j\omega) = H(s) |_{s=j\omega}$$

Comunemente, con abuso di notazione, la risposta armonica si indica con $H(\omega)$ e $H(f)$, salvo quando si vuole esplicitamente mettere in evidenza il legame fra risposte in frequenza e funzioni di trasferimento. Analogamente, la risposta in frequenza di un sistema discreto è:

$$H(e^{j\theta}) = H(z) |_{z=e^{j\theta}}$$

e le notazioni semplificate sono $H(\theta)$ e $H(\nu)$. Da quanto detto e dalla definizione di funzione di trasferimento segue che la risposta in frequenza può essere definita come:

$$H(f) = \left. \frac{y(t)}{x(t)} \right|_{x(t)=e^{j2\pi ft}} \quad (4.7)$$

nel caso di sistemi continui e rispettivamente

$$H(\nu) = \left. \frac{y(n)}{x(n)} \right|_{x(n)=e^{j2\pi\nu n}}$$

per i sistemi discreti. In altri termini la risposta in frequenza è il rapporto uscita/ingresso nel dominio del tempo in regime fasoriale.

Esempio 1: Risposta in frequenza di un filtro RC
Con riferimento al filtro RC, per $x(t) = e^{j\omega t}$, si ha:

$$y(t) = e^{j\omega t} \frac{\frac{1}{j\omega C}}{R + \frac{1}{j\omega C}} = e^{j\omega t} \frac{1}{1 + j\omega RC}$$

Quindi la risposta armonica di un filtro RC vale:

$$H(\omega) = \frac{1}{1 + j\omega RC} = \frac{1}{1 + j\omega/\omega_c} \quad (4.8)$$

dove

$$\omega_c = \frac{1}{RC}$$

è la pulsazione di taglio a 3 dB del filtro; il nome è giustificato dal fatto che per $\omega = \omega_c$ si ha $|H(\omega)| = 1/\sqrt{2}$, cioè il rapporto fra le ampiezze di uscita e di ingresso è pari a -3 dB ; in corrispondenza di tale pulsazione il sistema introduce uno sfasamento di $\pi/4$ in ritardo. La risposta in frequenza $H(f)$ si ottiene da quella in termini di pulsazione (4.8) effettuando il cambiamento di variabile $\omega = 2\pi f$; si ha pertanto

$$H(f) = \frac{1}{1 + jf/f_c}$$

con

$$f_c = \frac{\omega_c}{2\pi} = \frac{1}{2\pi RC}$$

frequenza di taglio a 3 dB . La risposta armonica di un filtro RC è riportata in modulo e fase in fig. 4.2.

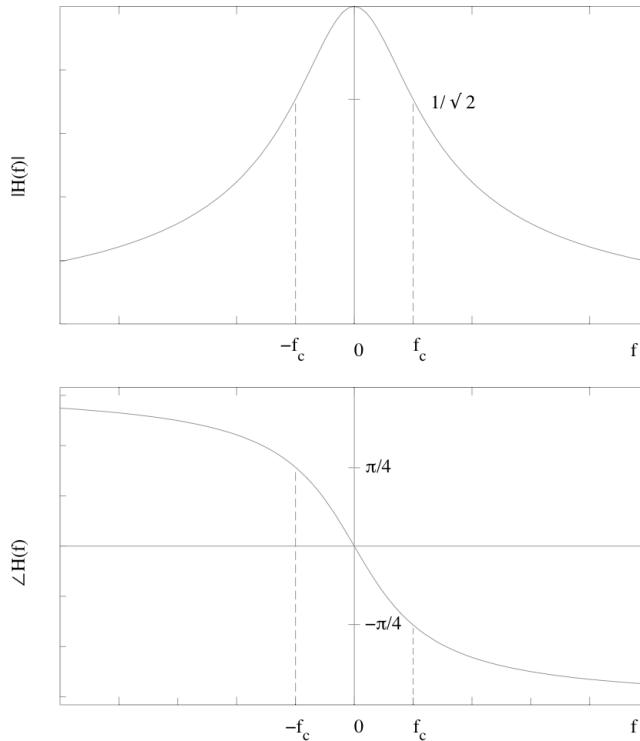


Figura 4.2: Risposta in frequenza di un filtro RC passa-basso.

Sebbene la definizione di risposta in frequenza faccia riferimento a fasori come segnali d'ingresso e quindi ad una sollecitazione complessa, è semplice, per sistemi reali, ricavare la

risposta in frequenza a partire da un ingresso di tipo sinusoidale. Si consideri quindi l'ingresso

$$x(t) = A_x \cos(2\pi f_0 t + \varphi_x)$$

da cui

$$y(t) = \underbrace{A|H(f_0)|}_{A_y} \cos[\underbrace{2\pi f_0 t + \varphi + \angle H(f_0)}_{\varphi_y}]$$

cioè la risposta ad una sinusoide di frequenza f_0 ampiezza A e fase iniziale φ è ancora una sinusoide della stessa frequenza f_0 , avente ampiezza A_y e fase iniziale φ_y date da:

$$A_y = |H(f_0)|A \quad \varphi_y = \angle[H(f_0)] + \varphi \quad (4.9)$$

Analogamente nel caso di sistemi discreti, con ovvio significato dei simboli, si ha:

$$A_y = |H(\nu_0)|A \quad \varphi_y = \angle[H(\nu_0)] + \varphi \quad (4.10)$$

Le (4.9) e (4.10) definiscono un metodo pratico, e largamente usato, per ricavare sperimentalmente modulo e fase della risposta in frequenza di un sistema incognito valutando il rapporto fra le ampiezze e lo sfasamento fra due sinusoidi rispettivamente in uscita e in ingresso al sistema; quindi, se si desidera la risposta in frequenza a diverse frequenze occorre variare la frequenza delle sinusoidi in ingresso esplorando tutto il campo desiderato.

4.2 TRASFORMATA DI FOURIER

L'analisi dei paragrafi precedenti evidenzia l'utilità di rappresentare i segnali come sovrapposizione di fasori. Per segnali continui l'equazione che definisce tale rappresentazione (*equazione di sintesi*) è:

$$x(t) = \int_{-\infty}^{+\infty} X(f) e^{j2\pi f t} df = \frac{1}{2\pi} \int_{-\infty}^{+\infty} X(\omega) e^{j\omega t} d\omega \quad (4.11)$$

e i coefficienti sono dati da (*equazione di analisi*):

$$X(f) = \int_{-\infty}^{+\infty} x(t) e^{-j2\pi f t} dt \quad X(\omega) = \int_{-\infty}^{+\infty} x(t) e^{-j\omega t} dt \quad (4.12)$$

Si noti che, con abuso di notazione, si è denotato con lo stesso simbolo $X(\cdot)$ due funzioni diverse affidando al nome della variabile indipendente, ω o f , l'identificazione della corretta interpretazione. Non esistono ragioni generali per preferire la notazione in f o la notazione in ω e pertanto le useremo più o meno indifferentemente o, in casi specifici, secondo convenienza; il passaggio dall'una all'altra è regolato dal semplice cambiamento di variabile

$\omega = 2\pi f$. Analogamente per segnali discreti l'equazione di sintesi è:

$$x(n) = \int_{-1/2}^{+1/2} X(\nu) e^{j2\pi\nu n} d\nu = \frac{1}{2\pi} \int_{-\pi}^{+\pi} X(\theta) e^{j\theta n} d\theta \quad (4.13)$$

e quella di analisi:

$$X(\nu) = \sum_{n=-\infty}^{+\infty} x(n) e^{-j2\pi\nu n} \quad X(\theta) = \sum_{n=-\infty}^{+\infty} x(n) e^{-j\theta n} \quad (4.14)$$

Anche in questo caso il passaggio dalla notazione in ν a quella in θ è regolato dal cambio $\theta = 2\pi\nu$.

La funzione $X(\cdot)$, indipendentemente dal parametro frequenziale utilizzato (ω , f , θ o ν), si chiama *Trasformata di Fourier* o *spettro* del segnale $x(\cdot)$. Notiamo esplicitamente che lo spettro è una funzione complessa e pertanto è spesso conveniente considerarne la parte reale e la parte immaginaria, ovvero il modulo e la fase: questi ultimi vengono comunemente denominati *spettro d'ampiezza* e, rispettivamente, *spettro di fase*. Nel seguito, per indicare sinteticamente il legame di trasformazione secondo Fourier useremo la scrittura:

$$X(\cdot) = \mathcal{F}\{x(\cdot)\} \quad x(\cdot) = \mathcal{F}^{-1}\{X(\cdot)\}$$

ovvero anche:

$$x(\cdot) \longleftrightarrow X(\cdot)$$

Le formule precedenti mettono in luce un notevole parallelismo nella rappresentazione spettrale di segnali continui e discreti: la principale differenza, che discende dalla periodicità in frequenza del fasore discreto $e^{j2\pi\nu n}$, è che per segnali continui lo spettro $X(f)$ è di norma aperiodico, mentre lo spettro $X(\nu)$ di una sequenza è sempre periodico di periodo 1 (ovvero $X(\theta)$ è periodico di periodo 2π), pertanto è sufficiente considerare $X(\nu)$ nell'intervallo $(-\frac{1}{2}, +\frac{1}{2})$ (ovvero $X(\theta)$ nell'intervallo $(-\pi, \pi)$). In altri termini, per ricostruire $x(n)$ si integrano i contributi spettrali nell'intervallo finito $(-\frac{1}{2}, \frac{1}{2})$ (o $(-\pi, \pi)$), ma, data la periodicità, un qualunque intervallo di ampiezza pari al periodo può essere usato alternativamente ed è molto comune anche la scelta degli intervalli $(0, 1)$ o $(0, 2\pi)$.

Il legame ingresso uscita nel dominio della frequenza per un sistema LTI segue immediatamente dal principio di sovrapposizione: infatti esprimendo l'ingresso $x(t)$, supposto tempo continuo, come sovrapposizione di fasori, cioè

$$x(t) = \int_{-\infty}^{+\infty} X(f) e^{j2\pi ft} df$$

e ricordando che i fasori sono autofunzioni per i sistemi LTI, si ottiene immediatamente la corrispondente rappresentazione dell'uscita in termini di fasori, cioè

$$y(t) = \int_{-\infty}^{+\infty} H(f) X(f) e^{j2\pi ft} df \quad (4.15)$$

Un analogo risultato vale per segnali e sistemi discreti. Da tale equazione segue che gli spettri dell'ingresso e dell'uscita sono legati alla risposta in frequenza del sistema da

$$Y(\cdot) = X(\cdot)H(\cdot) \quad (4.16)$$

Dalla (4.16) segue che il legame per gli spettri di ampiezza e di fase è:

$$|Y(\cdot)| = |X(\cdot)||H(\cdot)| \quad \angle Y(\cdot) = \angle X(\cdot) + \angle H(\cdot) \quad (4.17)$$

Pertanto lo spettro d'ampiezza dell'uscita è pari al prodotto dello spettro d'ampiezza dell'ingresso per la risposta in ampiezza del sistema $|H(\cdot)|$ e quello di fase è dato dalla somma dello spettro di fase dell'ingresso e della risposta in fase $\angle H(\cdot)$ del sistema.

Il legame ingresso/uscita nel dominio della frequenza evidenzia come i sistemi LTI abbiano un comportamento selettivo in frequenza: alcune componenti dell'ingresso sono amplificate, altre attenuate o eliminate del tutto; tale proprietà è denominata capacità di filtraggio e spiega l'impiego del termine filtro come sinonimo di sistema.

4.3 ESEMPI DI TRASFORMATE

Impulso esponenziale monolatero continuo

Consideriamo un impulso esponenziale monolatero di ampiezza A e costante di tempo T , cioè il segnale $x(t) = Ae^{-t/T} u(t)$. Dall'equazione di sintesi si ha:

$$X(f) = \int_0^{+\infty} Ae^{-t/T} e^{-j2\pi ft} dt = \frac{AT}{1 + j(2\pi fT)}$$

e quindi:

$$Ae^{-t/T} u(t) \longleftrightarrow \frac{AT}{1 + j(2\pi fT)} \quad (4.18)$$

In particolare spettro d'ampiezza e di fase sono dati da:

$$|X(f)| = \frac{AT}{\sqrt{1 + (2\pi fT)^2}} \quad \angle X(f) = -\arctan(2\pi fT)$$

e sono riportati in fig. 4.3.

Si noti che lo spettro d'ampiezza è massimo per $f = 0$ e decade come f^{-1} al crescere della frequenza, è cioè concentrato nell'intorno della frequenza nulla (spettro passabasso); inoltre la frequenza $f_0 = 1/(2\pi T)$ è la frequenza di taglio a 3 dB avendosi

$$20 \log \left| \frac{X(f)}{X(0)} \right|_{f=f_0} = 10 \log \frac{1}{1 + (2\pi fT)^2} \Big|_{f=f_0} = -3 \text{ dB}$$

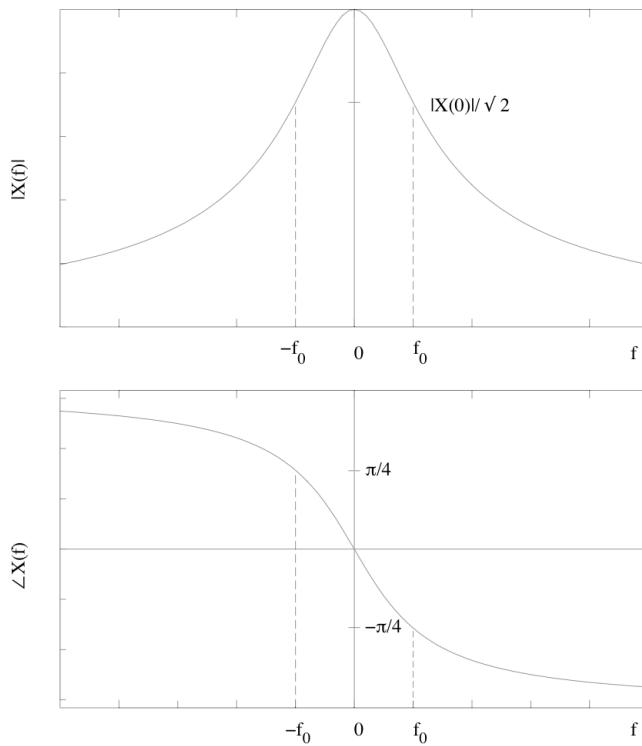


Figura 4.3: Spettro di un impulso esponenziale monolatero.

Definita la banda di un segnale come la gamma di frequenze occupata dalle componenti significative dello spettro, per il segnale in esame la banda può essere definita come un opportuno multiplo della frequenza di taglio a 3 dB. Ad esempio, posto $B = 10f_0$ (banda monolatera), le componenti esterne all'intervallo $(-B, B)$ sono attenuate di almeno 23 dB rispetto a quella di centrobanda; in altri termini si ha:

$$20 \log \left| \frac{X(f)}{X(0)} \right| \geq -23 \text{ dB} \quad \forall |f| > B$$

Si osservi infine che al diminuire della costante di tempo T diminuisce la durata del segnale, ma aumenta la sua banda.

Impulso esponenziale bilatero continuo

L'impulso esponenziale bilatero continuo è il segnale

$$x(t) = e^{-a|t|} \quad a > 0$$

ed il suo spettro vale:

$$X(f) = \int_{-\infty}^{+\infty} e^{-a|t|} e^{-j2\pi ft} dt = \frac{2a}{a^2 + (2\pi f)^2}$$

Pertanto risulta:

$$e^{-a|t|} \longleftrightarrow \frac{2a}{a^2 + (2\pi f)^2}$$

Lo spettro $X(f)$ è reale e pari ed è rappresentato in fig. 4.4. Anche l'impulso esponenziale è un segnale passa-basso ed a è la pulsazione di taglio a 6 dB.

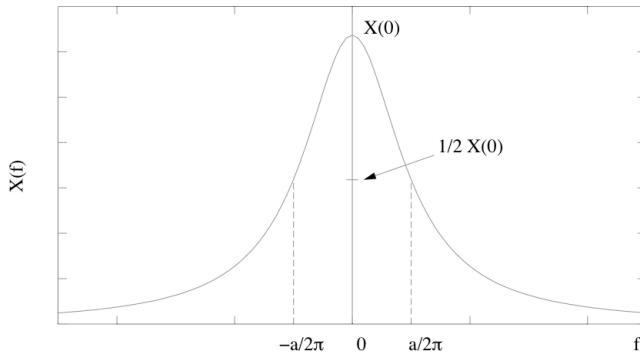


Figura 4.4: Spettro di un impulso esponenziale bilatero.

Sequenza esponenziale monolatera

La sequenza esponenziale monolatera è il segnale

$$x(n) = a^n u(n) \quad |a| < 1$$

ed il suo spettro vale:

$$X(\theta) = \sum_{n=-\infty}^{+\infty} a^n u(n) e^{-j\theta n} = \sum_{n=0}^{+\infty} (a e^{-j\theta})^n = \frac{1}{1 - a e^{-j\theta}}$$

In definitiva si ha:

$$a^n u(n) \longleftrightarrow \frac{1}{1 - ae^{-j\theta}}$$

ovvero

$$a^n u(n) \longleftrightarrow \frac{1}{1 - ae^{-j2\pi\nu}}$$

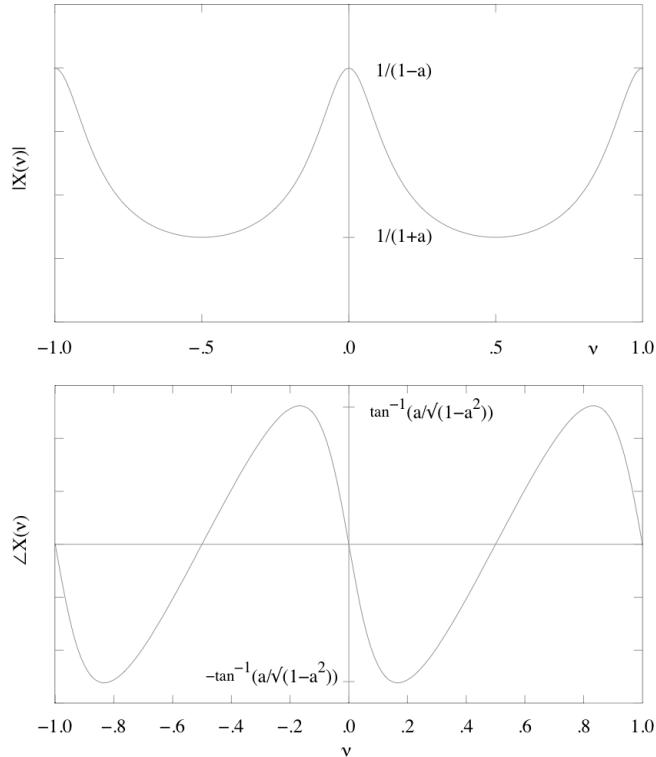


Figura 4.5: Spettro della sequenza esponenziale monolatera (passa-basso).

In fig. 4.5 e in fig. 4.6 sono riportati gli spettri d'ampiezza e fase rispettivamente per $a > 0$ e per $a < 0$. Per $a > 0$ la sequenza è un segnale passa-basso, mentre per $a < 0$ lo spettro è concentrato nell'intorno delle frequenze $\pm \frac{1}{2}, \pm \frac{3}{2}, \dots$ (segnale passa-alto). Il valore di $|a|$ regola la rapidità di decadimento della sequenza nel dominio del tempo e la larghezza di banda nel dominio della frequenza: precisamente al diminuire di $|a|$ la sequenza va sempre più rapidamente a zero mentre la sua banda aumenta.

Sequenza esponenziale bilatera

Si consideri la sequenza esponenziale bilatera cioè il segnale:

$$x(n) = a^{|n|} \quad |a| < 1$$

La sua trasformata vale:

$$X(\theta) = \sum_{n=0}^{+\infty} a^n e^{-j\theta n} + \sum_{n=-\infty}^{-1} a^{-n} e^{-j\theta n} = \sum_{n=0}^{+\infty} (ae^{-j\theta})^n + \sum_{k=1}^{+\infty} (a e^{j\theta})^k$$

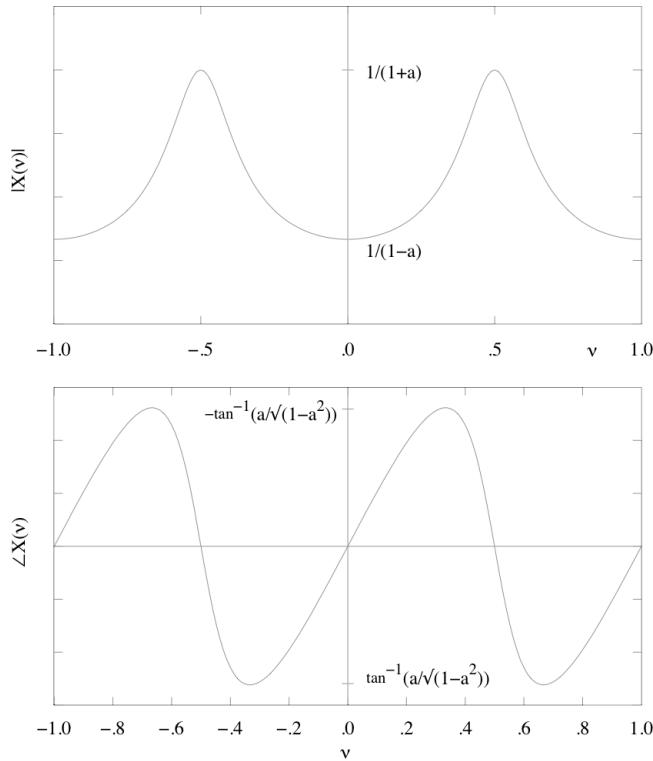


Figura 4.6: Spettro della sequenza esponenziale monolatera (passa-alto).

La prima sommatoria è la serie geometrica di ragione $ae^{-j\theta}$ e la seconda sommatoria è la serie geometrica di ragione $ae^{j\theta}$ mancante del primo termine e quindi:

$$X(\theta) = \frac{1}{1 - a e^{-j\theta}} + \frac{1}{1 - a e^{j\theta}} - 1 = \frac{1 - a^2}{1 - 2a \cos \theta + a^2}$$

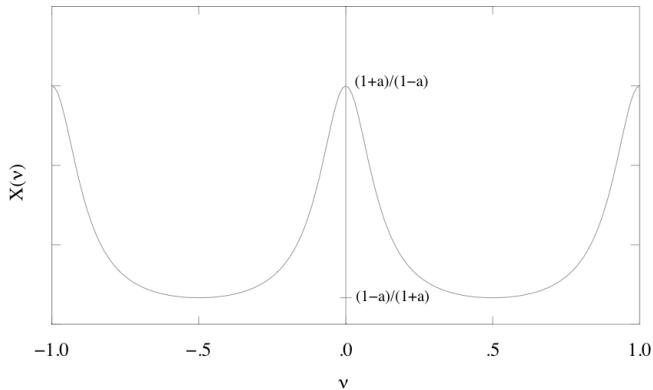
In definitiva risulta:

$$a^{|n|} \longleftrightarrow \frac{1 - a^2}{1 - 2a \cos \theta + a^2}$$

ovvero

$$a^{|n|} \longleftrightarrow \frac{1 - a^2}{1 - 2a \cos(2\pi\nu) + a^2}$$

Lo spettro della sequenza esponenziale bilatera è reale e pari; inoltre il segnale è passa-basso, per $0 < a < 1$, mentre per $a < 0$ è passa-alto. Tale trasformata è diagrammata in fig. 4.7 per $0 < a < 1$.

Figura 4.7: Spettro della sequenza esponenziale bilatera ($0 < a < 1$).*Impulso rettangolare*

Consideriamo un impulso rettangolare di ampiezza A e durata T

$$x(t) = A \Pi\left(\frac{t}{T}\right)$$

La sua trasformata è:

$$X(f) = A \int_{-\frac{1}{2}T}^{\frac{1}{2}T} e^{-j2\pi f t} dt = A \frac{\sin(\pi f T)}{\pi f} = AT \text{sinc}(fT)$$

Dunque:

$$\Pi\left(\frac{t}{T}\right) \longleftrightarrow AT \text{sinc}(fT) \quad (4.19)$$

Lo spettro dell'impulso rettangolare è reale e pari ed è riportato in fig. 4.8. Tale spettro esibisce una serie di lobi di larghezza $1/T$, salvo il centrale che è di larghezza $2/T$, e di altezza decrescente; in particolare il primo lobo laterale ha ampiezza 0,207 volte quella del lobo principale, corrispondente ad un'attenuazione di $\alpha_{SL} = 13.2$ dB. Il segnale è passa-basso e come larghezza di banda si può assumere la semiampiezza $1/T$ del lobo principale (banda monolatera); in tal caso, le componenti fuori banda sono attenuate rispetto a quella a frequenza zero di almeno 13.2 dB, si ha cioè

$$\alpha(f) = 20 \log \frac{|X(0)|}{|X(f)|} \geq \alpha_{SL} = 13.2 \text{ dB} \quad \forall |f| \geq \frac{1}{T}$$

e il loro inviluppo decade, al crescere della frequenza, come $1/f$, pari ad un decremento di 6 dB/ott. = 20 dB/dec.

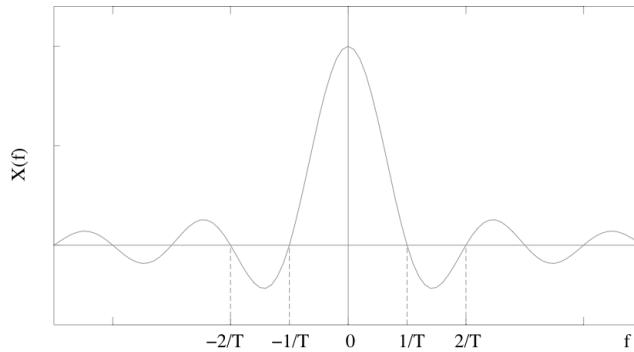


Figura 4.8: Spettro dell'impulso rettangolare.

Impulso sinc(·)

Consideriamo l'impulso sinc(·) di ampiezza A e banda B , cioè il segnale:

$$x(t) = A \text{sinc}(2Bt) \quad (4.20)$$

Ponendo nella trasformata di un impulso rettangolare (4.19) $T = 2B$, dalle proprietà di dualità e di linearità segue che:

$$A \text{sinc}(2Bt) \longleftrightarrow \frac{A}{2B} \Pi\left(\frac{f}{2B}\right)$$

Lo spettro di tale impulso è illustrato in fig. 4.9. Si osservi che l'impulso sinc(·) costituisce un primo esempio di segnale a banda rigorosamente limitata, cioè il suo spettro è identicamente zero al di fuori della gamma di frequenze $-B < f < B$ (banda).

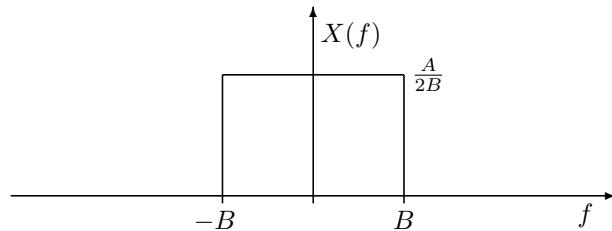


Figura 4.9: Spettro dell'impulso sinc(·).

Finestra rettangolare

Si consideri la finestra rettangolare $x(n) = \mathcal{R}_N(n)$; il suo spettro vale:

$$X(\nu) = \sum_{n=0}^{N-1} e^{-j2\pi\nu n} = \frac{1 - e^{-j2\pi\nu N}}{1 - e^{-j2\pi\nu}}$$

da cui, mettendo in evidenza $e^{-j\pi\nu N}$ al numeratore e $e^{-j\pi\nu}$ al denominatore, si ha:

$$\mathcal{R}_N(n) \longleftrightarrow \mathcal{D}_N(\nu) \quad (4.21)$$

ove si è introdotta la funzione di Dirichlet:

$$\mathcal{D}_N(\nu) = \frac{\sin(\pi\nu N)}{\sin(\pi\nu)} e^{-j(N-1)\pi\nu} \quad (4.22)$$

Gli andamenti dello spettro d'ampiezza e di quello di fase della finestra rettangolare sono riportati in fig. 4.10: da tale figura si notano alcune analogie, ma anche alcune differenze, con la trasformata dell'impulso rettangolare continuo.

Infatti, poiché la finestra rettangolare è stata definita in modo non simmetrico rispetto all'origine, la sua trasformata non è puramente reale, ma presenta una fase lineare a tratti. L'ampiezza A_1 del primo lobo laterale, rapportata all'ampiezza A_0 del lobo centrale, è leggermente dipendente da N , secondo la Tab. 4.1: per $N > 10$ il rapporto $(A_1/A_0)_{dB}$ si può

N	3	4	5	10	∞
$(A_1/A_0)_{dB}$	-9,54	-11,30	-12,04	-12,17	-13,26

Tabella 4.1

Valori in dB del rapporto A_1/A_0 al variare di N .

ritenere approssimativamente costante e pari a $-13 dB$, tendendo asintoticamente ($N \rightarrow \infty$) a $-13,26 dB$, cioè al valore che compete all'impulso rettangolare a tempo continuo.

Impulsi ideali

Lo spettro del δ -impulso discreto $\delta(n)$ vale:

$$X(\nu) = \sum_{n=-\infty}^{+\infty} \delta(n) e^{-j2\pi\nu n} = 1$$

e quindi:

$$\delta(n) \longleftrightarrow 1$$

Analogamente nel caso dell'impulso ideale continuo si ha:

$$\delta(t) \longleftrightarrow 1$$

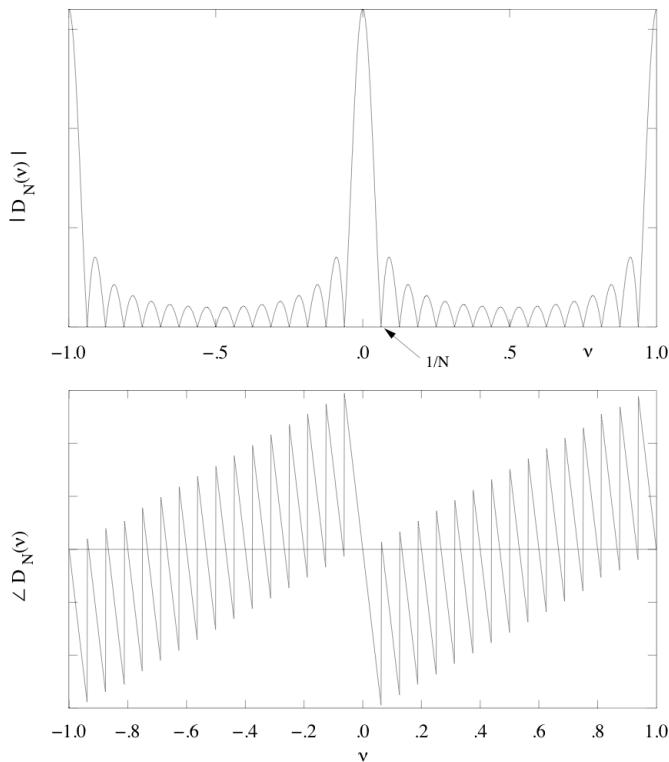


Figura 4.10: Spettro della finestra rettangolare.

Le corrispondenti equazioni di sintesi forniscono due proprietà degli esponenziali:

$$\delta(n) = \int_{-1/2}^{1/2} e^{j2\pi n\nu} d\nu \quad (4.23)$$

e rispettivamente:

$$\delta(t) = \int_{-\infty}^{+\infty} e^{j2\pi ft} df \quad (4.24)$$

Ovviamente le (4.23) e (4.24) valgono indipendentemente dall'interpretazione di n e t come tempi e di ν e f come frequenze: in particolare, invertendo i ruoli di t e di f nella (4.24) (dualità) e tenendo conto che l'impulso di Dirac è pari, si ha anche:

$$\delta(f) = \int_{-\infty}^{+\infty} e^{-j2\pi ft} dt \quad (4.25)$$

La (4.25) è utile perché consente di ottenere la trasformata di Fourier (in senso generalizzato) di un segnale continuo e costante nel tempo, il quale non è dotato di trasformata di Fourier in senso ordinario. Infatti, posto $x(t) = A$, si può scrivere:

$$\int_{-\infty}^{+\infty} Ae^{-j2\pi ft} dt = A\delta(f)$$

e quindi si ottiene la coppia di trasformate:

$$A \longleftrightarrow A\delta(f) = A2\pi\delta(\omega) \quad (4.26)$$

dove l'ultima uguaglianza segue dalla proprietà di cambiamento della scala degli impulsi di Dirac.

A commento dei vari esempi considerati osserviamo che in linea generale segnali lentamente variabili nel tempo presentano uno spettro molto concentrato in frequenza e viceversa segnali rapidamente variabili presentano componenti spettrali significative anche a frequenze molto elevate.

4.4 TRASFORMATA DI FOURIER AL LIMITE

L'ultimo esempio considerato mostra come l'impiego delle funzioni generalizzate consenta di ottenere coppie di trasformate di Fourier anche quando le condizioni di Dirichlet o di energia finita non sono soddisfatte; ciò evidenzia che è possibile definire una rappresentazione spettrale anche per i segnali di potenza e non solo per quelli di energia. L'estensione della trasformazione di Fourier in modo rigoroso richiede la *Teoria delle distribuzioni*, ma tale approccio esula dallo scopo di queste dispense; nel seguito ci limiteremo ad affrontare tale estensione in modo intuitivo.

Sia $x(t)$ il segnale di cui si vuole ottenere la trasformata e sia $w_T(t)$ una famiglia di segnali, usualmente denominati *finestre*, che al limite, quando il parametro T tende all'infinito, tende a diventare identicamente pari ad 1, cioè

$$\lim_{T \rightarrow \infty} w_T(t) = 1$$

Esempi di finestre sono l'impulso rettangolare $\Pi(t/T)$, e quello triangolare $\Lambda(t/T)$. Si consideri poi il segnale finestrato $x_T(t) = x(t)w_T(t)$ che, di norma, è dotato di spettro $X_T(f)$ in senso ordinario; si definisce trasformata di Fourier di $x(t)$ il limite (in senso generalizzato) di $X_T(f)$ per $T \rightarrow \infty$. Ricapitolando la generalizzazione della definizione di spettro per forme d'onda è

$$x(t) \xrightarrow{FIN.} x_T(t) \xrightarrow{F.T.} X_T(f) \xrightarrow{T \rightarrow \infty} X(f)$$

Analoga la definizione per sequenze

$$x(n) \xrightarrow{FIN.} x_N(n) \xrightarrow{F.T.} X_N(\nu) \xrightarrow{N \rightarrow \infty} X(\nu)$$

L'importanza dell'approccio utilizzato nel generalizzare la definizione di spettro è nel suo valore operativo: infatti lo spettro misurato con un analizzatore di spettro è di norma quello del segnale finestrato, con una finestra di durata sufficientemente grande.

Allo scopo di chiarire ulteriormente la definizione data consideriamo alcuni esempi.

Spettro di un segnale costante

Consideriamo un segnale costante $x(t) = A$; moltiplicandolo per la finestra rettangolare

$$w_T(t) = \Pi\left(\frac{t}{T}\right)$$

e ricordando che:

$$x_T(t) = A\Pi\left(\frac{t}{T}\right) \longleftrightarrow X_T(f) = AT \operatorname{sinc}(fT)$$

da cui, facendo divergere la durata della finestra si ricava

$$A \longleftrightarrow \lim_{T \rightarrow +\infty} AT \operatorname{sinc}(fT) = A\delta(f)$$

Si osservi che lo spettro di una costante è una riga (δ -impulso) a frequenza zero, mentre quello del segnale finestrato ha una banda non nulla (dispersione spettrale) tanto più piccola quanto maggiore è la durata della finestra; per una data durata, la precisione della misura dipende dal tipo di finestra utilizzato nel senso che, al variare del tipo di finestra utilizzata, variano la dispersione spettrale (banda) e la rapidità di decadimento a zero dello spettro del segnale finestrato.

Sequenza costante

Consideriamo la sequenza costante $x(n) = \tilde{\delta}(n)$, cioè la sequenza¹:

$$\tilde{\delta}(n) = \sum_{k=-\infty}^{+\infty} \delta(n - k)$$

quindi $\tilde{\delta}(n)$ è una successione d'impulsi unitari, cioè un segnale che vale identicamente uno, per qualsiasi n . Troncando tra $-N$ e N ($2N + 1$ campioni) ed utilizzando la procedura al

¹ Il simbolo \sim viene usato, qui e nel seguito, per indicare la versione periodica del segnale su cui opera.

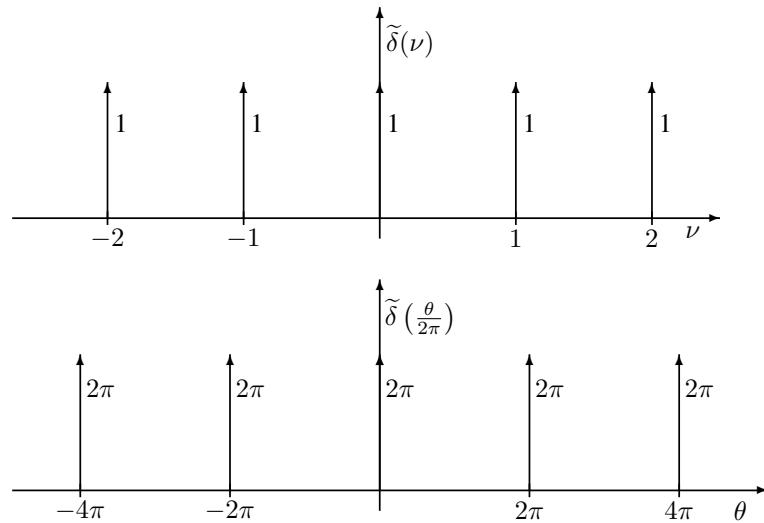


Figura 4.11: Spettro di una sequenza costante.

limite, si ottiene

$$\tilde{\delta}(n) = \sum_{k=-\infty}^{+\infty} \delta(n-k) \longleftrightarrow \tilde{\delta}(\nu) = \sum_{k=-\infty}^{+\infty} \delta(\nu-k) \quad (4.27)$$

come mostrato in figura 4.11

4.5 PROPRIETÀ DELLA TRASFORMATA DI FOURIER

In questo paragrafo analizziamo le proprietà della trasformata di Fourier che, per la maggior parte, valgono indifferentemente sia per le forme d'onda che per le sequenze; pertanto, per evitare inutili duplicazioni, le proprietà saranno espresse (quando possibile) con riferimento ad un argomento generico, denotato con il simbolo (\cdot) , che può essere specializzato nei vari casi d'interesse. Conformemente all'uso adottato sinora, i segnali nel dominio del tempo vengono denotati con la lettera minuscola e le loro trasformate con le corrispondenti maiuscole. Le proprietà vengono date senza dimostrazione, l'enfasi essendo piuttosto sulla loro applicazione.

4.5.1 Linearità

È immediato verificare che la trasformata di Fourier è lineare, cioè risulta:

$$a_1x_1(\cdot) + a_2x_2(\cdot) \longleftrightarrow a_1X_1(\cdot) + a_2X_2(\cdot) \quad (4.28)$$

In altri termini lo spettro di una combinazione lineare di segnali è la combinazione lineare, con gli stessi coefficienti, dei singoli spettri (principio di sovrapposizione).

4.5.2 Proprietà di simmetria della trasformata di Fourier

È facile controllare che ad una riflessione nel dominio del tempo corrisponde una riflessione nel dominio della frequenza e viceversa; cioè:

$$x(-(\cdot)) \longleftrightarrow X(-(\cdot)) \quad (4.29)$$

Pertanto un segnale pari ha spettro pari e viceversa. Inoltre ad una coniugazione in un dominio corrisponde una coniugazione più una riflessione nell'altro dominio; in altri termini si ha:

$$x^*(\cdot) \longleftrightarrow X^*(-(\cdot)) \quad x^*(-(\cdot)) \longleftrightarrow X^*(\cdot) \quad (4.30)$$

Come conseguenza si ha che se $x(\cdot)$ è reale, cioè $x(\cdot) = x^*(\cdot)$, allora per la (4.30) risulta $X(\cdot) = X^*(-(\cdot))$; una funzione che verifica questa proprietà si dice *Hermitiana* o *coniugata simmetrica*: quindi, in breve, se $x(\cdot)$ è reale, allora $X(\cdot)$ è Hermitiana. Evidentemente il carattere Hermitiano dello spettro è equivalente alla condizione che $\Re\{X(\cdot)\}$ sia pari e $\Im\{X(\cdot)\}$ sia dispari; ovvero è equivalente alla condizione che lo spettro d'ampiezza sia pari e quello di fase dispari. Quindi, quando si calcola o si rappresenta la trasformata di Fourier di un segnale reale, parte reale e immaginaria, ovvero modulo e fase, dello spettro possono essere calcolati o rappresentati solo per frequenze positive, poiché i valori per frequenze negative sono ricavabili per simmetria coniugata. Se $x(\cdot)$ oltre che reale è anche pari, allora dalla (4.29) segue che anche lo spettro è pari oltre che hermitiano e quindi lo spettro è puramente reale: pertanto se $x(\cdot)$ è reale e pari, anche $X(\cdot)$ è reale e pari.

4.5.3 Cambiamento di scala

Analizziamo l'effetto nel dominio della frequenza di un cambiamento di scala dell'asse dei tempi. Per segnali a tempo continuo è immediato verificare che si ha:

$$x(at) \longleftrightarrow \frac{1}{|a|}X\left(\frac{f}{a}\right) \quad (4.31)$$

La proprietà (4.31) comporta che ad una compressione nel dominio del tempo ($|a| > 1$) corrisponde un'espansione dello spettro e viceversa. Notiamo anche che oltre al cambiamento

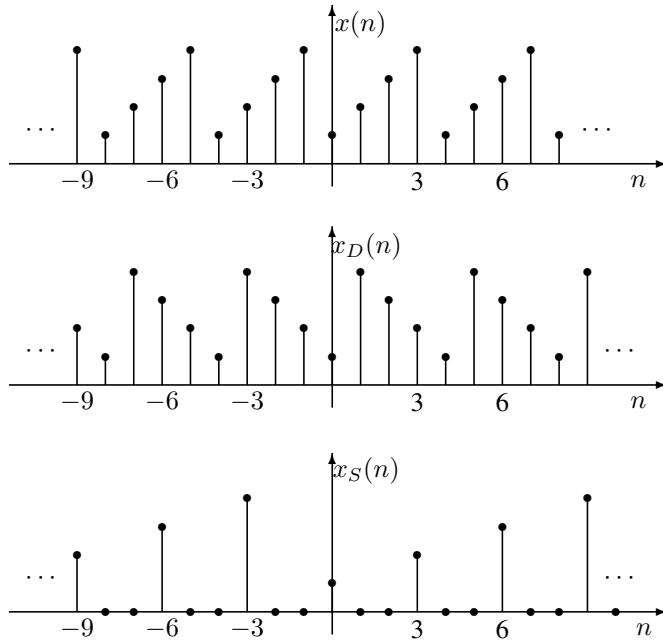


Figura 4.12: Sequenza originaria $x(n)$, decimata $x_D(n)$ ed espansa $x_S(n)$ per N ($N = 3$).

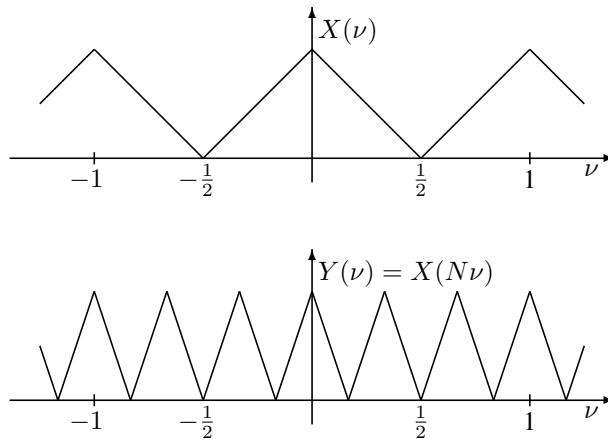
della scala delle frequenze si ha anche un cambiamento della scala delle ordinate (ampiezze) in modo che l'area sottesa dallo spettro resti costante: tale area infatti è pari al valore nell'origine $x(0)$, che non si modifica per effetto del cambiamento della scala dell'asse dei tempi.

Nel caso di sequenze, atteso il carattere discreto del tempo, l'operazione di cambiamento di scala va opportunamente interpretata. Precisamente, supposto il fattore di scala positivo, in quanto, analogamente a quanto si verifica per i segnali a tempo continuo, il segno del fattore di scala tiene conto di un eventuale ribaltamento in aggiunta al cambiamento di scala vero e proprio, la compressione dell'asse dei tempi porta a considerare la sequenza $x_D(n) = x(nN)$, cioè la sequenza:

$$\dots, x(-2N), x(-N), x(0), x(N), x(2N), \dots$$

ottenuta da $x(n)$ prendendo un campione ogni N : tale operazione è detta *decimazione* per N (fig. 4.12). Poiché la decimazione comporta in genere una perdita d'informazione, l'effetto sullo spettro non è un semplice cambiamento di scala come verrà mostrato più in seguito.

Consideriamo ora l'espansione dell'asse dei tempi, cioè prendiamo in esame la sequenza $x_S(n) = x[n/N]$, con N intero non nullo. La scrittura $x[n/N]$ è convenzionale e come debba essere interpretata è chiarito in fig. 4.12: per n multiplo di N , $x_S(n)$ è il campione

Figura 4.13: Effetto sullo spettro dell'espansione dell'asse dei tempi ($N = 3$).

della sequenza originaria all'istante n/N (ad esempio, con riferimento al caso di fig. 4.12, ove $N = 3$, $x_S(3) = x(1)$, $x_S(6) = x(2)$ e così via), mentre viene posto uguale a zero per gli altri valori di n ; in altri termini, quando il valore dell'argomento n/N non è intero, $x[n/N]$ è per convenzione zero; l'impiego delle parentesi quadre invece delle tonde nel denotare la sequenza espansa, con fattore di espansione N , e cioè la notazione $x[n/N]$ ricorda che trattasi di scrittura convenzionale. Trascurando i dettagli analitici, si ricava

$$x_S(n) = x\left[\frac{n}{N}\right] \longleftrightarrow X(N\nu)$$

La relazione ottenuta è quindi simile a quella valida per i segnali a tempo continuo, salvo il fatto che le ampiezze dello spettro non vengono alterate. L'effetto nel dominio della frequenza è illustrato nella fig. 4.13, sempre per il caso $N = 3$. Si noti che la trasformata, che originariamente è periodica di periodo 1, dopo il cambiamento di scala diventa periodica di periodo $1/3$; si noti inoltre che l'area totale nell'intervallo $-1/2 \leq \nu \leq 1/2$ resta immutata.

Come applicazione del cambiamento di scala è immediato valutare lo spettro del segnale campionatore ideale discreto di periodo qualsiasi.

Esempio 2: Segnale campionatore ideale discreto di periodo N

Il treno campionatore ideale discreto di periodo N è il segnale:

$$\tilde{\delta}_N(n) = \sum_{k=-\infty}^{+\infty} \delta(n - kN)$$

esso è cioè un treno d'impulsi discreti di ampiezza unitaria localizzati negli istanti kN , con k intero relativo (fig. 4.14a). Poiché tale segnale può essere riguardato come

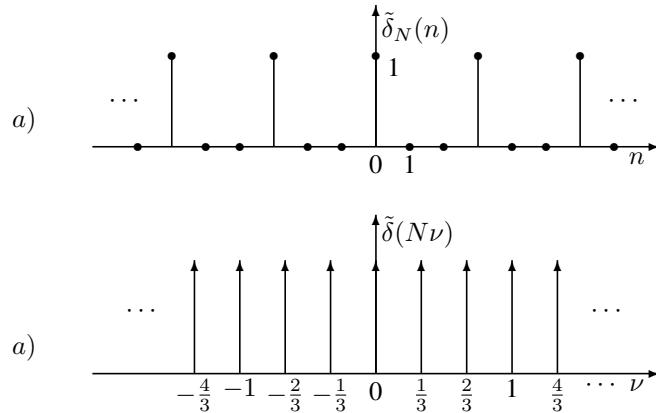


Figura 4.14: Il segnale campionatore discreto e il suo spettro.

un'espansione della sequenza costante $\tilde{\delta}(n)$ avendosi:

$$\tilde{\delta}_N(n) = \tilde{\delta}\left(\frac{n}{N}\right)$$

applicando la proprietà di cambiamento di scala alla coppia segnale-spettro (4.27) si ottiene:

$$\begin{aligned} \tilde{\delta}_N(n) &= \sum_{k=-\infty}^{+\infty} \delta(n - kN) \\ &\quad \downarrow \\ \tilde{\delta}(N\nu) &= \frac{1}{N} \tilde{\delta}_{\frac{1}{N}}(\nu) = \frac{1}{N} \sum_{k=-\infty}^{+\infty} \delta\left(\nu - \frac{k}{N}\right) \end{aligned} \quad (4.32)$$

cioè lo spettro di un segnale campionatore ideale discreto di periodo N è ancora un treno campionatore ideale, precisamente è un treno periodico di δ -impulsi di area e periodo entrambi pari a $1/N$ (fig. 4.14b). Il fattore d'ampiezza introdotto deriva dalla proprietà di cambiamento di scala degli impulsi di Dirac, e comunque è tale da lasciare inalterata l'area della trasformata nell'intervallo $-1/2 \leq \nu \leq 1/2$.

4.5.4 Traslazione nel dominio del tempo

È facile verificare che:

$$x(t - T) \longleftrightarrow X(f)e^{-j2\pi fT} \quad x(n - N) \longleftrightarrow X(\nu)e^{-j2\pi\nu N}$$

Quindi un ritardo introduce uno sfasamento lineare con ma non influisce sullo spettro d'ampiezza. Si osservi che nel caso di segnali discreti N è generalmente un numero intero,

tuttavia la proprietà rimane formalmente valida anche se N è un generico numero reale. Come applicazione di questa proprietà consideriamo il seguente importante esempio.

Esempio 3: *Segnale campionatore ideale di periodo T*

Iniziamo col considerare il *segnale campionatore ideale* di periodo unitario $\tilde{\delta}(t)$, cioè il treno d'impulsi di Dirac

$$\tilde{\delta}(t) = \sum_{k=-\infty}^{+\infty} \delta(t - k) \quad (4.33)$$

Applicando la procedura di trasformazione al limite, si ottiene

$$\tilde{\delta}(t) = \sum_{k=-\infty}^{+\infty} \delta(t - k) \iff \tilde{\delta}(f) = \sum_{m=-\infty}^{+\infty} \delta(f - m) \quad (4.34)$$

cioè ad un treno d'impulsi di Dirac di area e periodo unitari nel tempo corrisponde un treno d'impulsi di Dirac di area e periodo unitari nella frequenza.

Consideriamo ora il segnale campionatore ideale continuo di periodo T (fig. 4.15a), cioè un treno d'impulsi di Dirac di area unitaria localizzati negli istanti $t = kT$:

$$\tilde{\delta}_T(t) = \sum_{k=-\infty}^{+\infty} \delta(t - kT)$$

Poiché tale segnale può ottersi cambiando la scala sia delle ampiezze che dei tempi del segnale campionatore ideale di periodo unitario (4.33) avendosi:

$$\tilde{\delta}_T(t) = \frac{1}{T} \tilde{\delta}\left(\frac{t}{T}\right)$$

applicando la proprietà (4.31) alla coppia segnale-spettro (4.34) si ottiene:

$$\begin{aligned} \tilde{\delta}_T(t) &= \sum_{k=-\infty}^{+\infty} \delta(t - kT) \\ &\quad \downarrow \\ \tilde{\delta}(fT) &= \frac{1}{T} \tilde{\delta}_{\frac{1}{T}}(f) = \frac{1}{T} \sum_{k=-\infty}^{+\infty} \delta\left(f - \frac{k}{T}\right) \end{aligned} \quad (4.35)$$

Lo spettro del segnale campionatore ideale di periodo T è illustrato in fig. 4.15b. ed è costituito da una successione di righe spettrali, tutte di area $\frac{1}{T}$, equispaziate in frequenza di $\frac{1}{T}$.

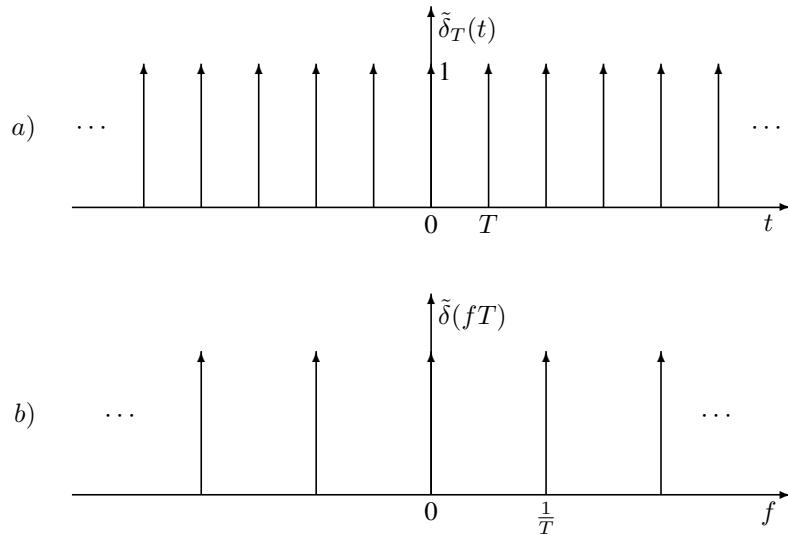


Figura 4.15: Il segnale campionatore ideale e il suo spettro.

4.5.5 Traslazione nel dominio della frequenza (modulazione)

È immediato verificare che:

$$x(t)e^{j2\pi f_c t} \longleftrightarrow X(f - f_c) \quad x(n)e^{j2\pi\nu_c n} \longleftrightarrow X(\nu - \nu_c) \quad (4.36)$$

Esempio 4: Trasformata di un fasore

Un fasore continuo $Ae^{j2\pi f_c t}$, tenuto conto della proprietà espressa dalla (4.36) e della trasformata di una costante (4.26), ha come spettro l'impulso di Dirac $A\delta(f - f_c)$; in altri termini si ha (fig. 4.16a):

$$Ae^{j2\pi f_c t} \longleftrightarrow A\delta(f - f_c)$$

Dunque lo spettro di un fasore è una riga alla frequenza f_c del fasore di ampiezza (area) pari all'ampiezza del fasore. Analogamente, un fasore discreto $Ae^{j2\pi\nu_c n}$ si trasforma traslando in frequenza lo spettro della sequenza costante $A\tilde{\delta}(n)$, pertanto si ha:

$$Ae^{j2\pi\nu_c n} \longleftrightarrow A\tilde{\delta}(\nu - \nu_c) \quad (4.37)$$

In modo analogo si può valutare lo spettro di un segnale modulato in ampiezza, cioè del prodotto di un segnale $x(t)$ per una sinusoida (*modulazione*). Infatti, dalla formula di Eulero

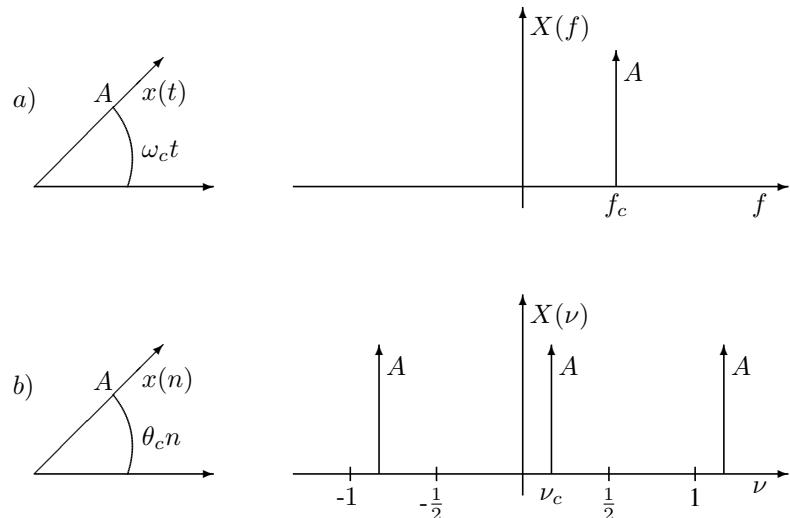


Figura 4.16: Spettro di un fasore: a) continuo b) discreto.

e dalla proprietà di linearità, segue facilmente che:

$$x(t) \cos(2\pi f_c t + \varphi_c) \longleftrightarrow \frac{1}{2} X(f - f_c) e^{j\varphi_c} + \frac{1}{2} X(f + f_c) e^{-j\varphi_c}$$

Nel contesto della modulazione il segnale \$x(t)\$ è detto segnale *modulante* e la sinusode segnale *portante*. L'effetto della modulazione è illustrata nella fig. 4.17 nel caso di un segnale modulante a banda limitata di tipo passa-basso, cioè \$X(f) = 0, \forall |f| > B\$, e una frequenza portante \$f_c\$ molto maggiore della banda (monolatera) \$B\$ del segnale modulante. Dopo la modulazione lo spettro risulta concentrato intorno alle frequenze \$+f_c\$ e \$-f_c\$, quindi il segnale modulato è di tipo *passa-banda* o, come anche si dice nel caso di \$f_c\$ sufficientemente elevata, *a radiofrequenza* (RF). Si noti inoltre che, per effetto della modulazione, l'occupazione totale di banda è raddoppiata, passando da \$B\$ (monolatera) per il segnale modulante a \$2B\$ per il segnale RF.

Esempio 5: Scambio di alte e basse frequenze

Consideriamo il segnale modulato \$x(n)(-1)^n\$. Poiché la portante \$(-1)^n\$ può essere equivalentemente riscritta come

$$(-1)^n = \cos(\pi n) = e^{j\pi n}$$

e quindi è un fasore di frequenza \$\nu = 1/2\$, dalla proprietà di traslazione in frequenza segue che:

$$x(n)(-1)^n \longleftrightarrow X\left(\nu - \frac{1}{2}\right)$$

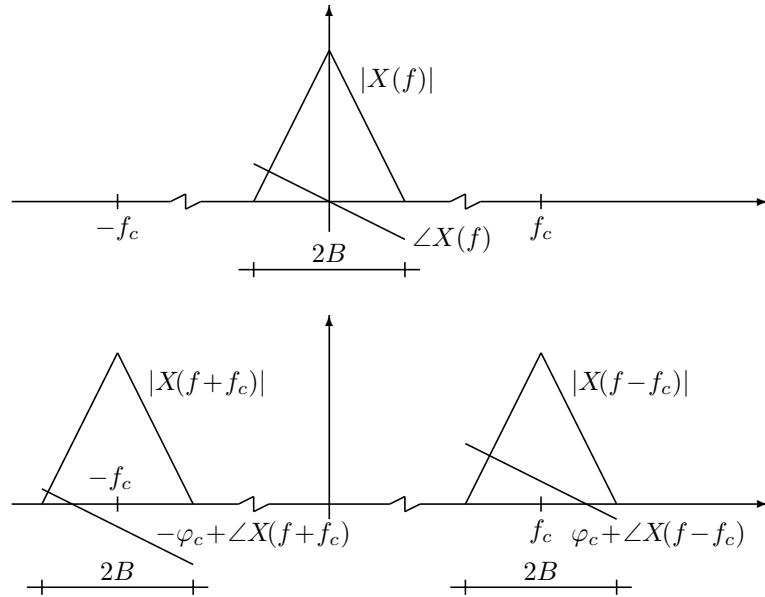


Figura 4.17: Spettri del segnale modulante (a) e del segnale modulato (b).

Nell'ipotesi che $x(n)$ sia il segnale passa-basso di fig. 4.18a, il segnale modulato ha lo spettro passa-alto di fig. 4.18b; viceversa, se $x(n)$ è il segnale passa-alto di fig. 4.18b, il segnale modulato ha lo spettro passa-basso di fig. 4.18a. In altri termini, se un segnale viene modulato con una portante di tipo $(-1)^n$, il che cambia di segno ai campioni del segnale negli istanti dispari, l'effetto in frequenza è di traslare lo spettro periodico del segnale di mezzo periodo, e quindi, tenendo conto della periodicità, di scambiare fra loro le regioni dello spettro corrispondenti alle basse ed alle alte frequenze.

Esempio 6: Impulso RF

L'impulso a radiofrequenza con inviluppo rettangolare, cioè il segnale

$$x(t) = A\Pi\left(\frac{t}{T}\right) \cos(2\pi f_c t)$$

è illustrato schematicamente in fig. 4.19; dalla (4.19) e dalla proprietà di modulazione segue che:

$$A\Pi\left(\frac{t}{T}\right) \cos(2\pi f_c t) \longleftrightarrow \frac{1}{2}AT \operatorname{sinc}[(f - f_c)T] + \frac{1}{2}AT \operatorname{sinc}[(f + f_c)T]$$

Lo spettro dell'impulso RF è riportato in fig. 4.19; al limite per $T \rightarrow +\infty$ si ottiene la trasformata del coseno, cioè due impulsi collocati in $f = \pm f_c$ di area $\frac{1}{2}A$, risulta

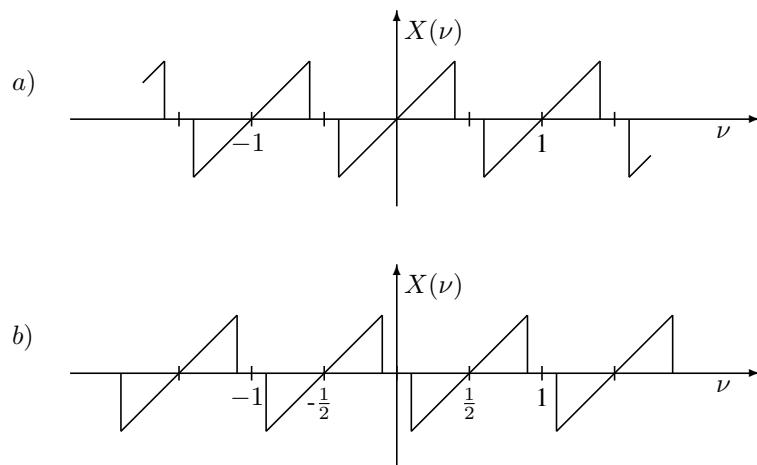


Figura 4.18: Scambio alte e basse frequenze.

pertanto:

$$A \cos(2\pi f_c t) \longleftrightarrow \frac{1}{2}A\delta(f - f_c) + \frac{1}{2}A\delta(f + f_c)$$

Esempio 7: L'impulso cosinusoidale

L'impulso cosinusoidale è il segnale:

$$x(t) = A\Pi\left(\frac{t}{T}\right) \cos\left(\pi\frac{t}{T}\right)$$

Pertanto, ponendo $f_c = \frac{1}{2T}$ nell'espressione dello spettro dell'impulso RF, si ottiene:

$$A\Pi\left(\frac{t}{T}\right) \cos\left(\pi\frac{t}{T}\right) \longleftrightarrow \frac{1}{2}AT \operatorname{sinc}\left(fT - \frac{1}{2}\right) + \frac{1}{2}AT \operatorname{sinc}\left(fT + \frac{1}{2}\right)$$

ovvero, esplicitando la funzione $\operatorname{sinc}(\cdot)$ e semplificando

$$A\Pi\left(\frac{t}{T}\right) \cos\left(\pi\frac{t}{T}\right) \longleftrightarrow \frac{2AT}{\pi} \frac{\cos(\pi fT)}{1 - (2fT)^2}$$

Questa coppia di trasformate è illustrata in fig. 4.20.

Rispetto ad un impulso rettangolare di pari durata la banda (larghezza del lobo principale) è maggiore ($3/T$ invece di $2/T$), ma, in compenso, i lobi secondari sono attenuati di almeno $\alpha_{SL} = 23$ dB rispetto all'ampiezza del lobo principale, e decadono a zero più rapidamente, precisamente come f^{-2} , corrispondente a 12 dB/ott.

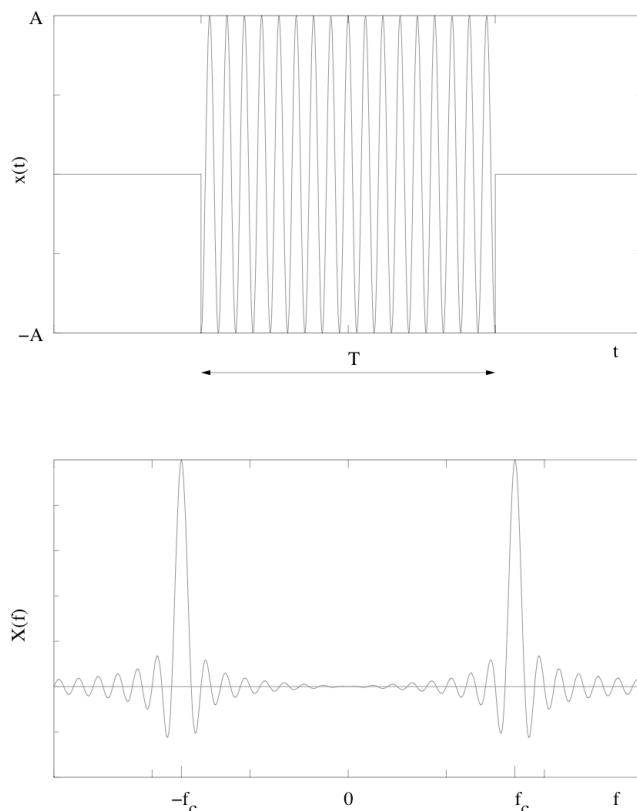


Figura 4.19: L'impulso RF e il suo spettro.

4.5.6 Convoluzione

Questa è una delle proprietà più importanti della trasformata di Fourier, per i suoi riflessi sull'analisi dei sistemi; la formulazione di tale proprietà, valida sia per sequenze che per segnali continui, è la seguente:

$$x(\cdot) * y(\cdot) \longleftrightarrow X(\cdot)Y(\cdot) \quad (4.38)$$

In altri termini, alla convoluzione di due segnali nel dominio del tempo corrisponde il prodotto dei rispettivi spettri nel dominio della frequenza; si osservi che, in virtù della proprietà associativa della convoluzione e del prodotto, la proprietà è valida anche per un numero di segnali maggiore di due, ma finito.

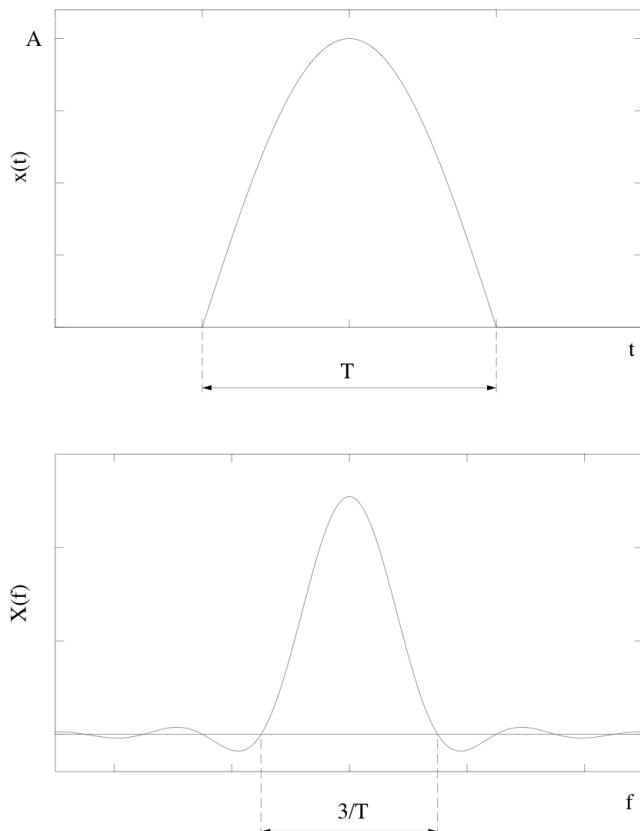


Figura 4.20: L'impulso cosinusoidale e il suo spettro.

La proprietà duale di quella della convoluzione è la proprietà di *modulazione generalizzata*: tale proprietà afferma che ad un prodotto nel dominio del tempo corrisponde una convoluzione nel dominio della frequenza. Precisamente per i segnali a tempo continuo risulta:

$$x(t)y(t) \longleftrightarrow X(f) * Y(f) = \frac{1}{2\pi} X(\omega) * Y(\omega) \quad (4.39)$$

Tale proprietà è sempre applicabile quando il prodotto nel dominio del tempo è ben definito (quindi può non essere applicabile quando ambedue i segnali contengono impulsi di Dirac) e quando essi sono separatamente trasformabili, eventualmente anche in senso generalizzato (infatti è comunque ben definita la convoluzione fra funzioni generalizzate).

Per segnali a tempo discreto la proprietà si enuncia allo stesso modo, ma l'operazione di convoluzione è diversa: precisamente è possibile dimostrare che lo spettro del prodotto di

due sequenze è legato a quello delle singole sequenze da:

$$x(n)y(n) \longleftrightarrow X(\nu) * Y(\nu) = \int_{-1/2}^{+1/2} X(u)Y(\nu - u)du \quad (4.40)$$

Pertanto nel caso di sequenze la convoluzione degli spettri è l'operazione definita a secondo membro della (4.40): tale operazione, anche se denotata con lo stesso simbolo della convoluzione, differisce da quest'ultima in quanto l'integrale va esteso al periodo e diviso per il periodo stesso: per sottolineare tale diversità la convoluzione relativa a segnali periodici dello stesso periodo è detta *convoluzione periodica*. L'impiego dello stesso simbolo sottolinea però la similitudine tra le due operazioni, ed invero la convoluzione periodica gode delle stesse proprietà della convoluzione, opportuna attenzione però deve essere posta sul fatto che i fattori devono essere segnali periodici dello stesso periodo, che è anche il periodo della convoluzione.

4.5.7 Replicazione e campionamento

Si definisce la replicazione di periodo T di $x(t)$ il segnale:

$$\tilde{x}(t) = \text{rep}_T[x(t)] = \sum_{k=-\infty}^{+\infty} x(t - kT) \quad (4.41)$$

e, analogamente, nel caso di sequenze, si pone:

$$\tilde{x}(n) = \text{rep}_N[x(n)] = \sum_{k=-\infty}^{+\infty} x(n - kN) \quad (4.42)$$

Il segnale $x(\cdot)$ che viene replicato è detto segnale generatore.

Si osservi che l'operazione di replicazione non è ben definita per un qualsiasi generatore $x(\cdot)$ come è immediato verificare utilizzando, per esempio, un segnale $x(\cdot)$ costante o periodico. Affinché l'operazione abbia senso è sufficiente che il generatore sia un segnale d'energia, ipotesi che riterremo senz'altro verificata nel seguito; tuttavia è di interesse utilizzare come generatore un impulso di Dirac: in tal caso infatti il segnale replicato è il segnale campionatore ideale, $\tilde{\delta}_T(t)$ o $\tilde{\delta}_N(n)$.

Notiamo esplicitamente che il segnale $\tilde{x}(\cdot)$, ottenuto per replicazione, è sempre un segnale periodico con periodo pari a quello di replica, T o N a seconda del caso. A scopo esemplificativo nella fig. 4.21 sono riportati i segnali ottenuti replicando l'impulso triangolare $\Lambda(t)$ di fig. 4.21a con periodo $T = 3/2$ (fig. 4.21b), $T = 2$ (fig. 4.21c) e $T = 3$ (fig. 4.21d).

Dato un segnale $x(\cdot)$, la sua versione campionata (idealmente) è il segnale (cfr. fig. 4.22):

$$x_\delta(t) = \sum_{k=-\infty}^{+\infty} x(kT)\delta(t - kT) \quad (4.43)$$

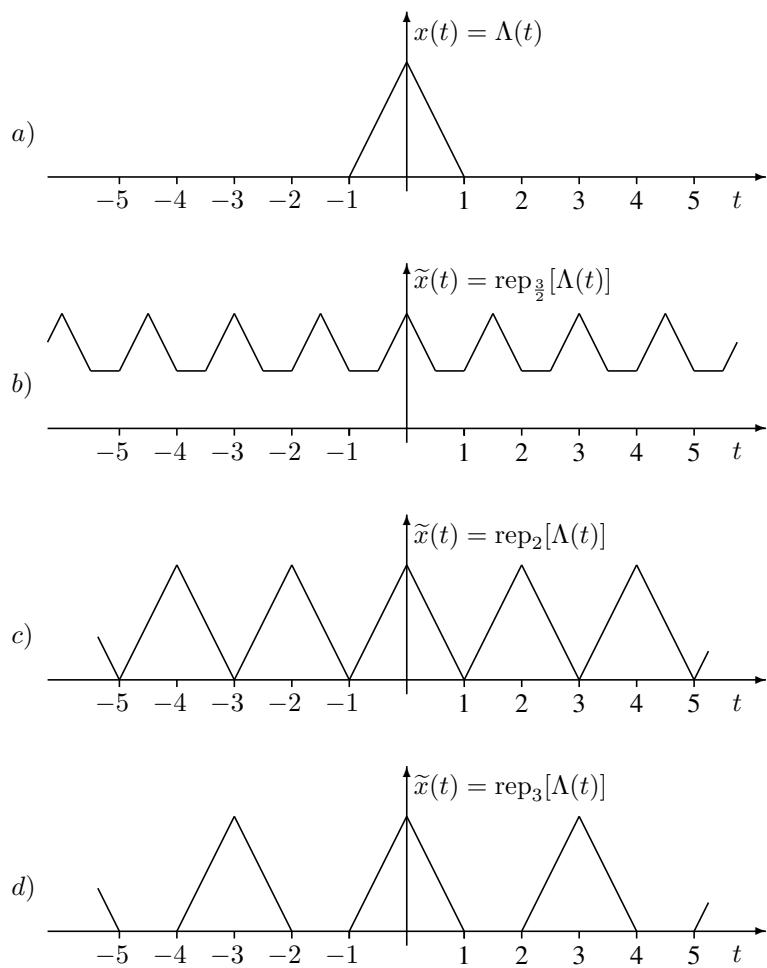


Figura 4.21: Replicazione di un impulso triangolare.

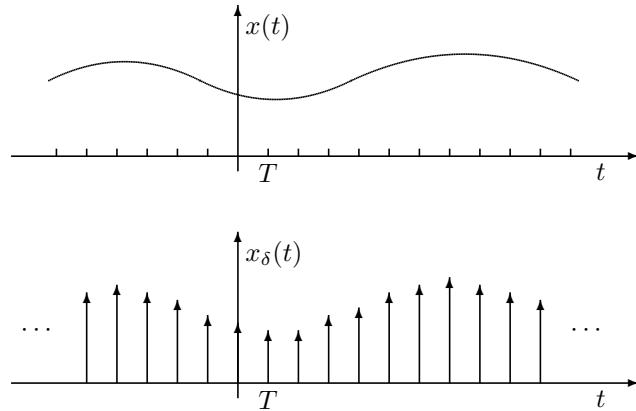


Figura 4.22: Il campionamento ideale di una forma d'onda.

nel caso di segnali a tempo continuo, mentre nel caso di sequenze si pone (cfr. fig. 4.23):

$$x_\delta(n) = \sum_{k=-\infty}^{+\infty} x(kN) \delta(n - kN) \quad (4.44)$$

T (risp. N) è detto periodo o passo di campionamento, ed il suo reciproco frequenza o cadenza di campionamento.

È utile riguardare le operazioni di replicazione e di campionamento come sistemi (fig. 4.24). La replicazione (fig. 4.24a) può essere realizzata con un sistema LTI in quanto le relazioni (4.41) e (4.42) possono essere riscritte come convoluzione; precisamente, come già accennato e come è immediato verificare, si ha:

$$\tilde{x}(t) = \text{rep}_T[x(t)] = x(t) * \tilde{\delta}_T(t) \quad \tilde{x}(n) = \text{rep}_N[x(n)] = x(n) * \tilde{\delta}_N(n) \quad (4.45)$$

cioè la replicazione del segnale $x(\cdot)$ la si ottiene effettuandone la convoluzione con un treno campionatore ideale, che pertanto è la risposta impulsiva. Anche il campionamento è un sistema lineare, ma temporalmente variante, avendosi

$$x_\delta(t) = x(t) \cdot \tilde{\delta}_T(t) \quad x_\delta(n) = x(n) \cdot \tilde{\delta}_N(n) \quad (4.46)$$

Usando la proprietà di convoluzione (Eq. 4.39 e Eq. 4.40) e la trasformata del segnale campionatore ideale (Eq. 4.35 e Eq. 4.32), trasformando la (4.45) si ottiene:

$$\sum_{k=-\infty}^{+\infty} x(t - kT) \longleftrightarrow \frac{1}{T} \sum_{k=-\infty}^{+\infty} X\left(\frac{k}{T}\right) \delta\left(f - \frac{k}{T}\right) \quad (4.47)$$

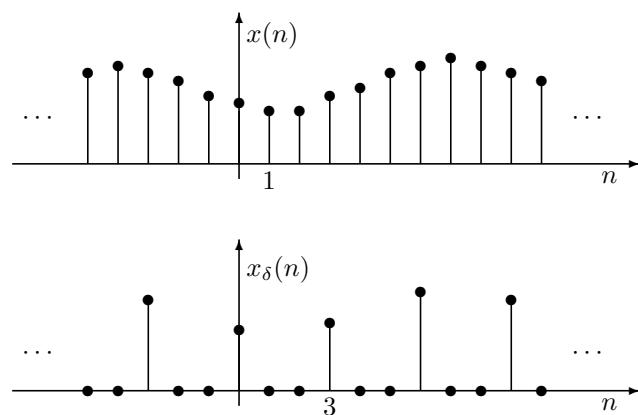


Figura 4.23: Il campionamento ideale di una sequenza.

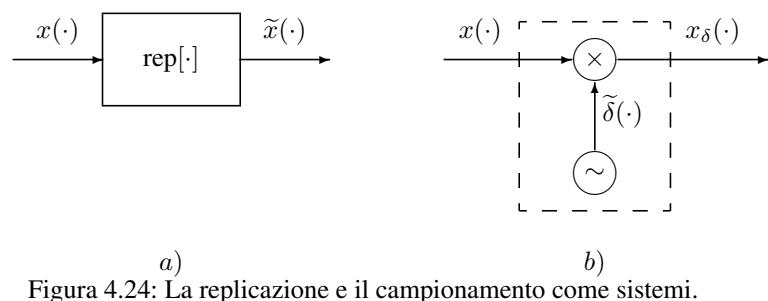


Figura 4.24: La replicazione e il campionamento come sistemi.

nel caso di segnali continui, e la:

$$\sum_{k=-\infty}^{+\infty} x(n - kN) \longleftrightarrow \frac{1}{N} \sum_{k=-\infty}^{\infty} X\left(\frac{k}{N}\right) \delta\left(\nu - \frac{k}{N}\right) \quad (4.48)$$

nel caso di sequenze. Le relazioni precedenti si esprimono sinteticamente dicendo che ad una replicazione nel dominio del tempo corrisponde un campionamento in frequenza.

Similmente, trasformando la relazione (4.46) si ottiene:

$$X_\delta(f) = X(f) * \frac{1}{T} \sum_{k=-\infty}^{+\infty} \delta\left(f - \frac{k}{T}\right) = \frac{1}{T} \sum_{k=-\infty}^{+\infty} X\left(f - \frac{k}{T}\right)$$

ovvero:

$$x_\delta(t) = \sum_{k=-\infty}^{+\infty} x(kT) \delta(t - kT) \longleftrightarrow X_\delta(f) = \frac{1}{T} \sum_{k=-\infty}^{+\infty} X\left(f - \frac{k}{T}\right)$$

nel caso di forme d'onda, e, analogamente nel caso di sequenze, si ha:

$$X_\delta(\nu) = X(\nu) * \frac{1}{N} \sum_{k=-\infty}^{\infty} \delta\left(\nu - \frac{k}{N}\right) = \frac{1}{N} \sum_{k=-\infty}^{\infty} X\left(\nu - \frac{k}{N}\right)$$

4.5.8 Spettro di segnali periodici

Si è osservato che la replicazione $\tilde{x}(\cdot)$ di $x(\cdot)$ è un segnale periodico di periodo pari a quello di replica; d'altra parte un qualunque segnale periodico $\tilde{x}(\cdot)$ può essere sempre considerato come la replicazione di un opportuno generatore $x(\cdot)$: infatti è sufficiente scegliere come generatore $x(\cdot)$ la restrizione di $\tilde{x}(\cdot)$ ad un periodo. È però importante notare che la corrispondenza così istituita fra segnale periodico e segnale generatore non è biunivoca: precisamente, dato un segnale $x(\cdot)$, è univocamente definito il segnale periodico $\tilde{x}(\cdot) = \text{rep}[x(\cdot)]$, ma, dato un segnale periodico, esistono diversi generatori $x(\cdot)$ tali che $\tilde{x}(\cdot) = \text{rep}[x(\cdot)]$. Ad esempio, il segnale periodico $\tilde{x}(t)$ di fig. 4.25c, oltre che pensarsi generato dalla replicazione del segnale triangolare di fig. 4.25a, può anche ottersi dalla replicazione della sua restrizione al periodo, rappresentata in fig. 4.25b. Dall'esempio riportato dovrebbe risultare chiaro che la non invertibilità della corrispondenza deriva dalla interazione, all'interno del periodo principale, delle repliche collocate nei periodi contigui. Tale interazione non è evidentemente possibile se imponiamo la condizione che il segnale generatore sia di durata rigorosamente limitata e minore o uguale al periodo del segnale periodico ed allocato tra 0 e T (o tra $-T/2$ e $T/2$): con questa restrizione la corrispondenza istituita è dunque invertibile.

L'utilità della rappresentazione (4.41) per una forma d'onda periodica (non importa se invertibile o meno) sta nel poterne determinare immediatamente lo spettro. Infatti, ricordando

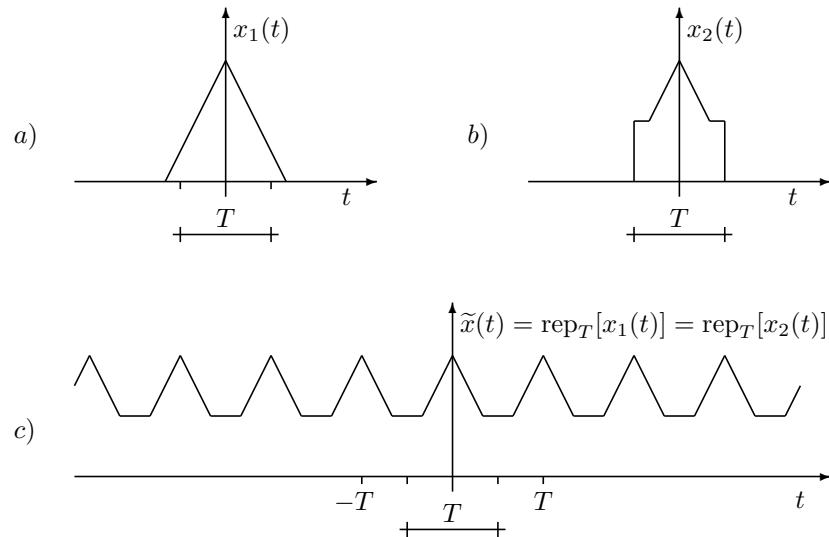


Figura 4.25: Rappresentazione di un segnale periodico come replicazione.

che ad una replicazione nel tempo corrisponde un campionamento in frequenza (4.47), si ha:

$$\tilde{x}(t) \longleftrightarrow \sum_{k=-\infty}^{+\infty} \frac{1}{T} X\left(\frac{k}{T}\right) \delta\left(f - \frac{k}{T}\right) \quad (4.49)$$

dove $X(f)$ è lo spettro del segnale generatore $x(t)$. La relazione (4.49) vale qualunque sia il generatore $x(t)$ di $\tilde{x}(t)$ secondo la (4.41): al variare di $x(t)$ anche $X(f)$ ovviamente varia, ma restano costanti i suoi valori alle frequenze $f = k/T$, con k intero, proprio in conseguenza del vincolo che $x(t)$ sia generatore di $\tilde{x}(t)$.

In maniera analoga, e con le stesse avvertenze, per una sequenza periodica di periodo N , tenuto conto che la si può rappresentare nella forma (4.42), dalla (4.48) si ha:

$$\tilde{x}(n) \longleftrightarrow \frac{1}{N} \sum_{k=-\infty}^{\infty} X\left(\frac{k}{N}\right) \delta\left(\nu - \frac{k}{N}\right) \quad (4.50)$$

Dunque lo spettro di un segnale periodico è uno spettro a righe, ogni riga essendo rappresentata da un impulso di Dirac, equispaziati in frequenza di $1/T$ (risp. $1/N$), le cui aree – a meno della costante di proporzionalità $1/T$ (risp. $1/N$) – seguono l'inviluppo $X(f)$ (risp. $X(\nu)$); in altri termini, lo spettro di un segnale periodico si ottiene campionando uniformemente in frequenza – con un treno campionatore ideale di periodo $1/T$ (risp. $1/N$) – lo spettro $X(f)$ (risp. $X(\nu)$) di un qualunque generatore e scalando i campioni secondo il fattore $1/T$ (risp. $1/N$).

Esempio 1: *Treno di impulsi rettangolari*

Si consideri il treno d'impulsi rettangolari di ampiezza A , durata Δ , periodo T , con $T > \Delta$ (fig. 4.26a): tale treno può riguardarsi come la replicazione, con periodo T , dell'impulso $A\Pi(t/\Delta)$, pertanto si ottiene

$$X\left(\frac{k}{T}\right) = A\Delta \text{sinc}(f\Delta) \Big|_{f=\frac{k}{T}} = A\Delta \text{sinc}\left(k\frac{\Delta}{T}\right)$$

ed il suo spettro vale

$$\tilde{x}(t) \longleftrightarrow \sum_{k=-\infty}^{+\infty} A\frac{\Delta}{T} \text{sinc}\left(k\frac{\Delta}{T}\right) \delta\left(f - \frac{k}{T}\right)$$

Tale spettro, è funzione essenzialmente del rapporto Δ/T , il cosiddetto *ciclo di servizio* (duty cycle) del treno. Lo spettro del treno d'impulsi rettangolari è riportato nella fig. 4.26b.

4.6 SPETTRI DI ENERGIA

Sia $x(t)$ un segnale di energia a tempo continuo; dalla relazione di Parseval si ha:

$$\mathcal{E}_x = \int_{-\infty}^{+\infty} |x(t)|^2 dt = \int_{-\infty}^{+\infty} |X(f)|^2 df$$

Analogamente per un segnale $x(n)$ tempo discreto si ha

$$\mathcal{E}_x = \sum_{-\infty}^{+\infty} |x(n)|^2 = \int_{-\frac{1}{2}}^{-\frac{1}{2}} |X(\nu)|^2 d\nu$$

L'importanza di tali relazioni non sta tanto nella possibilità di calcolare l'energia totale (anzi il calcolo nel dominio del tempo risulta quasi sempre più semplice) quanto nel riconoscere che la funzione $|X(f)|^2$ dà conto della distribuzione in frequenza dell'energia.

Dalla definizione data segue immediatamente che l' \mathcal{ESD} (Energy Spectral Density) è una funzione reale e non negativa; inoltre essa è pari per segnali reali. Sempre dalla definizione segue anche che l' \mathcal{ESD} è invariante rispetto a qualsiasi operazione sul segnale che ne lasci immutato lo spettro d'ampiezza. Pertanto la corrispondenza

$$\text{segnaile} \longrightarrow \mathcal{ESD}$$

non è biunivoca. Ad esempio una traslazione di un segnale nel dominio del tempo non ne altera l' \mathcal{ESD} ; analogamente, un ribaltamento seguito da una eventuale coniugazione non ha influenza sullo spettro di energia di un segnale reale.

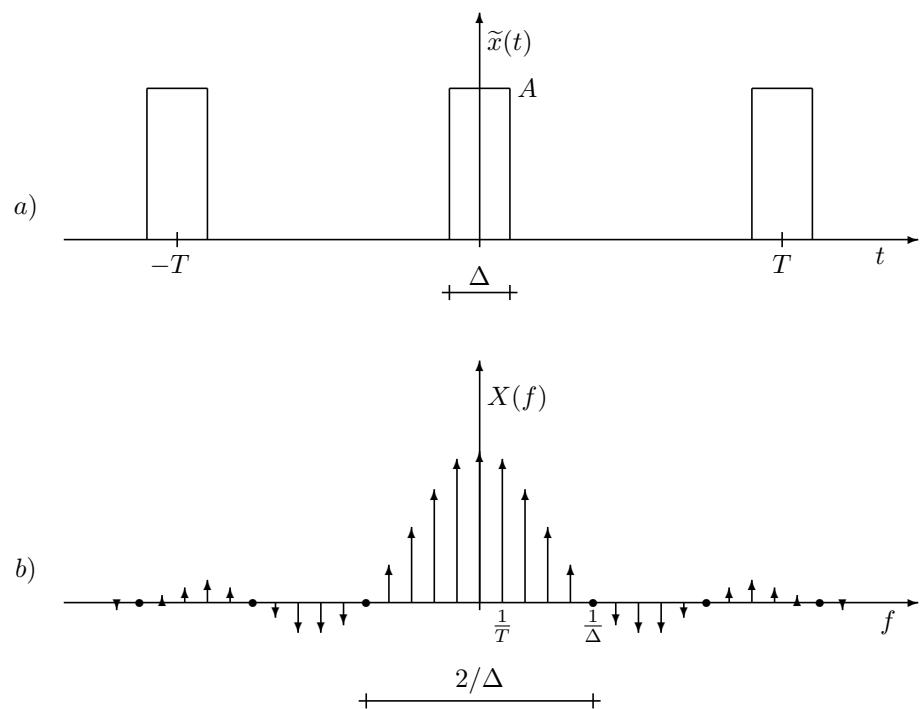


Figura 4.26: Treno di impulsi rettangolari e relativo spettro.

Dalla definizione data segue immediatamente che l' \mathcal{ESD} dell'ingresso $x(\cdot)$ e dell'uscita $y(\cdot)$ di un sistema LTI, di risposta armonica $H(\cdot)$, sono legate dalla relazione:

$$\mathcal{S}_y(\cdot) = |H(\cdot)|^2 \mathcal{S}_x(\cdot) \quad (4.51)$$

Nello scrivere tale equazione si è implicitamente supposto che i segnali in ingresso ed in uscita al sistema LTI siano entrambi di energia, ipotesi che, nel seguito, supporremo sempre soddisfatta; ad esempio tali ipotesi è certamente soddisfatta se il sistema è stabile, o, più in generale, se la sua risposta in frequenza è limitata.

La funzione $|H(\cdot)|^2$, che relaziona l' \mathcal{ESD} dell'ingresso e dell'uscita, prende il nome di funzione di trasferimento dell'energia.

Come accennato in precedenza la banda di un segnale viene valutata in base a considerazioni energetiche: precisamente, per i segnali di energia essa può essere definita come l'intervallo di frequenze in cui è allocata una determinata frazione dell'energia totale, ad esempio il 90%, il 95%, il 99%, ecc.; inoltre, se i segnali sono reali, comunemente si prendono in esame solo le frequenze positive (banda monolatera).

Una possibile ambiguità nella definizione di tale quantità deriva dal fatto che si possono considerare le sole frequenze positive ($f \geq 0$ per il caso continuo, $0 \leq \nu \leq 1/2$ per il caso discreto) oppure le frequenze positive e negative. Nel primo caso la banda è detta *monolatera* e, nel secondo, *bilatera*. Nel seguito, salvo avviso contrario, ci riferiremo alla banda monolatera, seguendo la convenzione più comune.

Per i segnali, la banda viene in genere determinata quantificando quali componenti possano essere trascurate rispetto al valore massimo dello spettro d'ampiezza: ad esempio si possono ritenere trascurabili le componenti che siano di almeno 20dB inferiori al massimo dello spettro d'ampiezza (componente più significativa). Naturalmente l'adeguatezza della definizione va verificata sperimentalmente in relazione alla particolare applicazione.

Per quanto riguarda i sistemi, la banda viene comunemente definita con riferimento alla risposta d'ampiezza, precisamente è l'intervallo di frequenza nel quale tale risposta si discosta dal valore assunto ad una frequenza di riferimento al più per una prefissata aliquota. Se tale aliquota è $1/\sqrt{2} = 0,707$, corrispondente a -3 dB , la banda si dice a 3 dB , ed è questo il caso più comune. In alcune applicazioni, ad esempio per valutare i sistemi che devono avere particolari caratteristiche di fedeltà di riproduzione ingresso-uscita, si sceglie un'aliquota maggiore, ad esempio 0,981, corrispondente a 1 dB . Nei sistemi passa-basso la frequenza di riferimento per la determinazione della banda è la frequenza zero. Nei sistemi passa-banda la scelta della frequenza di riferimento non è univoca e dipende dalle applicazioni.

Ad esempio, se il sistema è destinato a filtrare un segnale modulato da una portante sinusoidale a una data frequenza, allora questa si sceglie anche come frequenza di riferimento per il sistema, altrimenti si può scegliere il valore modale della risposta in ampiezza, o il valore baricentrico; la fig. 4.27 illustra la definizione di banda per sistemi passa-basso e passa-banda.

Esempio 1: *Impulso rettangolare*

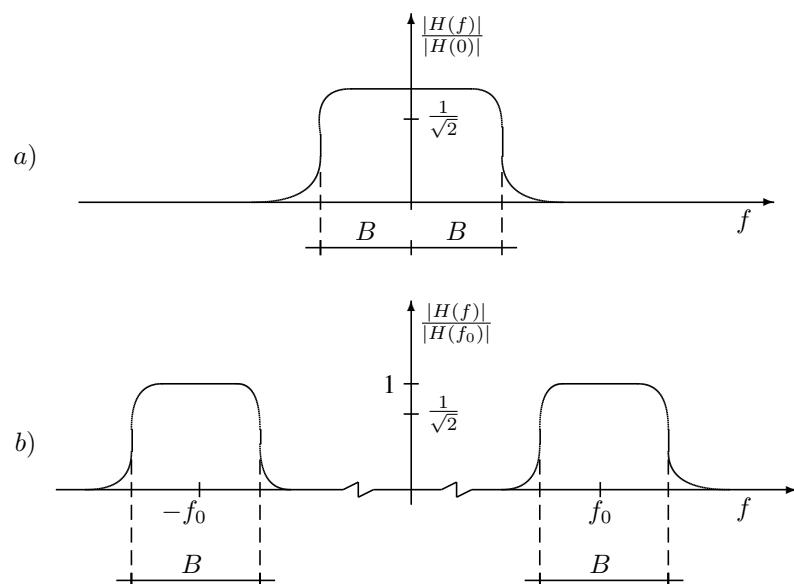


Figura 4.27: Definizione di banda a 3 dB per un sistema passa-basso (*a*) e passa-banda (*b*).

Sia $x(t)$ un impulso rettangolare di ampiezza A e durata T ; ricordando che

$$A\text{II}\left(\frac{t}{T}\right) \longleftrightarrow AT\text{sinc}(fT)$$

si ottiene immediatamente la seguente espressione per l' \mathcal{ESD} dell'impulso rettangolare

$$\mathcal{S}_x(f) = \mathcal{E}_x T \text{sinc}^2(fT)$$

ove $\mathcal{E}_x = A^2 T$ è l'energia di $x(t)$.

È interessante valutare la frazione di energia $\mathcal{E}(k)/\mathcal{E}$ dell'impulso rettangolare compresa nell'intervallo di frequenze

$$-\frac{k}{T} \leq f \leq \frac{k}{T} \quad k = 1, 2, \dots$$

Integrando l' \mathcal{ESD} si ha:

$$\frac{\mathcal{E}(k)}{\mathcal{E}} = \int_{-\frac{k}{T}}^{\frac{k}{T}} T \text{sinc}^2(fT) df = 2 \int_0^k \text{sinc}^2(\lambda) d\lambda$$

Valutando numericamente l'integrale si ha che per $k=1$, cioè nel lobo principale, la frazione di energia è pari al 90,3%; includendo anche una coppia di lobi laterali ($k=2$) la frazione di energia sale al 95 %, mentre per racchiudere il 99% dell'energia occorre portare in conto nove coppie di lobi laterali, cioè $k = 10$. Corrispondentemente la banda (monolatera) al 90% è T^{-1} , quella al 95% è $2T^{-1}$ e quella al 99% è $10T^{-1}$.

Esempio 2: Impulso esponenziale

Ricordando che:

$$x(t) = A e^{-2\pi f_0 t} u(t) \longleftrightarrow X(f) = \frac{A}{2\pi f_0 + j2\pi f}$$

si ottiene immediatamente la seguente espressione per l' \mathcal{ESD} di un impulso esponenziale:

$$\mathcal{S}(f) = \frac{A^2}{(2\pi f_0)^2 + (2\pi f)^2}$$

L'energia $\mathcal{E}(B)$ nella banda $-B \leq f \leq B$ è quindi:

$$\mathcal{E}(B) = \int_{-B}^B \frac{A^2}{(2\pi f_0)^2 + (2\pi f)^2} df = \frac{A^2}{2\pi^2 f_0^2} \arctan\left(\frac{B}{f_0}\right)$$

in particolare ponendo $B = \infty$ si ottiene l'energia totale $\mathcal{E}_x = \frac{A^2}{4\pi f_0}$.

Noto lo spettro di energia si può calcolare la banda del segnale. Precisamente la banda monolatera B_ϵ cui compete la frazione $1 - \epsilon$ dell'energia totale si ottiene risolvendo l'equazione

$$1 - \epsilon = \frac{2}{\pi} \arctan \left(\frac{B_\epsilon}{f_0} \right)$$

da cui si ricava:

$$B_\epsilon = f_0 \tan \left(\frac{\pi}{2} (1 - \epsilon) \right)$$

Così per esempio la banda al 95% dell'energia vale:

$$B_{0.05} = f_0 \tan \left(0.95 \frac{\pi}{2} \right) \text{ Hz} \simeq 12.7 f_0$$

4.6.1 Densità spettrale di potenza

Consideriamo ora i segnali di potenza per i quali l'energia è infinita, mentre è finita la potenza $\mathcal{P}_x = < |x(\cdot)|^2 >$. Nel seguito faremo riferimento prevalentemente ai segnali a tempo continuo considerato che le argomentazioni possono essere estese facilmente alle sequenze.

Per introdurre la funzione di densità spettrale di potenza (\mathcal{PSD}), e cioè la funzione che dà conto della distribuzione in frequenza della potenza, conviene riguardare i segnali di potenza come limite di segnali di energia. Precisamente sia $x_T(t)$ il segnale troncato all'intervallo $(-T, T)$ e sia $X_T(f)$ la sua trasformata, cioè:

$$x_T(t) = x(t) \Pi \left(\frac{t}{2T} \right) \longleftrightarrow X_T(f)$$

Procedendo come per i segnali di energia si ottiene

$$\begin{aligned} \mathcal{P}_x &= \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T |x(t)|^2 dt = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-\infty}^{+\infty} |x_T(t)|^2 dt = \\ &= \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-\infty}^{+\infty} |X_T(f)|^2 df = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-\infty}^{+\infty} |X_T(f)|^2 df \end{aligned}$$

Da cui, passando al limite sotto il segno di integrale si ha:

$$\mathcal{P}_x = \int_{-\infty}^{+\infty} \lim_{T \rightarrow \infty} \frac{1}{2T} |X_T(f)|^2 df$$

pertanto la funzione integranda è interpretabile come densità spettrale di potenza (abbreviazione \mathcal{PSD}) o, più semplicemente, spettro di potenza, di $x(t)$, che è dunque definita da:

$$\mathcal{S}_x(f) = \lim_{T \rightarrow \infty} \frac{1}{2T} |X_T(f)|^2 \quad (4.52)$$

Analogamente, per un segnale a tempo discreto, si pone:

$$\mathcal{S}_x(\nu) = \lim_{N \rightarrow \infty} \frac{1}{2N+1} |X_N(\nu)|^2 \quad (4.53)$$

dove $X_N(\nu)$ è lo spettro della sequenza troncata. Osserviamo esplicitamente che i limiti a secondo membro della definizione di PSD sono da intendersi in senso generalizzato.

La notazione utilizzata, come nel caso di segnali di energia, è $\mathcal{S}_x(\cdot)$ o $(\mathcal{S}(\cdot))$ se è chiaro dal contesto a quale segnale ci si riferisca essendo implicito nella natura di $x(\cdot)$ di quale densità spettrale, energia o potenza, si tratti.

Si osservi che la funzione di trasferimento dell'energia $|H(\cdot)|^2$ di un sistema LTI relaziona anche gli spettri di potenza, pertanto si ha:

$$\begin{aligned} \mathcal{S}_y(f) &= \lim_{T \rightarrow \infty} \frac{1}{2T} |Y_T(f)|^2 = \lim_{T \rightarrow \infty} \frac{1}{2T} |H(f)X_T(f)|^2 = \\ &= |H(f)|^2 \lim_{T \rightarrow \infty} \frac{1}{2T} |X_T(f)|^2 \end{aligned}$$

Analogamente a quanto visto per i segnali di energia, la PSD consente di definire la banda per i segnali di potenza in termini energetici come la gamma di frequenze in cui è allocata una prefissata frazione della potenza totale.

4.6.2 Spettri mutui di energia e di potenza

Innanzi tutto ricordiamo che se $x(\cdot)$ e $y(\cdot)$ sono due segnali di energia anche la loro somma $x(\cdot) + y(\cdot)$ lo è ed inoltre l'energia della somma vale

$$\mathcal{E}_{x+y} = \mathcal{E}_x + \mathcal{E}_y + \mathcal{E}_{xy} + \mathcal{E}_{yx} = \mathcal{E}_x + \mathcal{E}_y + 2\Re\{\mathcal{E}_{xy}\}$$

Anche l'energia mutua può essere analizzata in frequenza, invero dalla relazione di Parseval si ha:

$$\mathcal{E}_{xy} = \int_{-\infty}^{\infty} X(f)Y^*(f)df \quad \mathcal{E}_{xy} = \int_{-\frac{1}{2}}^{\frac{1}{2}} X(\nu)Y^*(\nu)d\nu$$

che porta immediatamente ad interpretare la funzione integranda come *spettro di energia mutua* o *densità spettrale di energia mutua* (\mathcal{ESD} mutua) di $x(\cdot)$ e $y(\cdot)$. Notiamo esplicitamente che in questa definizione l'ordine dei due segnali è importante, dal momento che il secondo dei due spettri deve essere coniugato; considerazioni analoghe sussistono per i segnali a tempo discreto. Dunque l' \mathcal{ESD} mutua è la funzione

$$\mathcal{S}_{xy}(\cdot) = X(\cdot)Y^*(\cdot)$$

e ovviamente si ha

$$\mathcal{S}_{xy}(\cdot) = \mathcal{S}_{yx}^*(\cdot)$$

Una relazione analoga a quella che esprime l'energia di una somma sussiste anche per l' \mathcal{ESD} . Invero sviluppando $|X(\cdot) + Y(\cdot)|^2$ è immediato mostrare che l' \mathcal{ESD} del segnale somma $x(\cdot) + y(\cdot)$ è data da:

$$\mathcal{S}_{x+y}(\cdot) = \mathcal{S}_x(\cdot) + \mathcal{S}_y(\cdot) + \mathcal{S}_{xy}(\cdot) + \mathcal{S}_{yx}(\cdot) = \mathcal{S}_x(\cdot) + \mathcal{S}_y(\cdot) + 2\Re\{\mathcal{S}_{xy}(\cdot)\} \quad (4.54)$$

Analogamente per i segnali di potenza ha interesse considerare la distribuzione in frequenza della potenza mutua $\mathcal{P}_{xy} = \langle x(\cdot), y(\cdot) \rangle$ di $x(\cdot)$ e $y(\cdot)$ nell'ordine. Per ottenere l'analisi spettrale della potenza mutua occorre troncare i segnali $x(\cdot)$ e $y(\cdot)$ mediante una finestra e poi far divergere la durata della finestra, così come si è fatto nel paragrafo precedente per la \mathcal{PSD} . In altri termini si definisce lo *spettro di potenza mutua* o *densità spettrale di potenza mutua* di $x(\cdot)$ e $y(\cdot)$ nell'ordine, e la si denota col simbolo $\mathcal{S}_{xy}(\cdot)$, il limite

$$\mathcal{S}_{xy}(f) = \lim_{T \rightarrow \infty} \frac{1}{2T} X_T(f) Y_T^*(f) \quad (4.55)$$

dove $X_T(f)$ e $Y_T(f)$ denotano al solito gli spettri dei segnali troncati. Analogamente nel caso di segnali a tempo discreto, con ovvio significato dei simboli, si pone

$$\mathcal{S}_{xy}(\nu) = \lim_{N \rightarrow \infty} \frac{1}{2N+1} X_N(\nu) Y_N^*(\nu)$$

Anche per i segnali di potenza, come è facile verificare, la \mathcal{PSD} del segnale somma è data dalla (4.54).

4.6.3 Processi bianchi

Il termine *processo bianco* viene utilizzato per descrivere un segnale in cui tutte le componenti spettrali hanno la stessa ampiezza, e dunque, la \mathcal{PSD} risulta essere una funzione costante della frequenza. La notevole importanza che i processi bianchi rivestono nelle applicazioni è in gran parte dovuta al fatto che la densità spettrale di potenza del rumore termico è sostanzialmente costante in tutta la banda di frequenze in cui i sistemi si trovano ad operare. D'altra parte, dal punto di vista rigorosamente formale, essendo

$$\mathcal{P}_x = \int_{-\infty}^{\infty} \mathcal{S}_x(f) df = \int_{-\infty}^{\infty} C df = \infty \quad (4.56)$$

il processo non ha rilevanza fisica. Il calcolo della densità spettrale di potenza del rumore termico può essere affrontato utilizzando considerazioni quanto-meccaniche da cui risulta che

$$\mathcal{S}_x(f) = \frac{\hbar f}{2(e^{\hbar f/KT} - 1)}$$

dove $\hbar = 6.610^{-34}$ Joules \times s. è la costante di Planck, $K = 1.3810^{-23}$ Joules/Kelvin è la costante di Boltzmann e T è la temperatura assoluta in gradi Kelvin. Lo spettro tende a zero

per $f \rightarrow \infty$ ma la convergenza è molto lenta. Infatti, per $T \simeq 300K$, ed $f \leq 10^{12}$ risulta $e^{\hbar f/KT} - 1 \simeq \hbar f/KT$ per cui, sostituendo nella (4.56) risulta

$$\mathcal{S}_x(f) \sim \frac{KT}{2} = \text{cost}$$

Quindi, per frequenze minori del terahertz la PSD del rumore termico può ritenersi con buona approssimazione costante e pari a $KT/2 = N_0/2$.

4.6.4 Teorema di Wiener-Kintchine

Consideriamo innanzi tutto due segnali di energia, ricordando la proprietà della trasformata di Fourier, per cui ad una coniugazione in un dominio corrisponde una coniugazione più un ribaltamento nell'altro dominio, antitrasformando lo spettro di energia mutua $\mathcal{S}_{xy}(\cdot) = X(\cdot)Y^*(\cdot)$, si ottiene:

$$x(\cdot) * y^*(-(\cdot)) \longleftrightarrow \mathcal{S}_{xy}(\cdot)$$

In altri termini la funzione di mutua correlazione e lo spettro di energia mutua sono legati fra loro dalla trasformata di Fourier, si ha cioè:

$$r_{xy}(\cdot) \longleftrightarrow \mathcal{S}_{xy}(\cdot) \quad (4.57)$$

Da cui, per $x(\cdot) \equiv y(\cdot)$, si ha che la funzione di autocorrelazione e la densità spettrale di energia sono legati fra loro dalla trasformata di Fourier, cioè:

$$r_x(\cdot) \longleftrightarrow \mathcal{S}_x(\cdot)$$

Il legame tra funzioni di correlazione e spettri di energia è una immediata conseguenza delle proprietà della trasformata di Fourier; la rilevanza del risultato è la sua validità anche per segnali di potenza, deterministici o aleatori che siano: in altri termini, sussiste il seguente:

Teorema di Wiener-Kintchine: Dati due segnali $x(\cdot)$ e $y(\cdot)$, il loro spettro di potenza mutua $\mathcal{S}_{xy}(\cdot)$ è la trasformata di Fourier della loro funzione di mutua correlazione.

Prova: Supposti i segnali a tempo continuo, sviluppando la media statistica a secondo membro della definizione di spettro mutuo (4.55) si ha:

$$X_T(f)Y_T^*(f) = \int_{-\infty}^{+\infty} x(t_1)\Pi\left(\frac{t_1}{2T}\right)e^{-j2\pi f t_1} \int_{-\infty}^{+\infty} y^*(t_2)\Pi\left(\frac{t_2}{2T}\right)e^{j2\pi f t_2} dt_1 dt_2$$

da cui si ricava:

$$\begin{aligned} X_T(f)Y_T^*(f) &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \Pi\left(\frac{t_1}{2T}\right)\Pi\left(\frac{t_2}{2T}\right)x(t_1)y^*(t_2)e^{-j2\pi f(t_1-t_2)} dt_1 dt_2 \\ &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \Pi\left(\frac{t_1}{2T}\right)\Pi\left(\frac{t_2}{2T}\right)r_{xy}(t_1, t_2)e^{-j2\pi f(t_1-t_2)} dt_1 dt_2 \end{aligned}$$

ed effettuando nell'integrale il cambio di variabile

$$\begin{cases} t = t_1 \\ \tau = t_1 - t_2 \end{cases}$$

si ottiene

$$X_T(f)Y_T^*(f) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \Pi\left(\frac{t}{2T}\right) \Pi\left(\frac{t-\tau}{2T}\right) r_{xy}(t, \tau) e^{-j2\pi f\tau} dt d\tau$$

da cui, dividendo per $2T$ e passando al limite per $T \rightarrow \infty$, si ha

$$\mathcal{S}_{xy}(f) = \int_{-\infty}^{+\infty} \left[\lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-\infty}^{+\infty} \Pi\left(\frac{t}{2T}\right) \Pi\left(\frac{t-\tau}{2T}\right) r_{xy}(t, \tau) dt \right] e^{-j2\pi f\tau} d\tau$$

Il limite tra parentesi quadre a secondo membro della relazione precedente formalmente differisce dalla media temporale in quanto il prodotto delle due finestre rettangolari di durata $2T$ è ancora una finestra rettangolare ma di durata $2T - |\tau|$; tuttavia, per $T \rightarrow \infty$ la riduzione della durata non ha effetto ($2T - |\tau| \simeq 2T$) e, conseguentemente, l'espressione tra parentesi quadre è la mutua correlazione $r_{xy}(\tau)$ tra i due segnali, onde l'asserto.

Si osservi infine che se due segnali sono incoerenti allora, a norma del teorema di Wiener-Kinchine, la loro \mathcal{PSD} mutua è nulla; pertanto non solo l'autocorrelazione di una somma di segnali incoerenti è la somma delle singole autocorrelazioni, ma anche la \mathcal{PSD} della somma di segnali incoerenti è la somma delle \mathcal{PSD} dei singoli addendi.

4.7 LEGAMI INGRESSO-USCITA PER LE \mathcal{PSD}

Sulla scorta del Teorema di Wiener-Kinchine è immediato stabilire i legami esistenti tra le \mathcal{PSD} , o le \mathcal{ESD} a seconda del tipo di segnali, in ingresso ed in uscita a sistemi LTI semplicemente trasformando gli analoghi legami tra le correlazioni stabiliti nel Cap.2. Nel seguito faremo riferimento alle \mathcal{PSD} che è il caso di maggiore interesse, anche se le relazioni stabilite, grazie alla simbologia unica per \mathcal{PSD} e \mathcal{ESD} , valgono senza alcuna modifica anche per gli spettri di energia.

Ricordando, ad esempio, che la mutua correlazione tra le uscite di due sistemi LTI (vedi fig. 4.28a) è data dalla relazione

$$r_{y_1 y_2}(\cdot) = r_{h_1 h_2}(\cdot) * r_{x_1 x_2}(\cdot)$$

trasformando secondo Fourier si ha

$$\mathcal{S}_{y_1 y_2}(\cdot) = H_1(\cdot)H_2^*(\cdot)\mathcal{S}_{x_1 x_2}(\cdot) \quad (4.58)$$

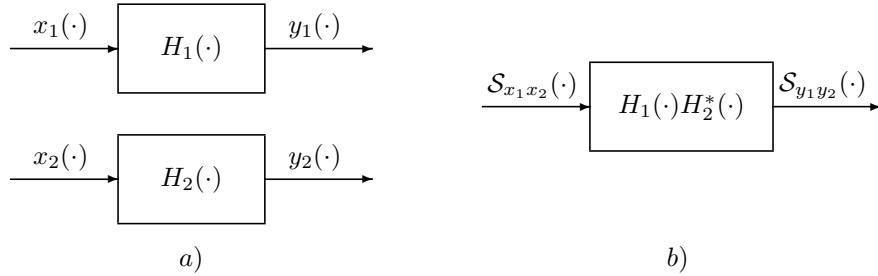


Figura 4.28: Legame I/O per le PSD mutue: a) sistemi effettivi, b) sistema equivalente.

In altri termini al fine del calcolo dello spettro di potenza incrociato tra le due uscite è sufficiente sostituire i due sistemi con un unico sistema di risposta armonica $H_1(\cdot)H_2^*(\cdot)$ sollecitato dalla PSD mutua tra i due ingressi, come schematizzato in fig. 4.28b. In modo analogo è possibile stabilire gli altri legami di interesse: così lo spettro di potenza dell'uscita è legato a quello dell'ingresso dalla relazione

$$\mathcal{S}_y(\cdot) = |H(\cdot)|^2 \mathcal{S}_x(\cdot) \quad (4.59)$$

cioè la PSD dell'uscita è pari a quella dell'ingresso per la funzione di trasferimento dell'energia, mentre le PSD mutue uscita-ingresso e ingresso-uscita sono date da

$$\mathcal{S}_{yx}(\cdot) = H(\cdot)\mathcal{S}_x(\cdot) \quad \mathcal{S}_{xy}(\cdot) = H(\cdot)^*\mathcal{S}_x(\cdot)$$

I vari legami considerati sono riassunti nella fig. 4.29. Una ulteriore conseguenza della (4.58)

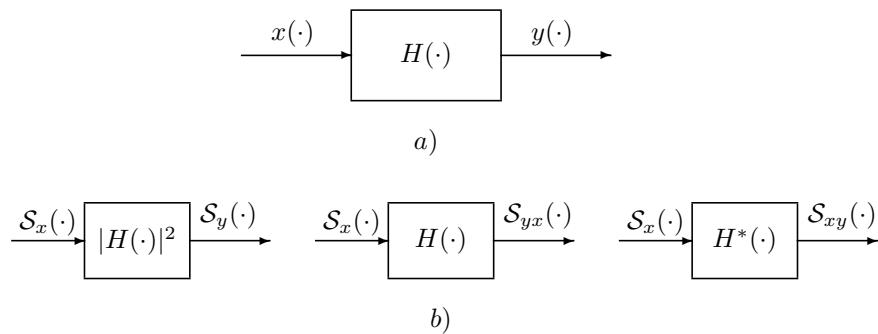


Figura 4.29: Legami I/O per PSD: a) sistema effettivo, b) sistemi equivalenti.

è l'incoerenza di due segnali qualsiasi che non si sovrappongono nel dominio della frequenza: è sufficiente infatti osservare che tali segnali restano inalterati per effetto di un filtraggio LTI

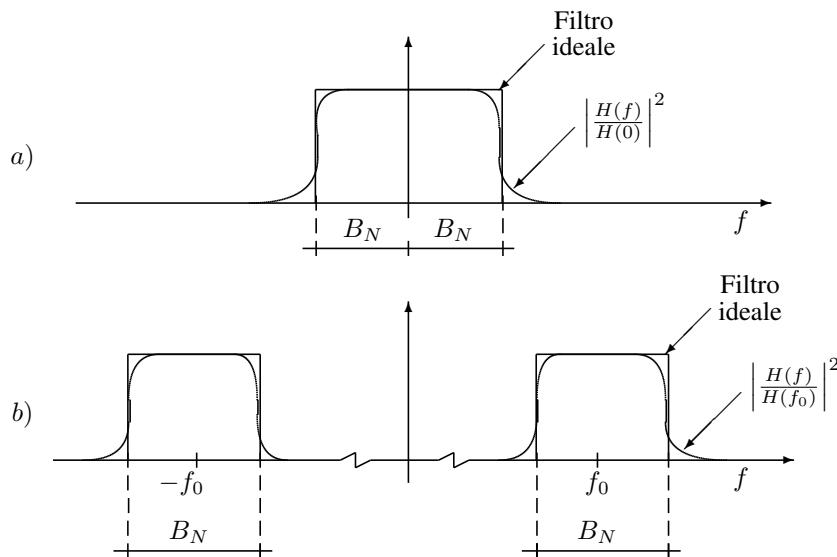


Figura 4.30: Definizione di banda di rumore: sistemi passa-basso (a) e passa-banda (b).

da parte di due sistemi le cui risposte armoniche valgano uno ove la \mathcal{PSD} del singolo segnale è non nulla e zero altrove.

La potenza di rumore in uscita ad un filtro LTI quando l'ingresso è rumore bianco è data da:

$$\mathcal{P}_x = \frac{\mathcal{N}_0}{2} \int_{-\infty}^{+\infty} |H(f)|^2 df$$

ove $\mathcal{N}_0/2$ denota la \mathcal{PSD} del rumore supposto, per fissare le idee, tempo continuo. Introdotta la banda di rumore (monolatera) del filtro

$$B_N = \int_0^{+\infty} \left| \frac{H(f)}{H(f_0)} \right|^2 df$$

l'espressione della potenza si semplifica in

$$\mathcal{P}_x = \mathcal{N}_0 |H(f_0)|^2 B_N$$

Pertanto la banda di rumore è la banda del filtro ideale avente lo stesso guadagno $|H(f_0)|^2$ di centro banda e la cui funzione di trasferimento dell'energia sottende la stessa area. Tale interpretazione è illustrata in fig. 4.30 sia nel caso di sistemi passa-basso che passa-banda. Dalla relazione di Parseval segue che la banda di rumore è calcolabile anche a partire dalla

risposta impulsiva con la relazione:

$$2B_N = \frac{\int_{-\infty}^{+\infty} |h(t)|^2 dt}{\left| \int_{-\infty}^{+\infty} h(t) e^{-j2\pi f_0 t} dt \right|^2}$$

Si osservi che la definizione data vale, con ovvie modifiche, anche per sistemi discreti; precisamente, in tal caso, si ha:

$$B_N = \int_0^{+\frac{1}{2}} \left| \frac{H(\nu)}{H(\nu_0)} \right|^2 d\nu \quad 2B_N = \frac{\sum_{m=-\infty}^{+\infty} |h(m)|^2}{\left| \sum_{m=-\infty}^{+\infty} h(m) e^{-j2\pi \nu_0 m} \right|^2}$$

Si osservi infine che, nel definire la banda di rumore si è implicitamente fatto riferimento ai sistemi reali, anche se la definizione può essere estesa con ovvie modifiche anche al caso di sistemi complessi.

Capitolo 5

Il campionamento dei segnali

I segnali a tempo discreto spesso sono una versione campionata di segnali a tempo continuo e analogamente i *segnali numerici* derivano da una quantizzazione di sequenze. Ciò è dovuto al fatto che l'elaborazione di segnali analogici si può eseguire vantaggiosamente operando preventivamente una *conversione analogica/numerica* o, sinteticamente *conversione A/D*, sulle forme d'onda, elaborando poi numericamente le sequenze così ottenute ed effettuando infine, se necessario, una *conversione numerico/analogica* o sinteticamente *conversione D/A*.

Sorprendentemente, in ipotesi molto bene approssimate in pratica, una forma d'onda è adeguatamente rappresentata dai suoi campioni: nel prosieguo del capitolo stabiliremo innanzi tutto quali debbano essere tali ipotesi e mostreremo come la forma d'onda possa essere ricostruita a partire dai propri campioni (Teorema del campionamento). Verranno poi analizzati gli effetti connessi a considerazioni di ordine pratico.

Si consideri una forma d'onda $x_a(t)$: ci proponiamo di stabilire se ed in quali ipotesi la sequenza dei suoi campioni

$$x(n) = x_a(nT), \quad -\infty < n < \infty \quad (5.1)$$

rappresenti il segnale tempo continuo $x_a(t)$, nel senso che sia possibile ricostruire esattamente $x_a(t)$ a partire da $x(n)$ (**Teorema del campionamento uniforme**): T è detto *periodo (o passo) di campionamento*, e il suo reciproco $f_c = 1/T$ è la *frequenza (o cadenza) di campionamento*.

5.1 CAMPIONAMENTO IDEALE

Ricordiamo che la versione campionata idealmente di una forma d'onda $x_a(t)$ è il segnale

$$x_\delta(t) = \sum_{n=-\infty}^{\infty} x_a(nT) \delta(t - nT) \quad (5.2)$$

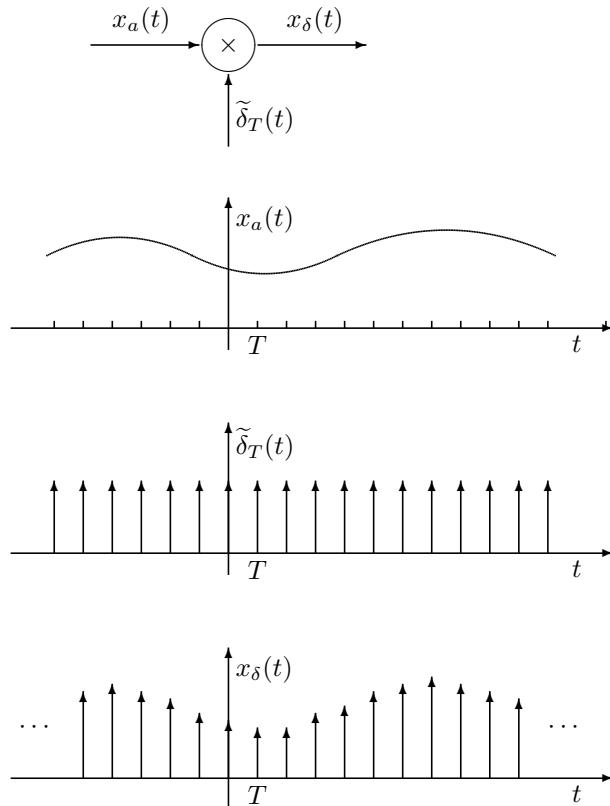


Figura 5.1: Campionamento ideale uniforme.

e che il campionamento ideale è schematizzabile come il prodotto di $x_a(t)$ per il treno periodico di impulsi di Dirac

$$\tilde{\delta}_T(t) = \sum_{n=-\infty}^{\infty} \delta(t - nT) \quad (5.3)$$

come illustrato in fig. 5.1. Ricordando inoltre che ad un campionamento nel dominio del tempo corrisponde una replicazione in frequenza, si ha che lo spettro di $x_\delta(t)$ vale:

$$X_\delta(f) = \frac{1}{T} \sum_{k=-\infty}^{\infty} X_a \left(f - \frac{k}{T} \right) = f_c \sum_{k=-\infty}^{\infty} X_a(f - kf_c) \quad (5.4)$$

Dunque lo spettro del segnale campionato idealmente è costituito da replicate dello spettro di $x_a(t)$ traslate in frequenza di $kf_c = k/T$ e scalate in ampiezza secondo il fattore $1/T = f_c$.

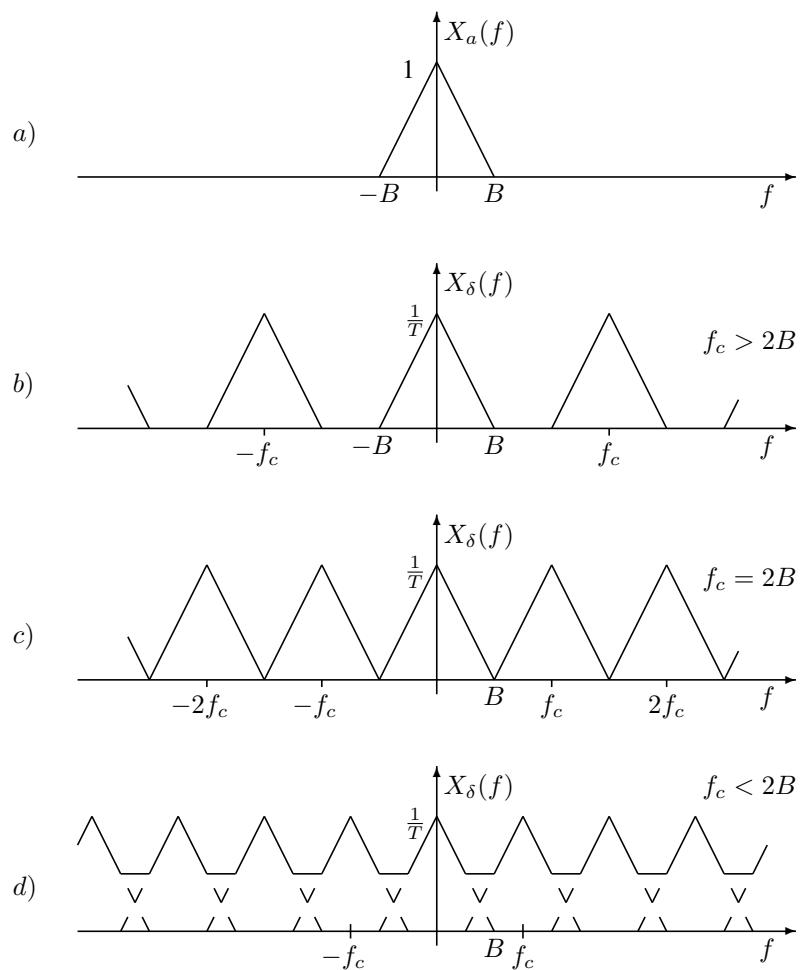


Figura 5.2: Analisi del campionamento nel dominio della frequenza.

La fig. 5.2 fornisce l'interpretazione grafica della precedente relazione: precisamente la fig. 5.2a mostra lo spettro di un segnale $x_a(t)$ con banda (monolatera) B strettamente limitata, la fig. 5.2b rappresenta lo spettro del segnale campionato nel caso che le repliche di $X_a(f)$ non si sovrappongono (sovracampionamento), essendo soddisfatta con un certo margine la *condizione di Nyquist*

$$f_c \geq 2B \quad \text{ovvero} \quad T \leq 1/2B \quad (5.5)$$

la fig. 5.2c si riferisce al caso di campionamento a frequenza di Nyquist $f_c = 2B$, mentre la fig. 5.2d è relativa al caso in cui tale condizione non sia soddisfatta (sottocampionamento).

Dall'analisi in frequenza del campionamento segue che, se il segnale è a banda limitata ed è soddisfatta la condizione di Nyquist, allora $x_a(t)$ può essere esattamente ricostruito dalla sua versione campionata $x_\delta(t)$ con un filtro passabasso ideale avente guadagno T (per compensare il fattore di scala delle ampiezze) e frequenza di taglio f_p soddisfacente la condizione:

$$B \leq f_p \leq f_c - B \quad (5.6)$$

avente cioè risposta armonica

$$H_r(f) = T \Pi \left(\frac{f}{2f_p} \right) \quad (5.7)$$

come illustrato in fig. 5.3. Se invece il segnale non è a banda limitata o se, pur essendolo, la disuguaglianza (5.5) non è soddisfatta, allora le repliche di $X_a(f)$ si sovrappongono, come mostrato in fig. 5.2d, e quindi la ricostruzione non è più possibile: si dice allora che il segnale campionato è affetto da *aliasing*. La minima frequenza di campionamento per cui un segnale con banda (monolatera) B può essere ricostruito senza dar luogo ad aliasing è pari a

$$f_c = f_N = 2B$$

e viene detta *frequenza o cadenza di Nyquist*. In conclusione sussiste il seguente:

Teorema del campionamento uniforme (Shannon): un segnale analogico $x_a(t)$ è rappresentato dai suoi campioni presi con passo costante T , ovvero con cadenza $f_c = 1/T$, se

1. il segnale $x_a(t)$ è a banda rigorosamente limitata, cioè il suo spettro $X_a(f)$ soddisfa la condizione

$$X_a(f) \equiv 0 \quad \forall |f| \geq B \quad (5.8)$$

2. la cadenza di campionamento è maggiore o uguale a quella di Nyquist, cioè $f_c \geq 2B$.

Inoltre il segnale può essere ricostruito dalla sua versione campionata $x_\delta(t)$ con un filtro passabasso ideale di guadagno T la cui frequenza di taglio f_p soddisfi la condizione (5.6).

Come precedentemente osservato la ricostruzione di $x_a(t)$ da $x_\delta(t)$ si ottiene con un filtro passabasso ideale di guadagno T e frequenza di taglio f_p soddisfacente la condizione (5.6); scegliendo in particolare $f_p = \frac{1}{2}f_c$, la risposta in frequenza del filtro di ricostruzione

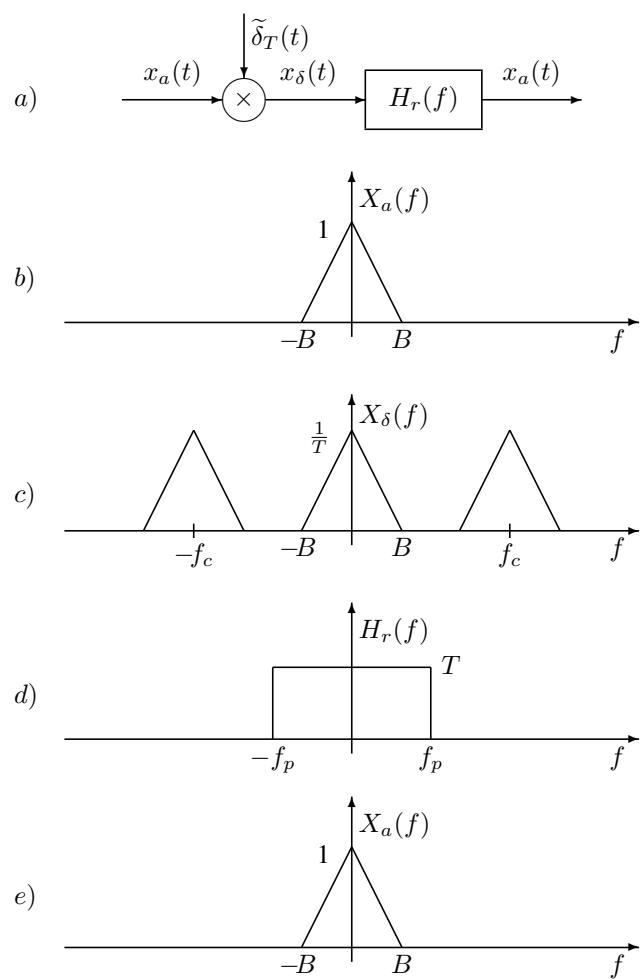


Figura 5.3: Ricostruzione di una forma d'onda dai suoi campioni con un filtro ideale.

vale:

$$H_r(f) = T\Pi(fT) = \frac{1}{f_c} \Pi\left(\frac{f}{f_c}\right) \quad (5.9)$$

cui corrispondente la risposta impulsiva:

$$h_r(t) = \frac{\sin \pi t/T}{\pi t/T} = \text{sinc}\left(\frac{t}{T}\right)$$

In tale ipotesi, valutando l'uscita col principio di sovrapposizione degli effetti, cioè come somma delle risposte ai singoli impulsi componenti il treno $x_\delta(t)$, si ottiene la seguente rappresentazione in serie (*serie di Shannon*) per un segnale con banda monolatera $B \leq \frac{1}{2T}$:

$$x_a(t) = \sum_{n=-\infty}^{\infty} x(n) \text{sinc}\left(\frac{t - nT}{T}\right) \quad (5.10)$$

Questa è la formula di *interpolazione ideale*: l'aggettivo *ideale* tiene conto del fatto che le funzioni interpolatrici $\text{sinc}(\cdot)$, che si dovrebbero sovrapporre per ricostruire il segnale desiderato, sono infinitamente estese nel tempo in ambedue le direzioni: in altre parole il filtraggio di ricostruzione ideale non è fisicamente realizzabile.

5.2 CAMPIONAMENTO REALE

Nella pratica applicazione del teorema del campionamento occorre portare in conto alcune limitazioni inerenti all'estensione spettrale dei segnali e alla fisica realizzabilità dei circuiti di campionamento e ricostruzione. Precisamente le modifiche da considerare rispetto al campionamento ideale sono le seguenti:

- il segnale campionato consiste di impulsi di ampiezza e durata finita;
- il filtro di ricostruzione non è ideale;
- il segnale da campionare non è a banda rigorosamente limitata.

Analizziamo separatamente gli effetti di tali modifiche, anche se ovviamente tali effetti in pratica si combinano tra loro.

5.2.0.1 Effetto della forma degli impulsi campionatori

Nei convertitori analogico/numerico spesso il corretto funzionamento dei circuiti a valle del campionamento, cioè i circuiti di quantizzazione e codifica, richiede che gli impulsi costituenti il segnale campionato abbiano durata non troppo piccola ed ampiezza costante. Per garantire tali condizioni al campionamento vero e proprio, sostanzialmente istantaneo, si fa seguire la memorizzazione del campione per un tempo τ , di norma, ma non necessariamente, pari

al passo di campionamento T , come illustrato in fig. 5.4 per due diversi valori del tempo di memorizzazione. In tal caso si parla di campionamento del tipo *campiona e trattienei* (*Sample & Hold*), o sinteticamente S&H, per sottolineare che il segnale campionato S&H è schematizzabile come:

$$x_{SH}(t) = \sum_{n=-\infty}^{+\infty} x(n)p(t - nT)$$

con

$$p(t) = \Pi\left(\frac{t - 0.5\tau}{\tau}\right)$$

cioè come un treno di impulsi rettangolari la cui ampiezza è pari, al generico campione. Pertanto da punto di vista funzionale il campionamento S&H può essere riguardato come:

$$x_{SH}(t) = x_\delta(t) * p(t)$$

cioè come una versione filtrata del segnale campionato idealmente con un filtro LTI di risposta impulsiva $p(t)$.

Per valutare la possibilità di ricostruire il segnale dalla sua versione campionata S&H è conveniente operare nel dominio della frequenza: trasformando la relazione precedente si ottiene

$$X_{SH}(f) = X_\delta(f)P(f) = P(f)\frac{1}{T} \sum_{n=-\infty}^{+\infty} X_a\left(f - \frac{n}{T}\right) \quad (5.11)$$

L'analisi in frequenza del campionamento S&H è illustrata in fig. 5.5: precisamente nella fig. 5.5a è riportato lo spettro del segnale da campionare, supposto a banda limitata B , nella fig. 5.5b è riportato lo spettro del segnale campionato idealmente, nella fig. 5.5c è riportato lo spettro dell'impulso formatore e nella fig. 5.5d è riportato lo spettro del segnale campionato (S&H).

La fig. 5.5 si riferisce al caso di campionamento a frequenza di poco superiore a quella di Nyquist e ad un tempo di memorizzazione pari al passo di campionamento T . In tali ipotesi, dall'espressione dello spettro del segnale S&H (5.11), segue che le varie repliche che costituiscono lo spettro del segnale campionato idealmente vengono distorte per effetto della memorizzazione; in particolare, per quanto concerne la replica fondamentale di $X_{SH}(f)$, si ha una attenuazione delle componenti a frequenza più elevata di $x_a(t)$ (*effetto apertura*). Tuttavia il segnale $x_a(t)$ è egualmente recuperabile a partire dalla sua versione campionata S&H mediante filtraggio passabasso: solo che in questo caso la risposta in frequenza del filtro deve essere proporzionale al reciproco dello spettro dell'impulso $p(t)$ per compensare la distorsione, cioè deve avversi

$$H_r(f) \propto \frac{1}{P(f)} \quad \forall |f| \leq B$$

In altri termini, la ricostruzione del segnale può realizzarsi ponendo in cascata un filtro ideale passabasso ed un egualizzatore.

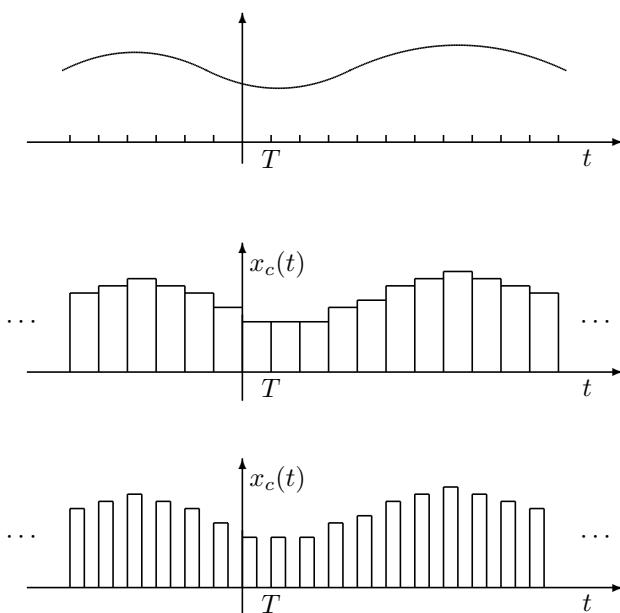


Figura 5.4: Campionamento S&H.

L'effetto apertura può essere ridotto o diminuendo il tempo di memorizzazione e/o aumentando la frequenza di campionamento. Infatti, per una data frequenza di campionamento, diminuendo a sufficienza il tempo di memorizzazione lo spettro dell'impulso $p(t)$ può essere considerato costante nella banda di $x_a(t)$, cioè

$$P(f) \simeq P(0) \quad \forall |f| \leq B \quad (5.12)$$

conseguentemente la replica fondamentale di $X_{SH}(f)$, cioè l'addendo per $n = 0$ nell'equazione (5.11), è una versione sostanzialmente indistorta dello spettro di $x_a(t)$: pertanto è possibile ricostruire il segnale a partire dalla sua versione campionata S&H con un filtro passabasso. Quando, come quasi sempre si verifica in pratica, il tempo di memorizzazione è pari al passo di campionamento T , la diminuzione della durata degli impulsi costituenti il segnale campionato SH, necessaria per realizzare la condizione (5.12), equivale ad incrementare la frequenza di campionamento f_c ; in tal caso le varie repliche costituenti lo spettro del segnale campionato sono tanto più separate quanto più elevata è l'entità del sovraccampionamento il che rende particolarmente semplice la realizzazione del filtro di ricostruzione.

In conclusione sia nel caso di campionamento a prodotto che nel caso di campionamento SH la forma degli impulsi non costituisce una effettiva limitazione, a patto di ricorrere, nel caso di campionamento S&H, ad un adeguato sovraccampionamento del segnale e/o ad un filtro di ricostruzione con risposta armonica opportuna.

5.2.0.2 Effetto del filtraggio non ideale

Il teorema del campionamento afferma che è possibile ricostruire il segnale a partire dalla sua versione campionata $x_\delta(t)$ mediante un filtro ideale: ovviamente in pratica tale tipo di ricostruzione non è realizzabile, ma solo approssimabile. I filtri passabasso che possono essere realizzati hanno una risposta armonica del tipo di fig. 5.6a: nella banda passante ($|f| \leq f_p$) ed in quella oscura ($|f| \geq f_s$) la risposta in frequenza del filtro approssima ottimamente quella del filtro ideale, mentre nella banda di transizione ($f_p \leq |f| \leq f_s$) lo scostamento dal comportamento ideale è notevole. Pertanto nella ricostruzione possono avversi due effetti

- una distorsione, tipicamente un'attenuazione, delle componenti a frequenza più elevata del segnale utile;
- la presenza nel segnale ricostruito di residui delle repliche centrate a frequenza multipla di quella di campionamento f_c .

Il primo effetto si ha se la banda del segnale è maggiore della frequenza di taglio del filtro (fig. 5.6b), cioè per $B > f_p$, mentre il secondo è prevalente se le repliche non capitano tutte in banda oscura (fig. 5.6c), cioè se $f_c - B < f_s$. Tali effetti possono essere ricondotti entro limiti accettabili aumentando la frequenza di campionamento, se è assegnato il filtro di ricostruzione, ovvero scegliendo un filtro di ricostruzione migliore (banda di transizione più stretta) per una data frequenza di campionamento. Un sovraccampionamento è comunque inevitabile per garantire una certa separazione delle varie repliche (*banda di*

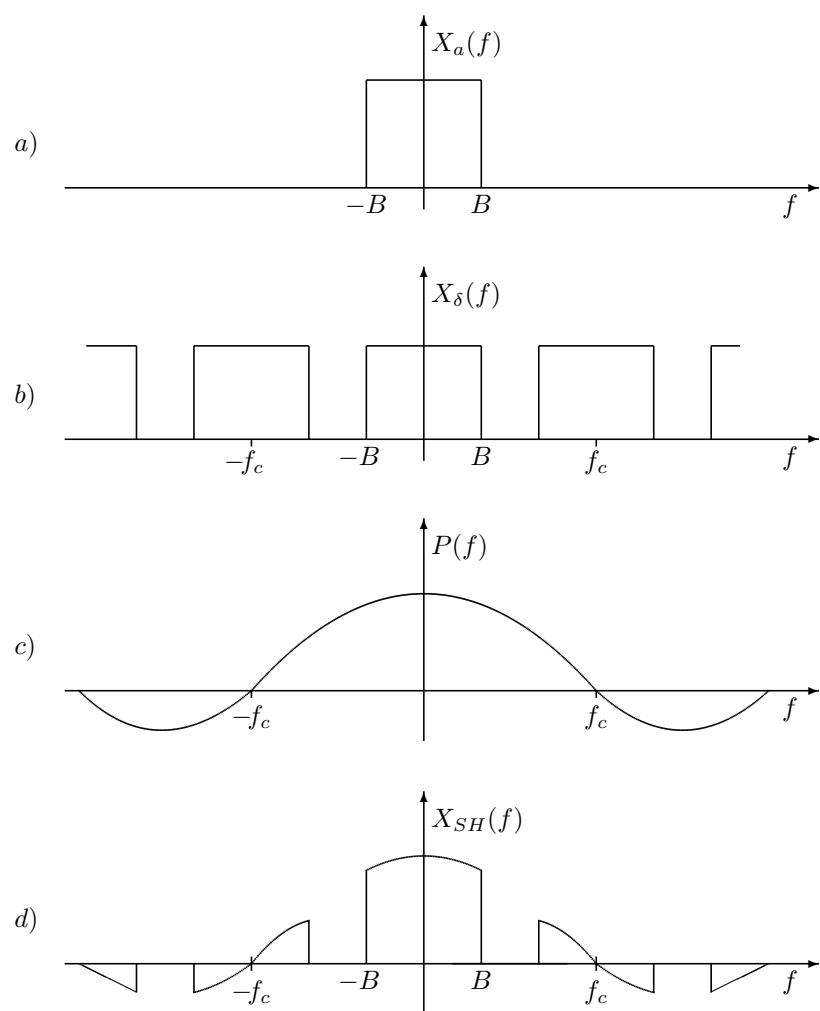


Figura 5.5: Analisi in frequenza del campionamento SH.

guardia) necessaria per allocare la banda di transizione del filtro (vedi fig. 5.6d): l'entità del sovraccampionamento dipende dalla qualità del filtro di ricostruzione nel senso che migliore è il filtro più bassa è la frequenza di campionamento utilizzabile.

In conclusione con un buon filtro e con un adeguato tasso di sovraccampionamento è possibile ricostruire col richiesto grado di fedeltà il segnale.

5.2.0.3 Effetto della banda non rigorosamente limitata

La maggior parte dei segnali hanno durata limitata, pertanto non possono essere a banda rigorosamente limitata, ma solo a banda praticamente limitata nel senso che le componenti frequenziali esterne alla banda nominale $-B < f < B$ sono piccole o comunque trascurabili ai fini dell'applicazione studiata. Ad esempio la voce ha un contenuto spettrale che si estende fino a frequenze di circa 20kHz, ma, tenuto conto che nella telefonia occorre garantire l'intellegibilità della conversazione e non la fedeltà della riproduzione del segnale vocale e tenuto conto delle caratteristiche dell'orecchio umano, per tale tipo di applicazione è sufficiente limitare il contenuto spettrale a 3,4kHz: pertanto la banda nominale del segnale telefonico è notevolmente inferiore alla banda effettiva.

Per effetto della banda non rigorosamente limitata le varie repliche di $X_a(f)$ presenti nel segnale campionato si sovrappongono come illustrato nella fig. 5.7: conseguentemente nel segnale ricostruito a partire da $x_\delta(t)$ le componenti spettrali originariamente esterne alla banda utile si presentano nel segnale ricostruito all'interno di tale banda sovrapponendosi a quelle utili (aliasing). Precisamente la componente a frequenza f_0 , con $\frac{1}{2}f_c \leq f_0 \leq f_c$ del segnale $x_a(t)$ nel segnale ricostruito è dislocata a frequenza $-f_a = f_0 - f_c$, con $0 \leq f_a \leq \frac{1}{2}f_c$, e analogamente la componente a frequenza $-f_0$ è dislocata a frequenza $f_a = f_c - f_0$; più in generale la componente a frequenza f_1 del segnale ricostruito è la risultante di tutte le componenti a frequenze $f_k = f_1 + kf_c$, k intero relativo, del segnale da campionare. Tali frequenze sono dette frequenze *alias* (\equiv altrove) della frequenza f_1 e la sovrapposizione delle repliche è comunemente denotata col termine inglese *aliasing*. Si noti che $f_0 = f_c - B$ è la frequenza più bassa che, per effetto dell'aliasing, rientra in banda utile

Allo scopo di ridurre la distorsione da *aliasing* è indispensabile attenuare notevolmente il contenuto spettrale esterno alla banda $-\frac{1}{2}f_c \leq f \leq \frac{1}{2}f_c$ prima di effettuare il campionamento, mediante un filtraggio passabasso detto appunto filtraggio anti-*aliasing*. Poiché un qualsiasi filtro fisicamente realizzabile inevitabilmente presenta una banda di transizione non nulla è anche necessario sovraccampionare il segnale in modo che non siano affette da *aliasing* le componenti del segnale nella banda di interesse $-B \leq f \leq B$, ma eventualmente solo quelle che capitano nella regione di transizione. In altri termini la frequenza f_p al limite di banda passante e quella f_s al limite di banda oscura del filtro anti-aliasing devono soddisfare le seguenti limitazioni:

$$f_p \geq B \quad f_s \leq f_c - B$$

In conclusione per effetto del carattere non ideale degli impulsi del treno campionatore, dell'impiego di filtri non-ideali e della banda non limitata le ipotesi del teorema del campionamento in pratica non si verificano; tuttavia, una frequenza di campionamento adeguatamente

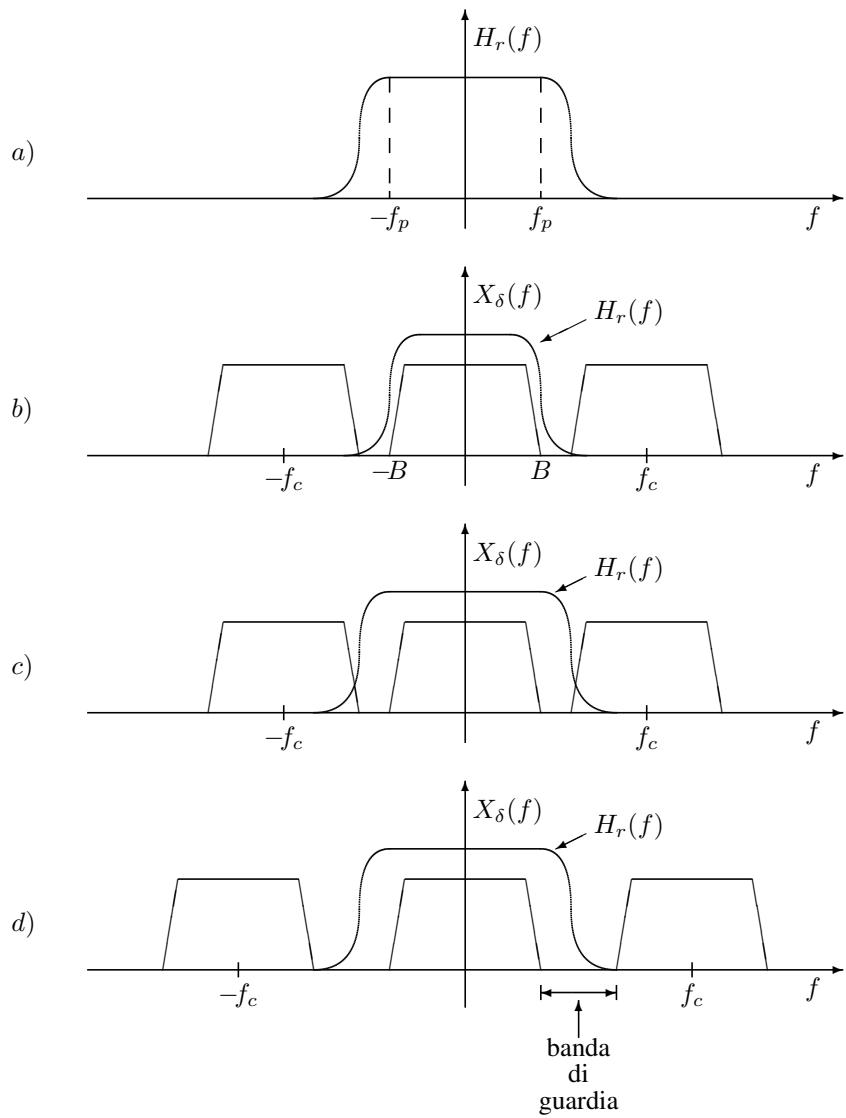


Figura 5.6: Ricostruzione con filtri non ideali.

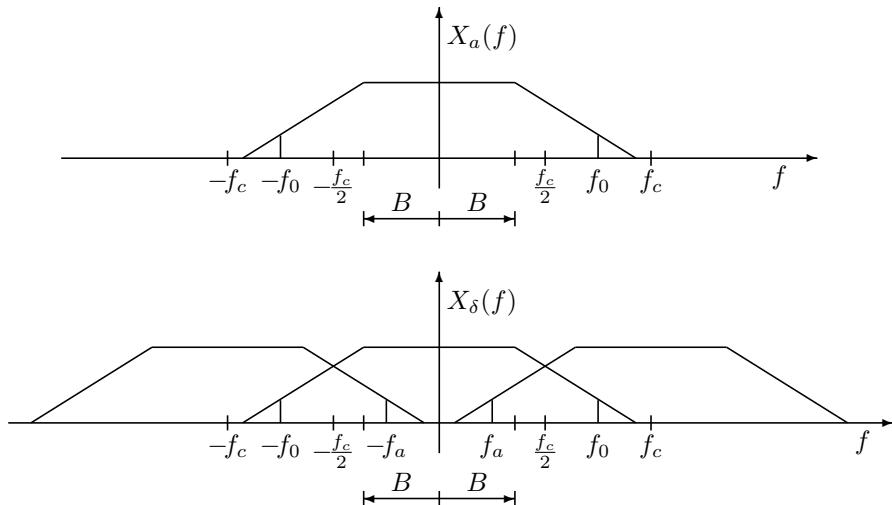


Figura 5.7: Effetto della banda non limitata.

maggiori di quella di Nyquist e l'uso di buoni filtri consentono di rendere piccolo quanto si vuole lo scarto tra il segnale originario e quello ricostruito: inoltre è possibile barattare la qualità dei filtri con l'entità del sovraccampionamento nel senso che un sovraccampionamento più elevato consente l'impiego di filtri meno costosi e viceversa.

5.3 CONVERSIONE ANALOGICO/NUMERICA

Come precedentemente accennato l'elaborazione numerica dei segnali richiede la conversione delle forme d'onda in segnali numerici o *conversione A/D*. Da un punto di vista funzionale la conversione A/D può essere suddivisa nei seguenti passi:

- *Conversione t/n*, cioè la conversione della forma d'onda $x_a(t)$ nella sequenza dei suoi campioni $x(n) = x_a(nT)$.
- *Quantizzazione e codifica*, cioè la rappresentazione dei singoli campioni con un numero finito di cifre significative, di norma binarie.

5.3.1 Conversione t/n

La conversione *t/n* consiste nell'estrare dalla forma d'onda $x_a(t)$ la sequenza dei suoi campioni $x(n) = x_a(nT)$: pertanto tale conversione è essenzialmente un campionamento, ma si preferisce denominarla conversione *t/n* per evitare confusione con il campionamento definito dalla (5.2). Sulla scorta del teorema del campionamento possiamo affermare che tale

operazione è invertibile: l'operazione inversa, *conversione n/t*, è la ricostruzione della forma d'onda a partire dai suoi campioni il che, almeno in linea di principio, può essere effettuato utilizzando la serie di Shannon (5.10).

A conferma di quanto detto è immediato stabilire il legame esistente tra lo spettro $X(\nu)$ della sequenza dei campioni e quello $X_\delta(f)$ del segnale campionato $x_\delta(t)$ e quindi con quello $X_a(f)$ della forma d'onda $x_a(t)$. Infatti, utilizzando la proprietà di linearità della Trasformata di Fourier e ricordando che

$$\delta(t - kT) \longleftrightarrow e^{-j2\pi fTk}$$

trasformando entrambi i membri della (5.2) si ottiene:

$$X_\delta(f) = \sum_{n=-\infty}^{\infty} x_a(nT) e^{-j2\pi fTn}$$

che, confrontata con la definizione di $X(\nu)$,

$$X(\nu) = \sum_{n=-\infty}^{\infty} x(n) e^{-j2\pi \nu n}$$

comporta

$$X(\nu) = X_\delta(f)|_{f=\nu/T} \quad X_\delta(f) = X(\nu)|_{\nu=fT} \quad (5.13)$$

Pertanto lo spettro della sequenza in uscita al convertitore t/n coincide con lo spettro del segnale campionato a meno di un cambiamento della scala dell'asse delle frequenze: tale cambiamento è una normalizzazione della frequenza analogica f a quella di campionamento f_c , cioè

$$\nu = fT = \frac{f}{f_c} \quad f = \frac{\nu}{T} = \nu f_c \quad (5.14)$$

In particolare la frequenza di campionamento $f = f_c$, che è il periodo di ripetizione in frequenza delle repliche di $X_a(f)$, corrisponde alla frequenza (numerica) $\nu = 1$, che costituisce il periodo delle trasformate di Fourier di sequenze.

Sostituendo nella (5.13) l'espressione (5.4) dello spettro del segnale campionato, si ottiene il legame cercato tra gli spettri dei segnali in ingresso ed in uscita al convertitore t/n :

$$X(\nu) = \frac{1}{T} \sum_{k=-\infty}^{\infty} X_a\left(\frac{\nu - k}{T}\right) = f_c \sum_{k=-\infty}^{\infty} X_a[(\nu - k)f_c] \quad (5.15)$$

5.3.2 Quantizzazione e codifica

In termini del tutto generali la quantizzazione senza memoria è una corrispondenza

$$Q : x \in \mathcal{R} \rightarrow \hat{x} \in \{\hat{x}_1, \hat{x}_2, \dots, \hat{x}_M\}$$

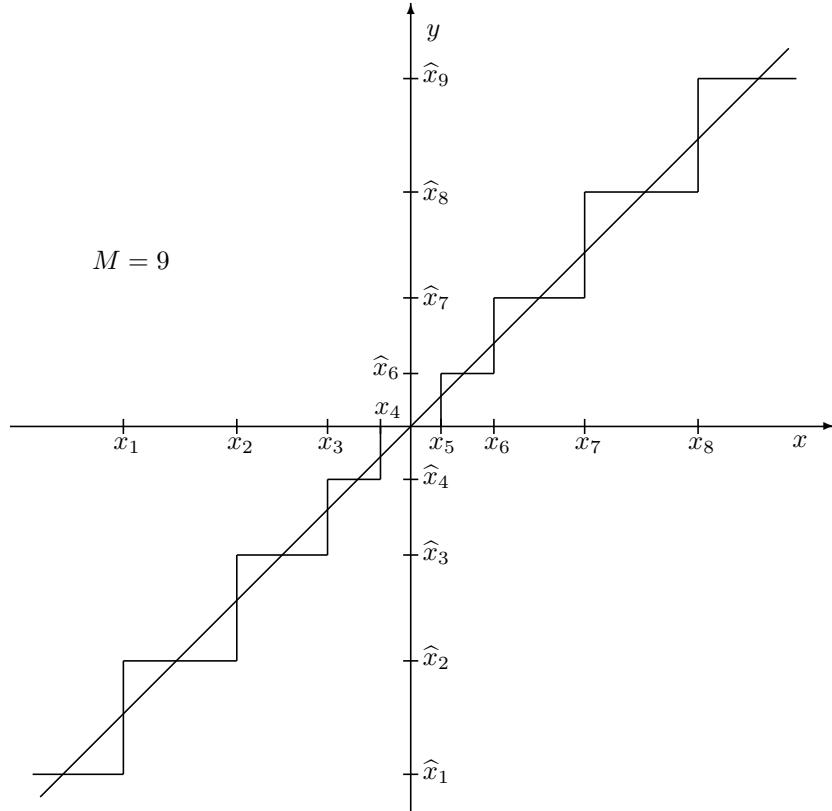


Figura 5.8: Legge di quantizzazione non uniforme.

che trasforma l'insieme continuo \mathcal{R} nell'insieme finito $\{\hat{x}_1, \hat{x}_2, \dots, \hat{x}_M\}$: come illustrato in fig. 5.8, tutti i valori di ingresso interni all'intervallo (x_{i-1}, x_i) , detti *intervallo di quantizzazione*, vengono rappresentati con l'unico valore \hat{x}_i detto *livello di restituzione*, coincidente con il punto medio dell'intervallo di quantizzazione, cioè

$$x \in (x_{i-1}, x_i) \quad \rightarrow \quad \hat{x}_i = \frac{1}{2}(x_{i-1} + x_i) \quad i = 2, \dots, M - 1$$

precisamente ciò è vero per gli intervalli interni, mentre i valori dell'ingresso appartenenti agli intervalli estremi $(-\infty, x_1)$ e $(x_{M-1}, +\infty)$ vengono restituiti rispettivamente con i valori \hat{x}_1 e \hat{x}_M .

I quantizzatori utilizzati nell'elaborazione numerica dei segnali sono in genere *uniformi*, caratterizzati dall'avere gli intervalli quantici interni tutti della stessa ampiezza Δ , detta *passo di quantizzazione*, come illustrato in fig. 5.8. Essi sono completamente definiti da due

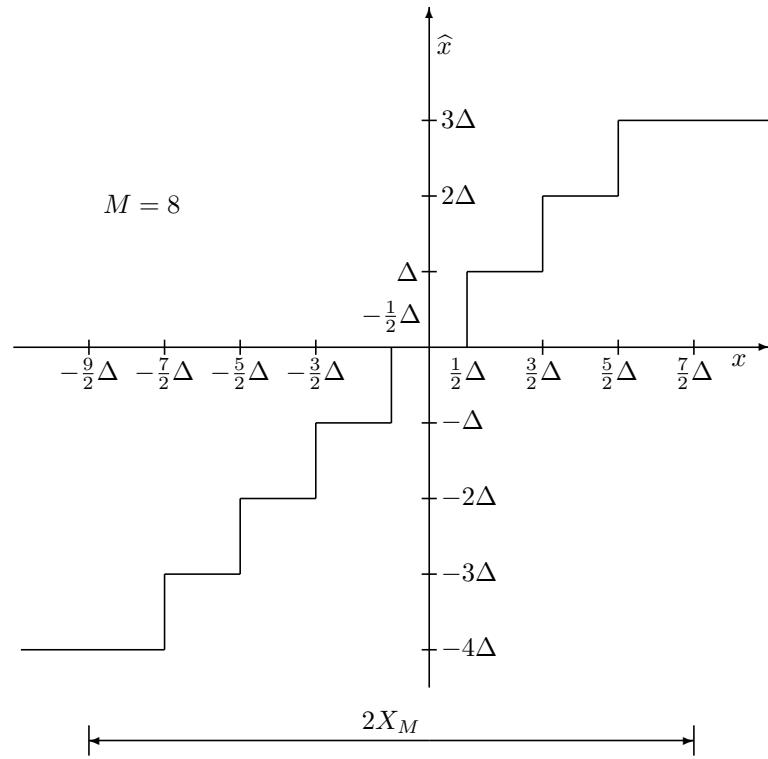


Figura 5.9: Quantizzatore uniforme.

parametri: il passo di quantizzazione Δ e il numero di livelli di restituzione M . Quest'ultimo è pari di norma ad una potenza di 2:

$$M = 2^b$$

dove b è il numero di cifre binarie, o bit, necessarie per codificare i livelli di restituzione. La fig. 5.8 si riferisce ad un quantizzatore a 3 bit, ma in pratica b è più elevato, essendo in genere compreso tra 8 e 16.

È conveniente schematizzare l'effetto della quantizzazione sul segnale $x(n)$ come la risultante dell'azione di un rumore additivo, l'*errore di quantizzazione*

$$e(n) = \hat{x}(n) - x(n)$$

come illustrato in fig. 5.10. È altresì utile suddividere tale rumore in due contributi, il rumore granulare e quello di quantizzazione, in quanto l'uscita $\hat{x}(n)$ del quantizzatore è affetta da errore per due possibili motivi:

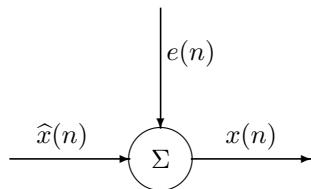


Figura 5.10: Modello di quantizzazione come rumore additivo.

- all'interno dell'intervallo $[-(M+1)\frac{\Delta}{2}, (M-1)\frac{\Delta}{2}]$ (per il quantizzatore di figura), la cui ampiezza

$$2X_M = (M-1)\frac{\Delta}{2} + (M+1)\frac{\Delta}{2} = M\Delta = 2^b\Delta$$

è denominata dinamica, si effettua un arrotondamento al più vicino livello di uscita; conseguentemente l'errore (rumore granulare) si mantiene limitato, e precisamente si ha:

$$|e(n)| \leq \frac{\Delta}{2}$$

- il segnale di ingresso può eccedere la dinamica del quantizzatore, ed in tal caso $e(n)$ non è limitato (rumore di sovraccarico).

La dinamica usualmente è espressa in dB rispetto al minimo livello di restituzione Δ ; in altri termini si pone

$$D = 20 \log_{10} \frac{2X_M}{\Delta} = 20 \log_{10} M = 6b \quad \text{dB}$$

Pertanto essa è proporzionale al numero di bit del quantizzatore ed aumenta di 6 dB per un incremento di un bit.

Per una assegnata densità di probabilità del segnale di ingresso, la dinamica incide sull'entità relativa del rumore di sovraccarico: infatti, per un fissato valore di M , se la dinamica $2X_M = M\Delta$ è elevata, anche Δ è grande, e, corrispondentemente, è notevole l'entità del rumore granulare; viceversa, nel caso di una bassa dinamica, aumenta la frequenza e l'ampiezza dei sovraccarichi.

Per rendere trascurabile la frazione di tempo per cui il quantizzatore è in sovraccarico è necessario che il valore efficace x_{rms} del segnale di ingresso sia sufficientemente inferiore alla dinamica, in altri termini è necessario che sia sufficientemente elevato il fattore di carico K_c definito da

$$K_c = \frac{X_M}{x_{rms}} = \frac{M\Delta}{2x_{rms}} = 2^{b-1} \frac{\Delta}{x_{rms}}$$

Una misura quantitativa dell'errore di quantizzazione è data dal rapporto segnale/rumore (SNR)

$$SNR = 10 \log \frac{E[x^2(n)]}{E[e^2(n)]}$$

Notiamo che il rumore di sovraccarico è essenzialmente l'errore di cimatura già introdotto, e, come tale, può essere analizzato. Per quanto riguarda la componente di rumore granulare si ha che, per Δ/x_{rms} sufficientemente piccolo, essa si può ritenere con ottima approssimazione uniforme in $(-\frac{\Delta}{2}, \frac{\Delta}{2})$ cioè

$$f_e(v) = \frac{1}{\Delta} \Pi\left(\frac{v}{\Delta}\right)$$

Infatti, tenendo presente la caratteristica segnale–errore $e = g(x)$ riportata in fig. 5.11 e derivata dalla caratteristica ingresso–uscita di fig. 5.9, dal teorema fondamentale sulle trasformazioni di variabili aleatorie, se l'ingresso non eccede la dinamica, si ha

$$f_e(v) = \begin{cases} \frac{1}{\Delta} \sum_{i=0}^{M-1} f_x(x_o + i\Delta)\Delta & \text{se } |v| \leq \frac{1}{2}\Delta \\ 0 & \text{altrimenti} \end{cases}$$

ove, $x_o + i\Delta$ $i = 0, \dots, M - 1$ sono le soluzioni dell'equazione $v = g(x)$. La sommatoria all'ultimo membro dell'equazione precedente è interpretabile come l'integrale per rettangoli della pdf del segnale da quantizzare: è quindi facile convincersi che, se Δ/x_{rms} è piccolo, il risultato è con ottima approssimazione uniforme in $(-\frac{\Delta}{2}, \frac{\Delta}{2})$, quale che sia la pdf del segnale da quantizzare, il che giustifica l'ipotesi di indipendenza del rumore di quantizzazione dal segnale di ingresso, comunemente fatta con riferimento al modello di rumore additivo.

Sebbene ciò non sia a rigore vero per il caso del rumore di quantizzazione, da quanto esposto segue che l'indipendenza tra segnale e rumore di quantizzazione è bene approssimata quando il passo di quantizzazione è molto piccolo rispetto a x_{rms} e quando è elevato il fattore di carico K_c in modo da rendere trascurabile il contributo dovuto al rumore di sovraccarico.

Con questa approssimazione, il valor quadratico medio dell'errore granulare vale:

$$E[e^2] = \frac{\Delta^2}{12} \quad (5.16)$$

e, tenuto conto di tale espressione, il rapporto segnale/rumore può essere riscritto come

$$SNR = 10 \log_{10} \left[12 \frac{x_{rms}^2}{\Delta^2} \right] \quad (5.17)$$

Esprimendo poi in tale relazione il rapporto Δ/x_{rms} in funzione di K_c e b in definitiva si ha

$$SNR = 10 \log_{10} \left[3 \frac{2^{2b}}{K_c^2} \right] \simeq 6b - 10 \log_{10} \left(\frac{K_c^2}{3} \right) \quad \text{dB} \quad (5.18)$$

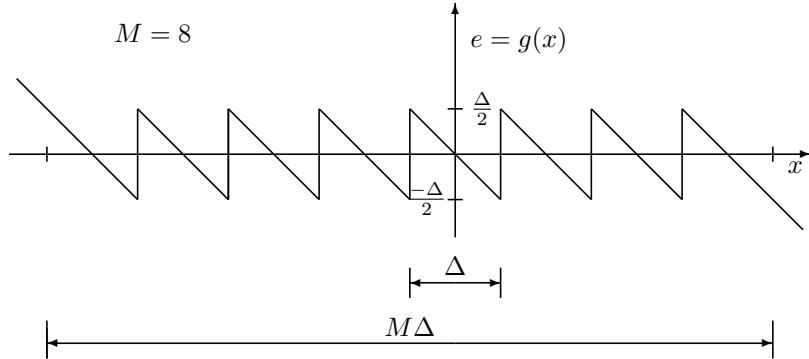


Figura 5.11: Legame segnale / rumore di quantizzazione.

Dalla (5.18) segue che per ogni bit aggiuntivo (raddoppio del numero di livelli di restituzione) il rapporto segnale-rumore si incrementa di 6 dB, indipendentemente da K_c : infatti, il valore di K_c influenza il valore della costante ma non la legge di variazione di SNR rispetto a b . Un elevato valore di K_c garantisce un piccolo tempo di sovraccarico; un valore comune è $K_c = 4$; effettuando tale scelta il SNR diventa

$$SNR = 6.02 b - 7.27 \text{dB}$$

In tal caso si ottiene un rapporto segnale-rumore di poco superiore a 40dB con un quantizzatore a 8 bit, mentre sono necessari 16 bit per avere un rapporto segnale-rumore di quasi 90dB.

Possando a considerare la codifica, si osservi che quest'ultima, essendo una corrispondenza biunivoca tra gli M livelli di restituzione e le sequenze di b bit (parole codice), non dà luogo ad alcuna degradazione delle prestazioni

Nella Tab. 5.1 sono anche indicate due (fra molte) possibili *tabelle di codifica* in binario. Il codice con spiazzamento (*offset*) assegna al generico livello di restituzione la rappresentazione binaria corrispondente all'ordine crescente dal più basso al più alto, con il più basso associato al numero zero. Il codice *complemento a due* si può pensare ottenuto dalla codifica con offset complementando il bit di sinistra (bit più significativo o MSB); con questa codifica il bit più significativo costituisce il bit *segno*, precisamente MSB=1 indica un livello negativo e MSB=0 uno positivo o nullo; in particolare il bit meno significativo (LSB) rappresenta proprio il passo di quantizzazione Δ . Il codice complemento a due è probabilmente il più usato nei convertitori A/D.

Livello di restituzione	Binario con off-set	Complemento a due
\hat{x}_1	000	100
\hat{x}_2	001	101
\hat{x}_3	010	110
\hat{x}_4	011	111
\hat{x}_5	100	000
\hat{x}_6	101	001
\hat{x}_7	110	010
\hat{x}_8	111	011

Tabella 5.1
Alcuni codici binari.

Capitolo 6

Ricezione ottima su canale AWGN

Nel seguito si affronterà lo studio del modulatore e del demodulatore di un sistema di trasmissione numerica. Il modulatore numerico trasmette una tra M possibili forme d'onda $\{s_1(t), s_2(t), \dots, s_M(t)\}$, di durata non superiore all'intervallo di segnalazione T , associabili, attraverso una legge biunivoca, ad una sequenza di $k = \log_2 M$ simboli binari. Al fine di rappresentare in maniera opportuna l'insieme di forme d'onda utilizzato dal modulatore, è necessario rappresentarle come elementi di uno spazio di segnali.

6.1 SPAZIO DEI SEGNALI TRASMESSI

La classe di segnali generati dal modulatore è quella dei segnali reali di quadrato sommabile in $[0, T]$. Tale insieme si denota con il simbolo $\mathbb{L}_{[0,T]}^2$. Un segnale $x(t) \in \mathbb{L}_{[0,T]}^2$ è tale che

$$\int_0^T x^2(t) dt < +\infty \quad (6.1)$$

Si noti che l'unica differenza rispetto alla definizione classica di sommabilità sta nel fatto che l'integrale è esteso a $[0, T]$ e non a tutto l'asse reale, essendo i segnali nulli esternamente a tale intervallo. L'integrale in (6.1) rappresenta l'energia del segnale $x(t)$.

Le proprietà dello spazio dei segnali sono quelle di chiusura rispetto all'operazione di somma e di moltiplicazione per uno scalare:

- $x(t), y(t) \in \mathbb{L}_{[0,T]}^2 \implies x(t) + y(t) \in \mathbb{L}_{[0,T]}^2$
- $\alpha \in \mathbb{R}, x(t) \in \mathbb{L}_{[0,T]}^2 \implies \alpha x(t) \in \mathbb{L}_{[0,T]}^2$

Il *prodotto scalare* sarà dato da

$$\langle x(t), y(t) \rangle = \int_0^T x(t) y(t) dt \quad (6.2)$$

dove, riferendoci solamente a segnali reali, si è omessa l'operazione di coniugazione. La condizione

$$\langle x(t), y(t) \rangle = \int_0^T x(t) y(t) dt = 0 \quad (6.3)$$

determina l'ortogonalità tra due segnali. Risulta evidente che un caso in cui si verifica la condizione di ortogonalità è quando due segnali non sono mai contemporaneamente diversi da zero, ciò vale, ad esempio, per i due segnali di Fig. 6.1:

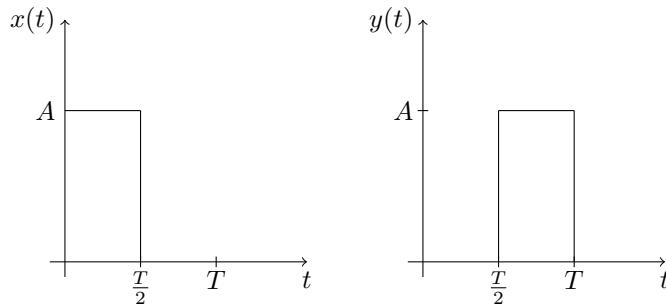


Figura 6.1: Esempio di segnali ortogonali

Si noti come la condizione di non sovrapposizione sia sufficiente per l'ortogonalità ma non necessaria; è possibile, infatti, costruire segnali che sono ortogonali pur essendo contemporaneamente diversi da zero, come per esempio i due segnali di Fig. 6.2.

Inoltre, dall'uguaglianza di Parseval:

$$\int_0^T x(t) y(t) dt = \int_{-\infty}^{+\infty} X(f) Y(f) df \quad (6.4)$$

si ricava che $x(t)$ e $y(t)$ sono ortogonali se le loro trasformate di Fourier non sono sovrapposte in frequenza. Ad esempio, se consideriamo le due \mathcal{F} -trasformate di Fig. 6.3, possiamo dedurre, per il teorema di Parseval (6.4), che $x(t)$ e $y(t)$ sono ortogonali.

La *norma* di un segnale è definita come

$$\|x(t)\| = \langle x(t), x(t) \rangle^{1/2} = \left(\int_0^T x^2(t) dt \right)^{1/2} = \sqrt{\mathcal{E}_x} \quad (6.5)$$

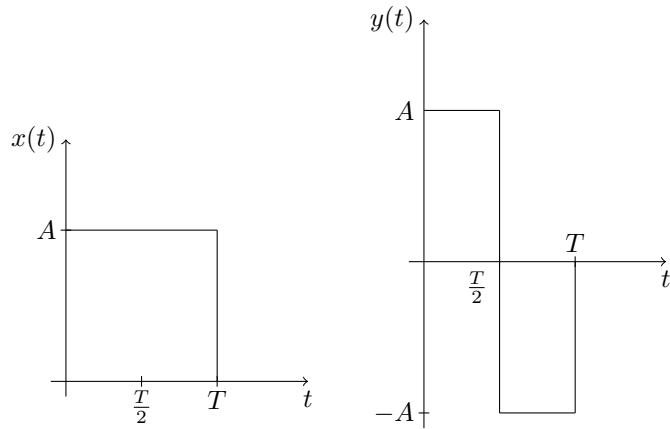


Figura 6.2: Esempio di segnali ortogonali pur essendo contemporaneamente diversi da zero

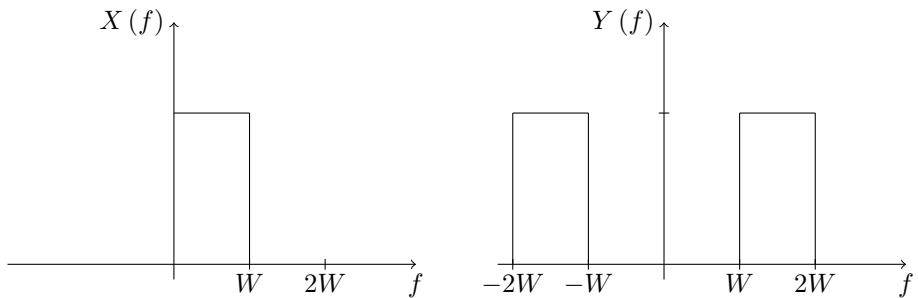


Figura 6.3: Esempio di segnali ortogonali in frequenza

dove con \mathcal{E}_x si indica l'energia del segnale $x(t)$ e la la distanza tra due segnali è data dalla norma della loro differenza

$$\begin{aligned}
 d^2(x(t), y(t)) &= \|x(t) - y(t)\| = \\
 &= \langle x(t) - y(t), x(t) - y(t) \rangle = \\
 &= \int_0^T (x(t) - y(t))^2 dt = \\
 &= \mathcal{E}_x + \mathcal{E}_y - 2\mathcal{E}_{xy}
 \end{aligned} \tag{6.6}$$

Base in $\mathbb{L}_{[0,T]}^2$

Il nostro obiettivo è studiare i segnali del modulatore e pertanto consideriamo solo un sottinsieme di tutti i possibili segnali di $\mathbb{L}_{[0,T]}^2$, cioè quelli generati dal modulatore $\mathcal{S} = \{s_1(t), s_2(t), \dots, s_M(t)\}$. Un insieme di N segnali linearmente indipendenti $\mathcal{B} = \{\psi_1(t), \psi_2(t), \dots, \psi_N(t)\}$ si dice base dello spazio dei segnali \mathcal{S} se ogni segnale $s_m(t) \in \mathcal{S}$ può essere espresso come combinazione lineare dei segnali $\psi_i(t) \in \mathcal{B}$, cioè

$$s_i(t) = \sum_{k=1}^N s_{ik} \psi_k(t), \quad i \in \{1, 2, \dots, M\}. \quad (6.7)$$

I coefficienti s_{ik} rappresentano le proiezioni del segnale $s_i(t)$ sul versore $\psi_k(t)$ che saranno date dal prodotto scalare

$$s_{ik} = \langle s_i(t), \psi_k(t) \rangle = \int_0^T s_i(t) \psi_k(t) dt \quad (6.8)$$

quindi il generico segnale $s_i(t)$ può essere rappresentato come

$$s_i(t) = \sum_{k=1}^N s_{ik} \psi_k(t) \quad (6.9)$$

Cerchiamo di capire perché i coefficienti s_{ik} sono proprio le proiezioni del segnale $s_i(t)$ lungo i versori $\psi_k(t)$. In genere, il modo di determinare una base è quello di scegliere i segnali in maniera tale che siano a due a due ortogonali e a norma unitaria. Per verificare queste proprietà della base si deve avere

$$\psi_{ij} = \langle \psi_i(t), \psi_j(t) \rangle = \begin{cases} 0 & \text{se } i \neq j \\ 1 & \text{se } i = j \end{cases} \quad (6.10)$$

La (6.10) può essere sinteticamente espressa come

$$\psi_{ij} = \langle \psi_i(t), \psi_j(t) \rangle = \delta(i - j) = \delta_{ij} \quad (6.11)$$

dove δ è il simbolo di Kroneker. Una base \mathcal{B} che soddisfa la (6.10) si dice ortonormale. Calcoliamo ora il prodotto scalare tra il segnale $s_i(t)$ e un vettore $\psi_{ij}(t)$ della base

$$\langle s_i(t), \psi_j(t) \rangle = \left\langle \sum_{k=1}^N s_{ik} \psi_k(t), \psi_j(t) \right\rangle = \sum_{k=1}^N s_{ik} \langle \psi_k(t), \psi_j(t) \rangle = s_{ij} \quad (6.12)$$

Quindi la generica forma d'onda $s_i(t)$ del modulatore può essere ottenuta come combinazione lineare delle funzioni di base. I coefficienti della combinazione sono proprio le proiezioni di $s_i(t)$ lungo i versori della base.

Analogamente a quanto fatto nello spazio vettoriale \mathbb{R}^N , è possibile identificare il segnale $s_i(t)$ con il vettore N -dimensionale \mathbf{s}_i , i cui elementi sono

$$\mathbf{s}_i = \begin{bmatrix} s_{i1} \\ s_{i2} \\ \vdots \\ s_{iN} \end{bmatrix} = \begin{bmatrix} \langle s_i(t), \psi_1(t) \rangle \\ \langle s_i(t), \psi_2(t) \rangle \\ \vdots \\ \langle s_i(t), \psi_N(t) \rangle \end{bmatrix} \quad (6.13)$$

Una volta passati dal segnale $s_i(t)$ alla sua rappresentazione vettoriale \mathbf{s}_i , è possibile verificare la conservazione delle proprietà.

A tal fine consideriamo due segnali $s_m(t)$ e $s_n(t)$ appartenenti e calcoliamone il prodotto scalare:

$$\begin{aligned} \langle s_m(t), s_n(t) \rangle &= \left\langle \sum_{i=1}^L s_{mi} \psi_i(t), \sum_{j=1}^L s_{nj} \psi_j(t) \right\rangle = \\ &= \sum_{i=1}^L \sum_{j=1}^L s_{mi} s_{nj} \langle \psi_i(t), \psi_j(t) \rangle = \\ &= \sum_{i=1}^L s_{mi} s_{ni} = \\ &= \langle \mathbf{s}_m, \mathbf{s}_n \rangle. \end{aligned} \quad (6.14)$$

Dalla (6.14) deduciamo che il prodotto scalare si conserva. Risulta quindi:

$$\langle s_m(t), s_n(t) \rangle = \int_0^T s_m(t) s_n(t) dt = \langle \mathbf{s}_m, \mathbf{s}_n \rangle \quad (6.15)$$

Questo risultato è di notevole importanza perché significa che possiamo calcolare il prodotto scalare tra due segnali (che comporterebbe il calcolo dell'integrale (6.2)) come il prodotto scalare tra due vettori (somma dei prodotti delle componenti omonime dei due vettori), il che è in generale più semplice.

Dalla conservazione del prodotto scalare discende la conservazione della norma, e quindi dell'energia. Risulta infatti:

$$\mathcal{E}_x = \int_0^T x^2(t) dt = \|x(t)\|^2 = \langle x(t), x(t) \rangle = \langle \mathbf{x}, \mathbf{x} \rangle = \|\mathbf{x}\|^2 \quad (6.16)$$

Inoltre, poiché che la norma di un vettore è la sua distanza dall'origine, deduciamo che l'energia del segnale è il quadrato della distanza tra il punto rappresentativo del segnale e l'origine, quindi l'energia è associata ad una grandezza geometrica.

Ad esempio, supponendo di avere una base bidimensionale $\mathcal{B} = \{\psi_1(t), \psi_2(t)\}$, possiamo rappresentare il segnale $x(t)$ con un punto di componenti (x_1, x_2) tale che $x(t) = x_1\psi_1(t) + x_2\psi_2(t)$.

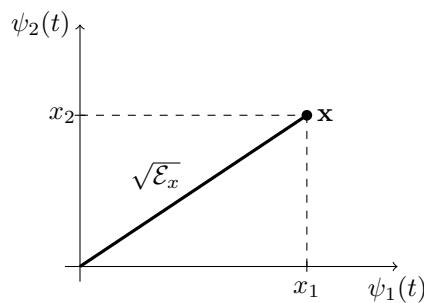


Figura 6.4: Rappresentazione del segnale $x(t)$ nello spazio dei segnali

6.1.0.1 Distanza

La distanza tra due segnali è definita come

$$d(x(t), y(t)) = \|x(t) - y(t)\| = \|\mathbf{x} - \mathbf{y}\| = d(\mathbf{x}, \mathbf{y}) \quad (6.17)$$

da cui deduciamo che anche la distanza può essere calcolata dai vettori rappresentativi, come mostrato in figura 6.5.

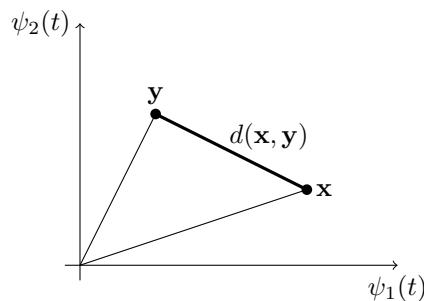


Figura 6.5: Distanza tra i vettori \mathbf{x} e \mathbf{y}

quindi la distanza tra due segnali è proprio la distanza euclidea tra i vettori rappresentativi dei segnali nello spazio dei segnali.

6.1.1 Costellazione dei segnali

Definita la base \mathcal{B} possiamo proiettare i segnali del modulatore lungo tale base ed averne così una rappresentazione geometrica nello spazio dei segnali. Passiamo così dall'insieme \mathcal{S} dei segnali all'insieme \mathcal{S}' dei corrispondenti vettori rappresentativi:

$$\mathcal{S} = \{s_1(t), s_2(t), \dots, s_M(t)\} \iff \mathcal{S}' = \{\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_M\} \quad (6.18)$$

Ricordiamo che la base \mathcal{B} ha dimensionalità L , mentre l'insieme \mathcal{S} dei possibili segnali del modulatore ha dimensionalità M , con $L \leq M$. L'uguaglianza si ha nel caso in cui il set di segnali \mathcal{S} del modulatore sono linearmente indipendenti, ossia costituiscono una base.

L'insieme \mathcal{S}' dei vettori rappresentativi dei segnali di \mathcal{S} si chiama **costellazione** dei segnali del modulatore.

6.2 PROCEDURA DI ORTONORMALIZZAZIONE DI GRAM-SCHMIDT

Si pone il problema di come individuare una base tale che, dalle combinazioni lineari dei suoi elementi, si ottengono i segnali del modulatore. Abbiamo visto che, nota una base per un dato insieme di segnali, è possibile rappresentare tutti i segnali per mezzo delle sole proiezioni sui versori della base. Per far ciò è necessario costruire la base a partire dall'insieme di segnali da rappresentare.

Per costruire la base si segue una procedura costruttiva di ortogonalizzazione: partendo da un certo insieme di segnali, si cerca un insieme di segnali a norma unitaria e a due a due ortogonali che ci consentono di rappresentare tutti i segnali dell'insieme di partenza.

Applichiamo inizialmente la procedura in uno spazio vettoriale in modo da poter illustrare anche geometricamente il procedimento. Estenderemo poi la procedura al caso di spazio dei segnali.

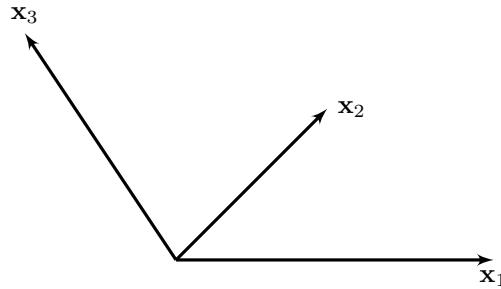
6.2.1 Ortonormalizzazione in \mathbb{R}^2

Si considerino i tre vettori $\{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3\}$ co-planari mostrati in fig. ??.

L'algoritmo di ortogonalizzazione si compone dei seguenti passi:

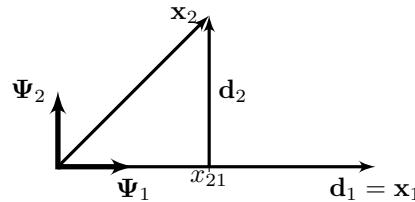
- 1.▷ Si definisce un vettore ausiliario $\mathbf{d}_1 = \mathbf{s}_1$ e si costruisce il primo versore della base, parallelo al vettore \mathbf{s}_1 e a norma unitaria

$$\Psi_1 = \frac{\mathbf{s}_1}{\|\mathbf{s}_1\|}$$

Figura 6.6: Vettori $\{x_1, x_2, x_3\}$ nello spazio vettoriale \mathbb{R}^2

2.▷ Sottraendo da s_2 la sua parte rappresentabile attraverso Ψ_1 (cioè la proiezione di x_2 su Ψ_1) otteniamo un vettore ortogonale a Ψ_1 (cfr. fig. 6.7).

$$\mathbf{d}_2 = \mathbf{s}_2 - s_{21}\Psi_1$$

Figura 6.7: Algoritmo di ortonormalizzazione di Gram-Schmidt – Individuazione del secondo vettore: Ψ_2

che normalizzato fornisce il secondo versore della base

$$\Psi_2 = \frac{\mathbf{d}_2}{\|\mathbf{d}_2\|}$$

3.▷ L'iterazione procede costruendo un terzo vettore \mathbf{d}_3 ottenuto sottraendo ad \mathbf{s}_3 le sue proiezioni lungo Ψ_1 e Ψ_2

$$\mathbf{d}_3 = \mathbf{s}_3 - s_{31}\Psi_1 - s_{32}\Psi_2$$

Nel caso specifico, essendo i due vettori co-planari risulta

$$\mathbf{s}_3 - (s_{31}\Psi_1 + s_{32}\Psi_2) = 0$$

In generale, il procedimento consiste nel calcolare il vettore ausiliario

$$\mathbf{d}_i = \mathbf{s}_i - \sum_{n=1}^{i-1} s_{in} \Psi_n$$

che normalizzato fornisce

$$\Psi_i = \frac{\mathbf{d}_i}{\|\mathbf{d}_i\|}$$

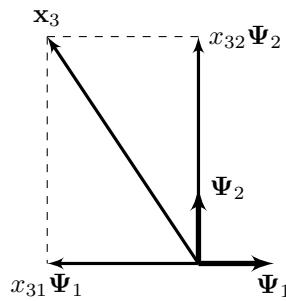


Figura 6.8: Vettore \mathbf{x}_3 ottenuto dalla combinazione lineare dei versori Ψ_1 e Ψ_2

Per i segnali la procedura si ripete in maniera del tutto analoga a quanto visto per lo spazio vettoriale \mathbb{R}^N :

$$d_i(t) = s_i(t) - \sum_{k=1}^{i-1} s_{ik} \psi_k(t), \quad i = 1, 2, \dots, M \quad (6.19)$$

da cui ricaviamo le funzioni di base

$$\psi_i(t) = \frac{d_i(t)}{\|d_i(t)\|} \quad (6.20)$$

che costituiscono la base $\mathcal{B} = \{\psi_1(t), \psi_2(t), \dots, \psi_L(t)\}$, con $L \leq M$.

La dimensione della base è minore o al più uguale ad M laddove qualcuno dei versori risulti nullo. Se $L = M$, gli M segnali del modulatore sono linearmente indipendenti.

Modulazione 4-PSK.

Consideriamo i quattro segnali

$$\begin{aligned} s_1(t) &= A \cos\left(\frac{2\pi}{T}t\right) \\ s_2(t) &= A \cos\left(\frac{2\pi}{T}t - \frac{\pi}{2}\right) = A \sin\left(\frac{2\pi}{T}t\right) \\ s_3(t) &= A \cos\left(\frac{2\pi}{T}t - \pi\right) = -A \cos\left(\frac{2\pi}{T}t\right) \quad t \in (0, T) \\ s_4(t) &= A \cos\left(\frac{2\pi}{T}t - \frac{3\pi}{2}\right) = -A \sin\left(\frac{2\pi}{T}t\right) \end{aligned} \quad (6.21)$$

rappresentati in fig. 6.9.

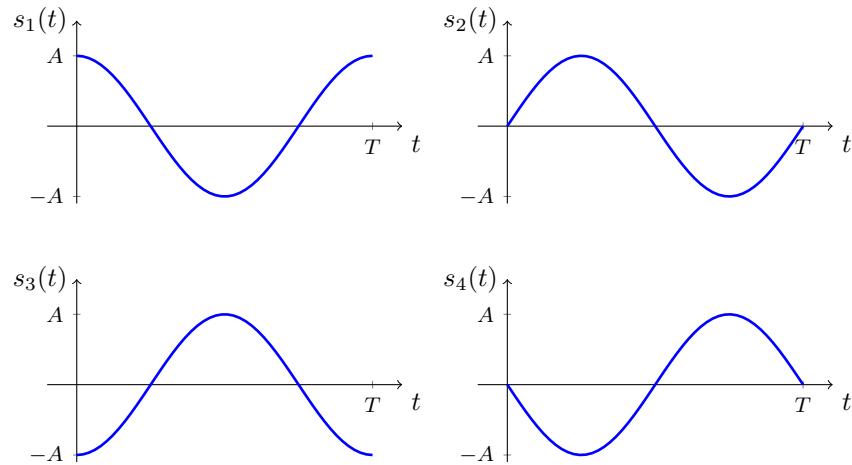


Figura 6.9: Forme d'onda generate per una segnalazione 4-PSK

Applichiamo il procedimento di Gram-Schmidt al fine di ottenere un sistema di versori ortonormali che permetta la rappresentazione dei quattro segnali di energia $s_1(t)$, $s_2(t)$, $s_3(t)$ ed $s_4(t)$ nello spazio generato dai versori.

1▷

$$\psi_1(t) = \frac{s_1(t)}{\|s_1(t)\|} = \frac{A \cos\left(\frac{2\pi}{T}t\right)}{\sqrt{\frac{A^2 T}{2}}} = \sqrt{\frac{2}{T}} \cos\left(\frac{2\pi}{T}t\right)$$

2▷

$$d_2(t) = s_2(t) - s_{21}\psi_1(t) = A \sin\left(\frac{2\pi}{T}t\right)$$

infatti risulta

$$\begin{aligned} s_{21} &= \langle s_2(t), \psi_1(t) \rangle = \int_0^T A \sin\left(\frac{2\pi}{T}t\right) \sqrt{\frac{2}{T}} \cos\left(\frac{2\pi}{T}t\right) dt \\ &= A \sqrt{\frac{2}{T}} \int_0^T \sin\left(\frac{4\pi}{T}t\right) dt = 0 \end{aligned}$$

da cui

$$\psi_2(t) = \frac{d_2}{\|d_2\|} = \sqrt{\frac{2}{T}} \sin\left(\frac{2\pi}{T}t\right)$$

3▷

$$d_3(t) = s_3(t) - s_{31}\psi_1(t) - s_{32}\psi_2(t) \quad (6.22)$$

$$\begin{aligned} s_{31} &= \langle s_3(t), \psi_1(t) \rangle = \int_0^T -A \cos\left(\frac{2\pi}{T}t\right) \frac{2}{T} \cos\left(\frac{2\pi}{T}t\right) dt \\ &= -\frac{A\sqrt{2}}{\sqrt{T}} \frac{T}{2} = -A\sqrt{\frac{T}{2}} \end{aligned}$$

$$s_{32} = \langle s_3(t), \psi_2(t) \rangle = \int_0^T -A \cos\left(\frac{2\pi}{T}t\right) \sqrt{\frac{2}{T}} \sin\left(\frac{2\pi}{T}t\right) dt = 0$$

per cui sostituendo nella (6.22) otteniamo

$$d_3(t) = -A \cos\left(\frac{2\pi}{T}t\right) + A \cos\left(\frac{2\pi}{T}t\right) = 0$$

$$\psi_3(t) = \frac{d_3}{\|d_3\|} = 0$$

quindi il segnale $s_3(t)$ è rappresentabile mediante i precedenti versori.

4▷

$$d_4 = s_4(t) - s_{41}\psi_1(t) - s_{42}\psi_2(t)$$

sviluppando i calcoli si ricava

$$s_{41} = \langle s_4(t), \psi_1(t) \rangle = 0$$

$$s_{42} = \langle s_4(t), \psi_2(t) \rangle = -A\sqrt{\frac{T}{2}}$$

$$d_4(t) = -A \sin\left(\frac{2\pi}{T}t\right) + A \sin\left(\frac{2\pi}{T}t\right) = 0$$

Dalla procedura di Gram-Schmidt deduciamo che la dimensione dello spazio dei segnali è uguale a due in quanto i segnali $s_3(t)$ ed $s_4(t)$ possono essere espressi come combinazione lineare dei versori $\psi_1(t)$ e $\psi_2(t)$. In particolare, dati i due versori $\psi_1(t) = \sqrt{\frac{2}{T}} \cos\left(\frac{2\pi}{T}t\right)$ e $\psi_2(t) = \sqrt{\frac{2}{T}} \sin\left(\frac{2\pi}{T}t\right)$, le coordinate dei quattro segnali risultano essere

$$\begin{aligned}\mathbf{s}_1 &= \left(A\sqrt{\frac{T}{2}}, 0\right) = \left(\sqrt{\mathcal{E}_s}, 0\right) \\ \mathbf{s}_2 &= \left(0, A\sqrt{\frac{T}{2}}\right) = \left(0, \sqrt{\mathcal{E}_s}\right) \\ \mathbf{s}_3 &= \left(-A\sqrt{\frac{T}{2}}, 0\right) = \left(-\sqrt{\mathcal{E}_s}, 0\right) \\ \mathbf{s}_4 &= \left(0, -A\sqrt{\frac{T}{2}}\right) = \left(0, -\sqrt{\mathcal{E}_s}\right)\end{aligned}$$

Da ciò si deduce che i segnali sono rappresentabili in uno spazio bidimensionale e risultano situati su una circonferenza equidistanziati di un angolo pari a $\frac{\pi}{2}$ come mostrato in fig. ??.

In generale, uno schema di segnalazione M-PSK si ottiene generalizzando le (6.21) cioè ponendo

$$s_i(t) = A \cos\left(\frac{2\pi}{T}t - \frac{2\pi i}{M}\right) \quad (6.23)$$

I segnali sono ancora rappresentabili su uno spazio a due dimensioni per cui risulta $N = 2$ e sono equispaziati di un angolo $2\pi/M$ su una circonferenza di raggio pari a $\sqrt{\mathcal{E}_s}$.

Per il caso 8-PSK, la costellazione di segnali è mostrata in fig. 6.11.

6.3 RICEZIONE OTTIMA IN RUMORE AWGN

Il problema della ricezione ottima di segnali consiste nell'individuare la struttura del ricevitore che assicura le migliori prestazioni possibili. Affinchè il problema sia ben posto, occorre individuare un criterio di ottimalità, considerare alcuni vincoli e scegliere un modello di canale. Il criterio di ottimalità che sceglieremo consiste nella minimizzazione della probabilità di errore cioè della probabilità di decidere per un simbolo diverso da quello trasmesso. Adotteremo

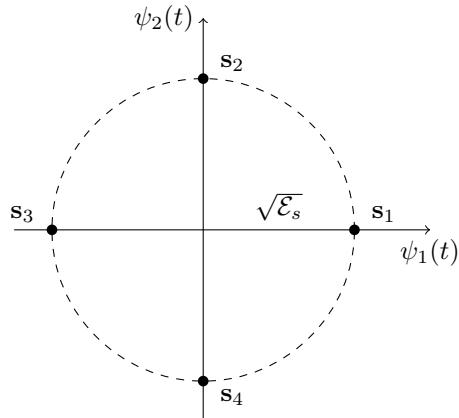


Figura 6.10: Costellazione 4-PSK

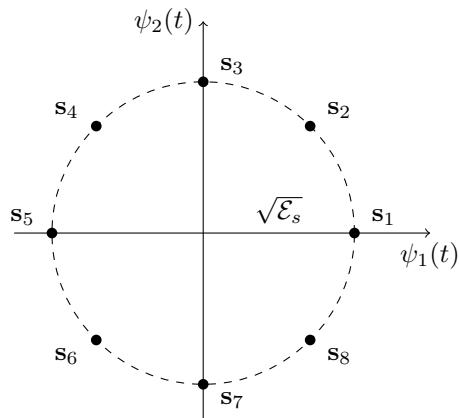


Figura 6.11: Costellazione 8-PSK

inoltre, per il momento, il modello AWGN per il canale di trasmissione, riservandoci di introdurre successivamente un modello di canale rumoroso distorcente. Consideriamo inoltre le seguenti ipotesi semplificative

1. Il modulatore è senza memoria; cioè, in ogni intervallo di durata T , il modulatore trasmette una forma d'onda, scelta tra quelle disponibili nell'alfabeto, in maniera indipendente da quelle trasmesse in precedenza.
2. I segnali trasmessi sono reali e hanno durata rigorosamente limitata ed uguale a T .

Dunque, non si ha interferenza fra i segnali trasmessi in un dato intervallo di tempo $(nT, (n+1)T)$ e quelli trasmessi negli intervalli precedenti fig. ???. Se la condizione di durata limitata del segnale viene a cadere, ad esempio a causa della presenza di un canale distorcente, si ha una situazione come quella mostrata in fig. ?? in cui il segnale in un dato intervallo di simbolo subisce interferenza da parte dei segnali trasmessi precedentemente. Si noti che, la mancanza di interferenza tra il simbolo attuale e quelli precedenti è essenziale affinchè la decisione ottima possa essere effettuata *simbolo a simbolo*.

3. Il canale è schematizzabile con un modello AWGN. Si noti ancora come l'ipotesi di indipendenza tra i campioni del rumore sia essenziale in quanto una correlazione tra i campioni del rumore introdurrebbe una correlazione sul segnale ricevuto $r(t) = s_i(t) + n(t)$ e quindi una dipendenza statistica del simbolo ricevuto da quelli precedenti.

Segnalazione 8-PSK

Consideriamo lo schema di segnalazione mostrato in fig. ?? ed analizziamo in dettaglio la funzione dei singoli blocchi.

- ▷ Sorgente binaria: genera una sequenza di simboli binari indipendenti.
- ▷ Modulatore: associa ad ogni blocco di tre simboli binari un segnale del tipo

$$s_i(t) = A \cos\left(\frac{2\pi}{T}t - \phi_i\right) \quad \phi_i = \frac{2\pi i}{8} \quad i = 0, \dots, 7; \quad t \in (nT, (n+1)T)$$

- ▷ Proiettore: effettua la proiezione del segnale ricevuto $r(t)$ sui versori della base. In altri termini, ciò significa effettuare i prodotti scalari $\langle r(t), \psi_i(t) \rangle$, con

$$\begin{aligned} \psi_1(t) &= \sqrt{\frac{2}{T}} \cos\left(\frac{2\pi}{T}t\right) & t \in (0, T) \\ \psi_2(t) &= \sqrt{\frac{2}{T}} \sin\left(\frac{2\pi}{T}t\right) \end{aligned}$$

In assenza di rumore abbiamo otto possibili punti (fig. 6.11) nello spazio a due dimensioni. In presenza di rumore, all'uscita del proiettore avremo

$$r_{ij} = \langle s_i(t), \psi_j(t) \rangle + \langle n(t), \psi_j(t) \rangle = s_{ij} + n_j$$

Nell'ipotesi in cui il segnale trasmesso sia $s_1(t)$, possiamo scrivere

$$r_{11} = \langle s_1(t), \psi_1(t) \rangle + \langle n(t), \psi_1(t) \rangle = s_{11} + n_1$$

$$r_{12} = \langle s_1(t), \psi_2(t) \rangle + \langle n(t), \psi_2(t) \rangle = s_{12} + n_2$$

dove

$$n_1 = \int_0^T n(t)\psi_1(t) dt$$

$$n_2 = \int_0^T n(t)\psi_2(t) dt$$

In termini vettoriali, il vettore risultante \mathbf{r} è dato dalla somma dei due vettori rappresentativi del segnale trasmesso e del rumore additivo. Gli effetti del rumore consistono dunque nel modificare la posizione del segnale ricevuto in maniera tale che le coordinate lungo i versori dello spazio dei segnali risultano essere la somma delle proiezioni della componente di segnale e della componente di rumore lungo gli stessi assi. Graficamente, gli effetti del rumore sono mostrati in fig. 6.12.

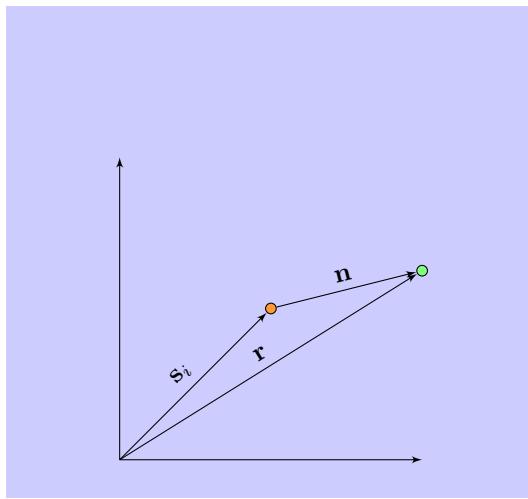


Figura 6.12: Rappresentazione dei segnali in presenza di rumore

Il compito del ricevitore è, dunque, quello di effettuare una decisione su quale segnale è stato trasmesso a partire dalla proiezione del segnale ricevuto sui versori dello spazio in cui vengono rappresentati i segnali trasmessi.

L'esempio precedente evidenzia come, a valle del sistema di proiezione, sia possibile associare al segnale ricevuto un vettore appartenente allo spazio dei segnali trasmessi. Si noti tuttavia che, mentre la procedura di costruzione dello spazio dei segnali ne garantisce per definizione la loro rappresentabilità, ciò non si può dire per il rumore, che in generale presenterà componenti esterne a tale spazio. Mostreremo in seguito che, nel caso di canale AWGN, è possibile ottenere una struttura di ricezione ottima a partire dalle sole proiezioni del segnale ricevuto.

A tal fine è vantaggioso suddividere il ricevitore in due blocchi funzionali, come mostrato in fig. 6.13:

- ◊ Proiettore: effettua la proiezione del segnale ricevuto $r(t)$ nello spazio dei segnali ed in tal modo fornisce una *statistica sufficiente* ai fini della decisione ottima.
- ◊ Decisore: sulla base della statistica sufficiente fornita dal proiettore, il blocco di decisione individua quale tra gli M possibili segnali trasmessi garantisce la minima probabilità di errore.

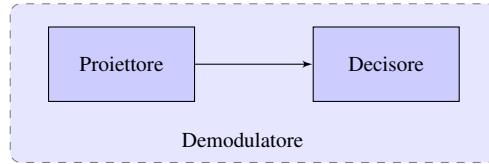


Figura 6.13: Struttura del demodulatore

Nel descrivere la strategia di ricezione procederemo nel modo seguente:

1. Individueremo una struttura generale del proiettore.
2. Ricaveremo una caratterizzazione statistica congiunta delle proiezioni.
3. Mostreremo che le proiezioni rappresentano una statistica sufficiente.
4. Ricaveremo una struttura generale del decisore ottimo.

Dati M segnali $\{s_i(t)\}$, $i = 1 \dots M$ ed una base ortonormale $\{\psi_i(t)\}$, $i = 1 \dots N$ nella quale essi sono rappresentabili, le proiezioni r_k del segnale ricevuto risultano essere

$$r_k = \langle r(t), \psi_k(t) \rangle = \int_0^T r(t) \psi_k(t) dt \quad k = 1, \dots, N \quad N \leq M \quad (6.24)$$

L'equazione (6.24) definisce la struttura mostrata in fig. 6.14.

Consideriamo ora il vettore delle proiezioni $\mathbf{r} = (r_1, r_2, \dots, r_N)$ le cui componenti, partendo dall'eq. (6.24) sono esprimibili come

$$\begin{aligned} r_k &= \int_0^T r(t) \psi_k(t) dt = \int_0^T [s_i(t) + n(t)] \psi_k(t) dt \\ &= \int_0^T s_i(t) \psi_k(t) dt + \int_0^T n(t) \psi_k(t) dt = s_{ik} + n_k \end{aligned}$$

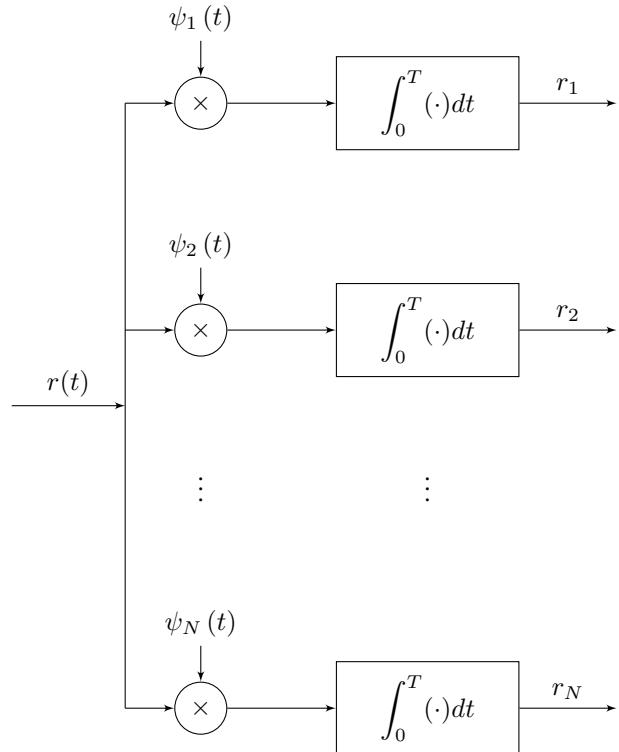


Figura 6.14: Struttura del proiettore

dove $n(t)$ è un processo di rumore gaussiano bianco con densità spettrale di potenza pari ad $N_0/2$. Cominciamo con l'osservare che \mathbf{r} è un vettore gaussiano in quanto ottenuto mediante una elaborazione lineare di un processo gaussiano. Per caratterizzare completamente il vettore è quindi sufficiente valutarne la media e le covarianze tra i singoli elementi. Con riferimento alle componenti di rumore si ha

$$E[n_k] = E \int_0^T n(t)\psi_k(t) dt = \int_0^T E[n(t)]\psi_k(t) dt = 0$$

Essendo le medie statistiche tutte nulle, la covarianza si può calcolare come

$$\begin{aligned}
 E[n_k n_h] &= E \left\{ \int_0^T n(t) \psi_k(t) dt \int_0^T n(\tau) \psi_h(\tau) d\tau \right\} \\
 &= \int_0^T \int_0^T E[n(t) n(\tau)] \psi_k(t) \psi_h(\tau) dt d\tau \\
 &= \int_0^T \int_0^T r_n(t - \tau) \psi_k(t) \psi_h(\tau) dt d\tau \\
 &= \int_0^T \int_0^T \frac{N_0}{2} \delta(t - \tau) \psi_k(t) \psi_h(\tau) d\tau dt \\
 &= \frac{N_0}{2} \int_0^T \psi_k(\tau) \psi_h(\tau) d\tau \\
 &= \frac{N_0}{2} \delta_{kh}
 \end{aligned} \tag{6.25}$$

dove δ_{kh} è il simbolo di Kroneker, definito come

$$\delta_{kh} = \begin{cases} 1 & \text{se } k = h \\ 0 & \text{se } k \neq h \end{cases} \tag{6.26}$$

In definitiva, quindi, le variabili aleatorie n_k , $k = 1, \dots, N$, risultano essere mutuamente in-correlate, gaussiane a media nulla e varianza $N_0/2$. Di conseguenza, le proiezioni r_k , condizionatamente alla trasmissione del segnale s_i , sono variabili aleatorie gaussiane, indipendenti a media s_{ik} e varianza $N_0/2$, sinteticamente si ha $r_k \sim \mathcal{N}(s_{ik}, \frac{N_0}{2})$ oppure, in forma esplicita,

$$f_{\mathbf{r}|s_i}(\mathbf{r}|s_i) = \frac{1}{(\pi N_0)^{\frac{N}{2}}} \exp \left\{ -\frac{1}{N_0} \sum_{k=1}^N (r_k - s_{ik})^2 \right\} \tag{6.27}$$

Da quanto precede ne deriva che il rumore non è rappresentabile nello spazio dei segnali per cui, detta $n'(t)$ l'aliquota di rumore che giace nello spazio dei segnali ed $n''(t)$ quella che giace al di fuori di tale spazio, possiamo scrivere

$$\begin{aligned}
 r(t) &= s_i(t) + n(t) = s_i(t) + [n'(t) + n''(t)] \\
 &= \sum_{k=1}^N (s_{ik} + n_k) \psi_k(t) + n''(t)
 \end{aligned} \tag{6.28}$$

È possibile mostrare che il rumore che giace al di fuori dello spazio dei segnali è *irrilevante* ai fini della decisione statistica.

Il secondo blocco funzionale che compone il ricevitore è il decisore che, come indicato in fig. ??, effettua una decisione sul simbolo trasmesso a partire dalla statistica fornita dal proiettore. Ci proponiamo quindi ricavare il decisore che rende minima la probabilità di errore che, utilizzando il teorema della probabilità totale può essere scritta come

$$P(e) = \sum_{i=1}^M P(e|s_i)P(s_i) \quad (6.29)$$

dove $P(s_i)$ sono le probabilità *a priori* di trasmettere i simboli s_i . Osserviamo che il vettore $\mathbf{r} = (r_1, \dots, r_N)$ individua un punto nello spazio ad N dimensioni R^N e che ogni punto di tale spazio corrisponde ad una configurazione di segnale più rumore ricevuti. Lo spazio può essere partizionato in M regioni di decisione \mathcal{R}_i definite come l'insieme dei punti in cui il decisore sceglie per il segnale s_i , cioè

$$\mathbf{r} \in \mathcal{R}_i \Leftrightarrow \hat{s} = s_i \quad (6.30)$$

La figura 6.15 mostra come esempio lo spazio R^N partizionato in quattro regioni di decisione corrispondenti ad altrettanti possibili segnali trasmessi. Riferendoci ad una generica partizione dello spazio dei segnali, possiamo riscrivere l'espressione della probabilità di errore come

$$\begin{aligned} P(e) &= \sum_{i=1}^M P(e|s_i)P(s_i) = \sum_{i=1}^M (1 - P(c|s_i))P(s_i) \\ &= 1 - \sum_{i=1}^M P(c|s_i)P(s_i) = 1 - \sum_{i=1}^M P\{\mathbf{r} \in \mathcal{R}_i | s_i\} P(s_i) \\ &= 1 - \sum_{i=1}^M \int_{\mathcal{R}_i} P(s_i) f_{\mathbf{r}|s_i}(\mathbf{r}|s_i) d\mathbf{r} \end{aligned} \quad (6.31)$$

dove $P(c|s_i) = 1 - P(e|s_i)$ è la probabilità di corretta decisione condizionata alla trasmissione del segnale s_i . Osserviamo che gli addendi nell'ultima equazione della (6.31) sono tutti positivi in quanto probabilità, per cui minimizzare la probabilità di errore equivale a massimizzare i singoli addendi della sommatoria. Ciò si ottiene assegnando alla regione di decisione \mathcal{R}_i tutti i punti per cui risulta $P(s_i) f_{\mathbf{r}|s_i}(\mathbf{r}|s_i) \geq P(s_j) f_{\mathbf{r}|s_j}(\mathbf{r}|s_j)$, $j \neq i$. Poichè assegnare un punto ad una regione di decisione equivale a decidere per il corrispondente segnale trasmesso, si ottiene la seguente regola di decisione ottima

$$\text{decidi per } s_i \text{ se } P(s_i) f_{\mathbf{r}|s_i}(\mathbf{r}|s_i) \geq P(s_j) f_{\mathbf{r}|s_j}(\mathbf{r}|s_j), \quad j = 1, M, \quad j \neq i. \quad (6.32)$$

Nel caso in cui i segnali sono equiprobabili, risultando $P(s_i) = P(s_j) = \frac{1}{M}$ le probabilità a priori sono indipendenti dall'indice j rispetto al quale opera il criterio di

Figura Mancante

Figura 6.15: Regioni di decisione

massimizzazione per cui risulta

$$\text{decidi per } s_i \text{ se } f_{\mathbf{r}|s_i}(\mathbf{r}|s_i) \geq f_{\mathbf{r}|s_j}(\mathbf{r}|s_j), \quad j = 1, M, \quad j \neq i. \quad (6.33)$$

La regola (6.33) è nota come *regola di decisione ML* (Maximum Likelihood). Il decisore a massima verosimiglianza è impiegato dunque quando i simboli trasmessi sono equiprobabili o più in generale, anche se in modo sub-ottimo, quando le probabilità a priori non sono note.

Nel caso di rumore gaussiano, introducendo la caratterizzazione statistica del vettore \mathbf{r} , possiamo ulteriormente sviluppare la regola di decisione ottima. Infatti, sostituendo l'espressione della densità di probabilità congiunta (6.27) nella regola (6.33) ricaviamo che il decisore deve massimizzare rispetto ad i la quantità

$$f(\mathbf{r}|s_i) = \frac{1}{(\pi N_0)^{\frac{N}{2}}} \exp \left\{ -\frac{1}{N_0} \sum_{j=1}^N (r_j - s_{ij})^2 \right\}$$

Tralasciando i fattori che non dipendono da i ed osservando che una trasformazione monotona crescente non altera il punto in cui una funzione attinge il massimo otteniamo, dopo una trasformazione logaritmica, la quantità

$$\frac{1}{N_0} \sum_{j=1}^N (r_j - s_{ij})^2 = -D_i^2 \quad (6.34)$$

dove con D_i si è indicata la distanza euclidea tra il vettore \mathbf{r} ed il vettore rappresentativo del segnale s_i . L'equazione (6.34) evidenzia che la regola di decisione ottima consiste nello scegliere il segnale della costellazione che presenta una distanza euclidea minima dal segnale ricevuto. Se i segnali non sono equiprobabili le distanze vanno corrette con un termine additivo che dipende dalla probabilità a priori di trasmettere il segnale $s_i(t)$. L'immediata interpretazione geometrica del criterio ML (la cui applicabilità è più generale e quindi non limitata al solo caso di segnali gaussiani) dà anche il nome di *criterio di decisione a minima distanza* alla regola (6.34). La fig. ?? mostra graficamente come il criterio a minima distanza agisce nello spazio dei segnali. Relativamente al caso rappresentato in figura, il decisore, in presenza di segnali equiprobabili deciderà per il segnale s_4 che presenta la minima distanza euclidea dal segnale ricevuto \mathbf{r} . La struttura del decisore ottimo ML è mostrata in fig. ???. A partire dalle proiezioni $r_j, \quad j = 1, \dots, N$, vengono calcolate le statistiche $D_j, \quad j = 1, \dots, M$ e successivamente si decide per quel segnale che presenta la statistica di valore minimo. Una seconda struttura è ricavabile osservando che la statistica D_i può essere

semplificata come segue, risultando

$$D_i = \sum_{k=1}^N (r_k - s_{ik})^2 = \sum_{k=1}^N r_k^2 + \sum_{k=1}^N s_{ik}^2 - 2 \sum_{k=1}^N r_k s_{ik}$$

dove, poichè il primo termine non dipende da i si ottiene

$$\min_i D_i = \max_i \left\{ 2 \sum_{k=1}^N r_k s_{ik} - |\mathbf{s}_i|^2 \right\} \quad (6.35)$$

a cui corrisponde la struttura mostrata in fig. ???. Si noti che la quantità $\sum_{k=1}^N r_k s_{ik}$ non è altro che il prodotto scalare tra i segnali \mathbf{r} ed \mathbf{s}_i mentre la quantità $|\mathbf{s}_i|^2$ è l'energia del segnale \mathbf{s}_i .

6.4 SEGNALAZIONE BINARIA ANTIPODALE

La segnalazione binaria antipodale è una modulazione di ampiezza in cui i segnali trasmessi sono del tipo

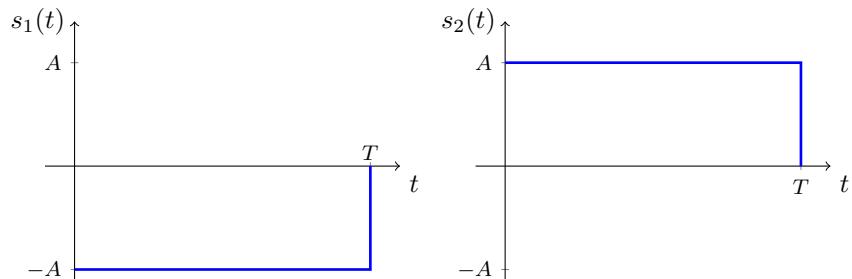
$$s_1(t) = g_T(t) \quad (6.36)$$

$$s_2(t) = -g_T(t) \quad t \in (0, T) \quad (6.37)$$

dove $g_T(t)$ è la forma d'onda generata dal modulatore avente energia

$$\mathcal{E}_g = \int_0^T |g_T(t)|^2 dt .$$

In figura 6.16 si riporta un esempio di segnali antipodali:



Utilizzando la procedura di ortogonalizzazione di Gram-Schmidt ricaviamo

$$\psi_1(t) = \frac{g_T(t)}{\sqrt{\mathcal{E}_g}}$$

$$d_2(t) = -g_T(t) - s_{21}\psi_1(t)$$

dove

$$s_{21} = \frac{1}{\sqrt{\mathcal{E}_g}} \int_0^T -g_T(t)g_T(t) dt = -\sqrt{\mathcal{E}_g}$$

per cui sostituendo nell' espressione di $d_2(t)$ otteniamo $d_2(t) = 0$. Lo spazio dei segnali è dunque monodimensionale con i segnali posizionati simmetricamente rispetto al punto di coordinata nulla e distanti tra loro $2\sqrt{\mathcal{E}_g}$ (cfr. fig.6.16).

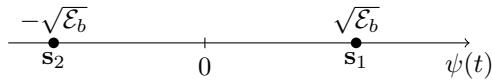


Figura 6.16: Segnalazione antipodale

Il proiettore presenta un solo ramo in cui esegue il prodotto scalare tra il segnale ricevuto e l'unico segnale di base ed è, dunque, formato da un'unica sezione la cui struttura è mostrata in fig. ??.

Al fine di individuare il criterio di decisione ottima, supponiamo di avere simboli equiprobabili e calcoliamo la distanza di \mathbf{r} da \mathbf{s}_1 e da \mathbf{s}_2 :

$$\begin{aligned} d(\mathbf{r}, \mathbf{s}_1) &= \sqrt{(\mathbf{r} - \sqrt{\mathcal{E}_g})^2} = |\mathbf{r} - \sqrt{\mathcal{E}_g}| \\ d(\mathbf{r}, \mathbf{s}_2) &= \sqrt{(\mathbf{r} + \sqrt{\mathcal{E}_g})^2} = |\mathbf{r} + \sqrt{\mathcal{E}_g}| \end{aligned}$$

confrontiamo, per semplicità le distanze al quadrato,

$$(\mathbf{r} - \sqrt{\mathcal{E}_g})^2 \underset{\mathbf{s}_1}{\gtrless} (\mathbf{r} + \sqrt{\mathcal{E}_g})^2 \quad (6.38)$$

Sviluppando i quadrati ed eliminando i termini comuni, si ottiene la regola di decisione

$$\mathbf{r} \underset{\mathbf{s}_2}{\gtrless} 0 \quad (6.39)$$

Da quanto detto se ne deduce che le regioni di decisione sono le due semirette alla destra dello zero (regione \mathcal{R}_1) ed alla sinistra dello zero (regione \mathcal{R}_2) in fig. ??.

Applicando direttamente il criterio a massima verosimiglianza, si ottiene In figura 6.17 sono rappresentate le densità di probabilità condizionate alla trasmissione dei segnali $s_1(t)$ ed $s_2(t)$. È utile notare come nella regione di decisione \mathcal{R}_1 risultino $f(r_1|s_1) > f(r_1|s_2)$ e

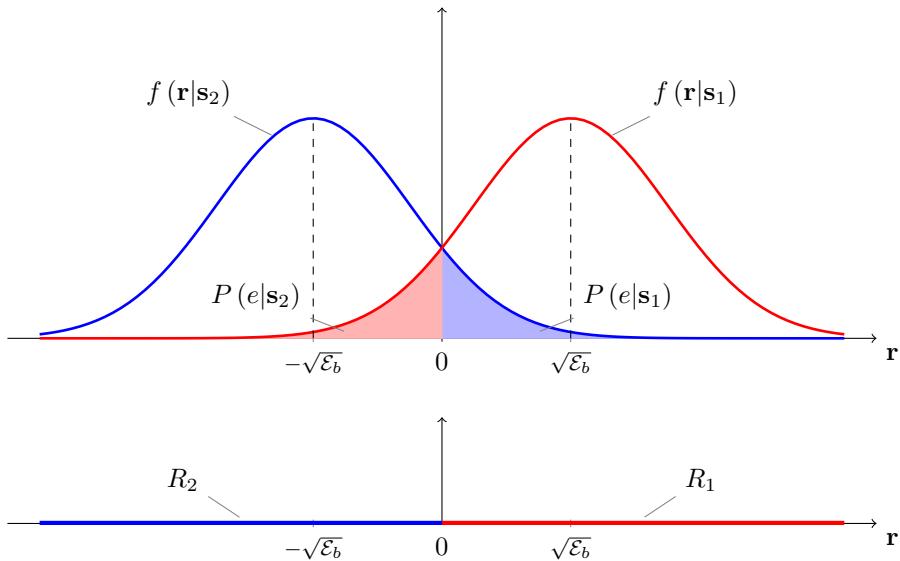


Figura 6.17: Grafico in alto: p.d.f. condizionali dei due segnali e probabilità di errore. Grafico in basso: regioni di decisione

viceversa nella regione di decisione \mathcal{R}_2 . Ovviamente, il punto in cui le due pdf condizionate assumono lo stesso valore coincide con l'origine che è anche la frontiera delle due regioni di decisione.

Per valutare la probabilità di errore si scomponete, di norma, la probabilità di errore totale nelle probabilità di errore condizionate alla trasmissione dei singoli segnali e si sommano poi queste probabilità con i rispettivi pesi dati dalle probabilità a priori; ciò in applicazione del teorema della probabilità totale. Per la segnalazione antipodale, si ottiene

$$P(e) = P(e|s_1)P(s_1) + P(e|s_2)P(s_2) \quad (6.40)$$

Ricordando che le pdf marginali delle proiezioni di $r(t)$ sui versori $\psi_k(t)$ sono gaussiane a media s_{ik} e varianza $N_0/2$ (sinteticamente $f(r_k|s_i) \sim \mathcal{N}(s_{ik}, \frac{N_0}{2})$), si ottiene, essendo $s_{11} = \sqrt{\mathcal{E}_g}$ ed $s_{21} = -\sqrt{\mathcal{E}_g}$

$$\begin{aligned} P(e|s_1) &= P(r_1 < 0|s_1) = P\left(\sqrt{\frac{N_0}{2}}X_0 + \sqrt{\mathcal{E}_g} < 0\right) \\ &= P\left(X_0 < -\sqrt{\frac{2\mathcal{E}_g}{N_0}}\right) = Q\left(\sqrt{\frac{2\mathcal{E}_g}{N_0}}\right) \end{aligned}$$

$$\begin{aligned} P(e|s_2) &= P(r_1 > 0|s_2) = P\left(\sqrt{\frac{N_0}{2}}X_0 - \sqrt{\mathcal{E}_g} > 0\right) \\ &= P\left(X_0 > \sqrt{\frac{2\mathcal{E}_g}{N_0}}\right) = Q\left(\sqrt{\frac{2\mathcal{E}_g}{N_0}}\right) \end{aligned}$$

La probabilitá di errore totale é dunque data dalla quantitá

$$P(e) = Q\left(\sqrt{\frac{2\mathcal{E}_g}{N_0}}\right) \quad (6.41)$$

riportata in Fig. (6.18).

Si osserva che la probabilitá di errore diminuisce all'aumentare del contrasto di energia \mathcal{E}_b/N_0 . É possibile riscrivere l'espressione della probabilitá di errore in termini della distanza euclidea d_{12} tra i due segnali, osservando che

$$d(s_1, s_2) = d_{12} = 2\sqrt{\mathcal{E}_g} = d_{12}$$

da cui si ottiene

$$P(e) = Q\left(\sqrt{\frac{2}{N_0}} \cdot \frac{d_{12}}{2}\right) = Q\left(\frac{d_{12}}{\sqrt{2N_0}}\right) \quad (6.42)$$

É utile evidenziare alcune importanti caratteristiche dell'eq. (6.41). Essendo la funzione $Q(x)$ una funzione decrescente, all'aumentare del rapporto $\frac{2\mathcal{E}_g}{N_0}$ le prestazioni del sistema migliorano. La precedente quantitá, posta nella forma $\frac{\mathcal{E}_g}{N_0/2}$ rappresenta il rapporto tra l'energia del segnale trasmesso e la PSD del rumore noto come *contrasto di energia*, da cui essenzialmente dipendono le prestazioni del sistema di modulazione.

6.5 SEGNALAZIONE PAM M-ARIA

Nella segnalazione PAM M-aria i segnali da trasmettere presentano un'ampiezza A_i appartenente ad un insieme di M valori,

$$s_i(t) = A_i g_T(t), \quad t \in [0, T] \quad (6.43)$$

dove:

$$A_i = 2i - 1 - M, \quad i = 1, \dots, M \quad (6.44)$$

Applicando la procedura di Gram-Schmidt, é facile verificare che la base necessaria alla rappresentazione degli M segnali é di tipo monodimensionale. Il versore della base é dato da

$$\psi(t) = \frac{g_T(t)}{\sqrt{\mathcal{E}_g}} \quad (6.45)$$

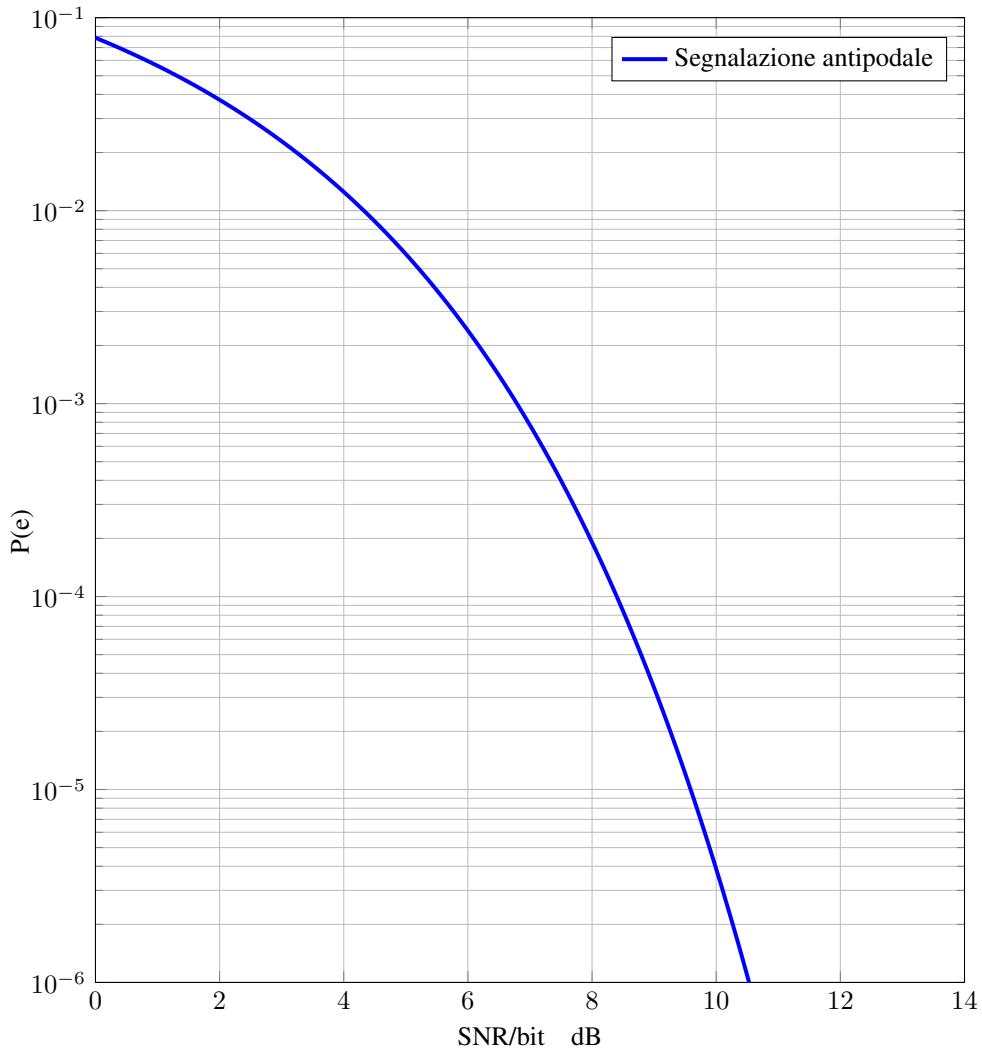
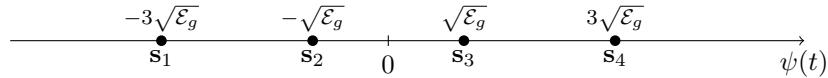


Figura 6.18: Probabilità di errore per la segnalazione binaria antipodale

per cui, i segnali possono essere espressi come:

$$s_m(t) = A_m \sqrt{\mathcal{E}_g} \psi(t) \quad (6.46)$$

Per semplicità di trattazione, ma senza perdere di generalità, consideriamo un 4-PAM, la cui costellazione è mostrata in figura ??.



I segnali hanno energia pari a $(2i - M - 1)^2 \mathcal{E}_g$. In tal caso si usa come parametro l'energia media definita come

$$\begin{aligned} \mathcal{E}_{av} &= \frac{1}{M} \sum_{i=1}^M \mathcal{E}_i = \frac{1}{M} \sum_{i=1}^M (2i - 1 - M)^2 \mathcal{E}_g \\ &= \frac{\mathcal{E}_g (M-1)M(M+1)}{3} = \frac{M^2 - 1}{3} \mathcal{E}_g \end{aligned} \quad (6.47)$$

Il ricevitore ottimo decide a minima distanza per cui le $M - 1$ soglie sono poste in corrispondenza dei punti $\tau_i = (2i - M)\sqrt{\mathcal{E}_g}$ $i = 1, \dots, M - 1$. Per $i = 1$ è facile verificare che la soglia $\tau_1 = -2\sqrt{\mathcal{E}_g}$, come riportato appunto in fig. ???. Lo schema del ricevitore ottimo è quindi quello di fig. ???.

La scelta delle regioni di decisione è fatta col criterio a minima distanza. Quindi stiamo dando per scontato che il canale è AWGN e i segnali sono equiprobabili.

Calcoliamo la Probabilità di errore.

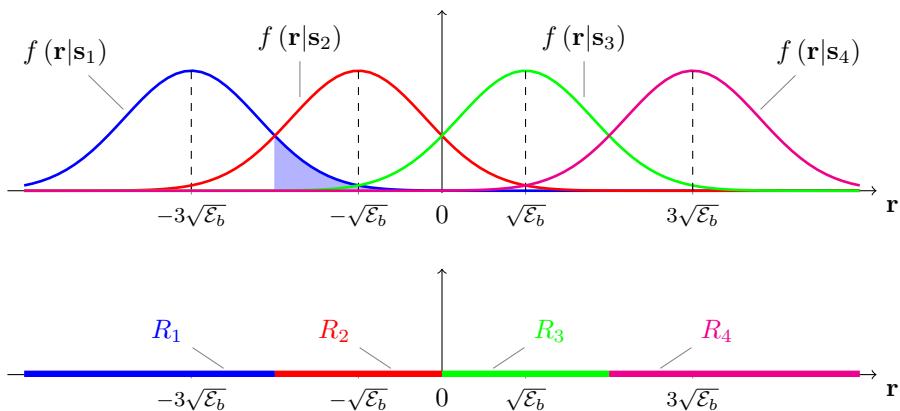


Figura 6.19: Grafico in alto: p.d.f. condizionali dei quattro segnali del 4-PAM. La regione in blu è la $P(e|s_1)$. Grafico in basso: regioni di decisione.

Per i segnali posti agli estremi della costellazione (s_1 e s_4 in questo caso) si commette errore se il rumore è negativo e maggiore di $\sqrt{\mathcal{E}_g}$ (segna s_1) oppure se il rumore è positivo

e maggiore di $\sqrt{\mathcal{E}_g}$ (segnalet \mathbf{s}_4), per cui

$$P(e|s_i) = P_r \left\{ n_1 > \sqrt{\mathcal{E}_g} \right\} = Q \left(\sqrt{\frac{2\mathcal{E}_g}{N_0}} \right) \quad \text{Segnali esterni} \quad i = 1, 4$$

Per i segnali interni (\mathbf{s}_2 e \mathbf{s}_3 in questo caso), si commette errore se il rumore è maggiore di $\sqrt{\mathcal{E}_g}$ o minore di $-\sqrt{\mathcal{E}_g}$ per cui si ottiene

$$P(e|s_i) = P_r \left\{ |n_1| > \sqrt{\mathcal{E}_g} \right\} = 2Q \left(\sqrt{\frac{2\mathcal{E}_g}{N_0}} \right) \quad \text{Segnali interni} \quad i = 2, 3$$

In definitiva la probabilità di errore, nell'ipotesi di segnali equiprobabili, sarà data da

$$\begin{aligned} P(e) &= \frac{1}{M} \sum_{i=1}^M P(e|s_i) \\ &= \frac{1}{M} \left[(M-2)2Q \left(\sqrt{\frac{2\mathcal{E}_g}{N_0}} \right) + 2Q \left(\sqrt{\frac{2\mathcal{E}_g}{N_0}} \right) \right] \\ &= \frac{2(M-1)}{M} Q \left(\sqrt{\frac{2\mathcal{E}_g}{N_0}} \right) \end{aligned} \quad (6.48)$$

6.5.1 Probabilità di errore in funzione del contrasto di energia

Per esprimere in maniera generale la $P(e)$ dobbiamo far riferimento al contrasto di energia. Essendo i segnali non equiergetici, dobbiamo far riferimento all'energia media. Questa dipende dal numero di segnali, e pertanto dobbiamo cercare di ricondurci all'energia media per bit:

$$\mathcal{E}_b = \frac{\mathcal{E}_s}{\log_2 M} \quad (6.49)$$

dove l'energia per segnale è in realtà una energia media, perché i segnali non sono equiergetici. Quindi:

$$\begin{aligned} \mathcal{E}_{avb} &= \sum_{m=1}^M \mathcal{E}_m P(\mathbf{s}_m) = \\ &= \frac{1}{M} \sum_{m=1}^M (2m-1-M)^2 \mathcal{E}_g = \\ &= \frac{1}{M} \mathcal{E}_g \frac{(M-1)M(M+1)}{3} = \\ &= \frac{M^2-1}{3} \mathcal{E}_g \end{aligned} \quad (6.50)$$

da cui

$$\mathcal{E}_g = \frac{3}{M^2 - 1} \mathcal{E}_{avb} = \frac{3}{M^2 - 1} (\log_2 M) (\mathcal{E}_b)_{av} \quad (6.51)$$

La $P(e)$ può allora essere riscritta come

$$\begin{aligned} P(e) &= 2 \frac{M-1}{M} Q\left(\sqrt{\frac{2\mathcal{E}_g}{N_0}}\right) = \\ &= 2 \frac{M-1}{M} Q\left(\sqrt{\frac{6 \log_2 M}{M^2 - 1} \cdot \frac{\mathcal{E}_b}{N_0}}\right) \end{aligned} \quad (6.52)$$

Quindi $P(e)$ dipende, oltre che dal contrasto di energia media per bit, anche dalla cardinalità del modulatore ($\log_2 M$).

Dobbiamo vedere come il parametro M influisce sulle prestazioni.

Supponiamo di fissare $(\mathcal{E}_b)_{av}/N_0$ e vediamo come varia la $P(e)$. All'aumentare di M , il termine $\frac{6 \log_2 M}{M^2 - 1}$ diminuisce perché al denominatore c'è una funzione quadratica di M che cresce più rapidamente del $\log_2 M$. Poiché la funzione $Q(\cdot)$ cresce al decrescere del suo argomento, la $P(e)$ aumenta.

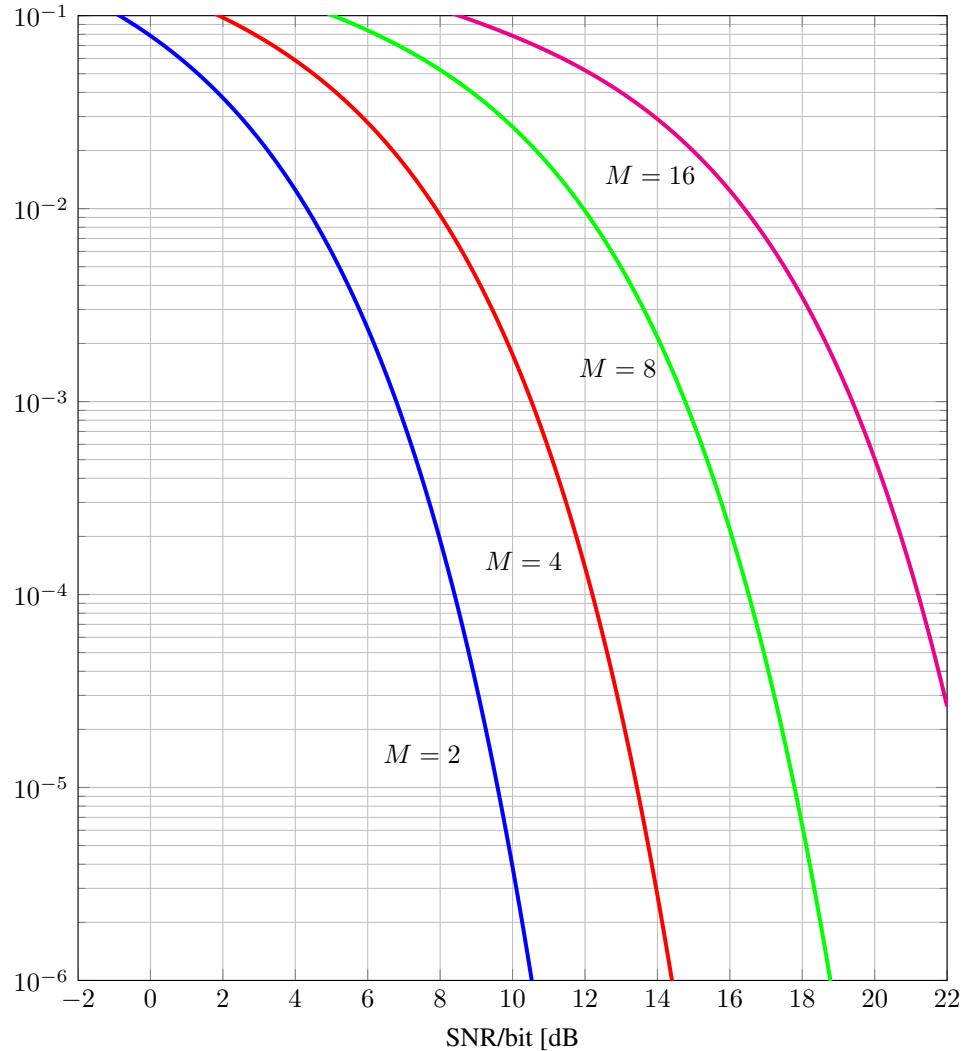
Per $M = 2$ otteniamo la segnalazione antipodale. D'altra parte, lasciando \mathcal{E}_b fissato, asintoticamente la $P(e)$ diventa:

$$\lim_{M \rightarrow +\infty} P(e) = 2 \cdot Q(0) = 1 \quad (6.53)$$

Dall'espressione di \mathcal{E}_g in funzione di \mathcal{E}_b osserviamo che se \mathcal{E}_b è costante all'aumentare di M allora \mathcal{E}_g diminuisce ed asintoticamente tende a zero. Pertanto, essendo \mathcal{E}_g rappresentativa della distanza, abbiamo una costellazione con segnali sempre più vicini.

Se invece vogliamo ragionare in termini di $P(e)$ costante dobbiamo mantenere costante l'argomento della Q ed osservare che se aumenta M il termine $\frac{6 \log_2 M}{M^2 - 1}$ diminuisce e quindi per mantenere il radicando costante dobbiamo aumentare l'altro termine, cioè l'energia media per bit.

In figura ?? si riporta la $P(e)$ per diversi valori di M .



Tale segnalazione non è efficiente in potenza perché all'aumentare di M dobbiamo aumentare l'energia per bit. Quindi questa segnalazione non va bene quando abbiamo dei vincoli sull'energia perché in quel caso non possiamo scegliere M grande a piacere.

Il coefficiente $\frac{6 \log_2 M}{M^2 - 1}$ è detto **coefficiente di efficienza in potenza** e si indica con γ . Questo coefficiente dà una misura dell'efficienza in potenza della segnalazione. Nel nostro caso γ diminuisce all'aumentare di M e quindi la segnalazione non è efficiente in potenza. Se invece γ aumentasse all'aumentare di M avremmo una segnalazione efficiente in potenza perché all'aumentare di M potremmo risparmiare in potenza.

6.5.1.1 Esempio

Vediamo come varia l'energia per bit tenendo fissata la $P(e)$ nel passaggio da M_2 a M_1 , con $M_2 > M_1$. Dobbiamo eguagliare gli argomenti delle funzioni Q nei due casi:

$$\frac{6 \log_2 M_2}{M_2^2 - 1} \left(\frac{\mathcal{E}_b}{N_0} \right)_{M_2} = \frac{6 \log_2 M_1}{M_1^2 - 1} \left(\frac{\mathcal{E}_b}{N_0} \right)_{M_1} \quad (6.54)$$

da cui

$$\left(\frac{\mathcal{E}_b}{N_0} \right)_{M_2} = \frac{6 \log_2 M_1}{6 \log_2 M_2} \cdot \frac{M_2^2 - 1}{M_1^2 - 1} \left(\frac{\mathcal{E}_b}{N_0} \right)_{M_1} \quad (6.55)$$

Il coefficiente $\frac{6 \log_2 M_1}{6 \log_2 M_2} \cdot \frac{M_2^2 - 1}{M_1^2 - 1}$ ci dà la misura di quanto dobbiamo aumentare il contrasto di energia per mantenere costante la probabilità di errore.

Per riportarci al grafico ?? dobbiamo calcolare l'incremento in dB, e quindi:

$$10 \log_{10} \left(\frac{6 \log_2 M_1}{6 \log_2 M_2} \cdot \frac{M_2^2 - 1}{M_1^2 - 1} \right) \quad (6.56)$$

Supponendo, per esempio, $M_1 = 2$ e $M_2 = 4$, si ha:

$$10 \log_{10} \left(\frac{1}{2} \cdot \frac{15}{3} \right) = 10 \log_{10} \left(\frac{5}{2} \right) \cong 4 \text{dB} \quad (6.57)$$

Questo incremento non è sempre lo stesso tra un passaggio e l'altro ma va calcolato di volta in volta. Il passaggio da un M all'altro è tipicamente un raddoppio perché dobbiamo sempre considerare M potenza di 2, cioè passiamo da $M_1 = 2^k$ ad $M_2 = 2^{k+1}$.

Particularizzando l'espressione di prima a questo tipo di passaggio si ha:

$$\frac{\log_2 2^k}{\log_2 2^{k+1}} \cdot \frac{2^{2(k+1)} - 1}{2^{2k} - 1} = \frac{k}{k+1} \cdot \frac{4 \cdot 4^k - 1}{4^k - 1} \quad (6.58)$$

Asintoticamente, per $M \rightarrow \infty$ e quindi $k \rightarrow \infty$ si ottiene 4, che è pari a circa 6 dB.

6.5.2 Probabilità di errore in funzione della distanza

Abbiamo visto che la $P(e)$ espressa in funzione di $(\mathcal{E}_b)_{av}$ e della distanza è:

$$P(e) = 2 \frac{M-1}{M} Q \left(\sqrt{\frac{6 \log_2 M}{M^2 - 1} \cdot \frac{(\mathcal{E}_b)_{av}}{N_0}} \right) = 2 \frac{M-1}{M} Q \left(\frac{d_{min}}{\sqrt{2 N_0}} \right) \quad (6.59)$$

La $P(e)$ in funzione della distanza può essere calcolata anche per una qualunque segnalazione binaria e otteniamo un valore preciso. In generale per qualunque segnalazione si può trovare

la $P(e)$ in funzione della di d_{min} , ma dobbiamo fare una approssimazione o meglio una massimizzazione, cioè troviamo un limite superiore (upper bound) della $P(e)$ in funzione della distanza.

Per calcolarlo usiamo la union bound:

$$P\left(\bigcup_{i=1}^M A_i\right) \leq \sum_{i=1}^M P(A_i) \quad (6.60)$$

Questa relazione vale per eventi A_i non mutuamente esclusivi.

La probabilità di errore condizionata è:

$$\begin{aligned} P(e|\mathbf{s}_m) &= P(\{\mathbf{r} \notin R_m\}|\mathbf{s}_m) = \\ &= P\left(\bigcup_{i \neq m} \{\mathbf{r} \in R_i\}|\mathbf{s}_m\right) = \\ &= P\left(\bigcup_{i \neq m} \{d(\mathbf{r}, \mathbf{s}_i) < d(\mathbf{r}, \mathbf{s}_m)\}|\mathbf{s}_m\right) \leq \\ &\leq \sum_{i \neq m} P(\{d(\mathbf{r}, \mathbf{s}_i) < d(\mathbf{r}, \mathbf{s}_m)\}|\mathbf{s}_m) \leq \\ &\leq \sum_{i \neq m} Q\left(\frac{d_{im}}{\sqrt{2N_0}}\right) \end{aligned} \quad (6.61)$$

Quindi la $P(e)$ totale sarà:

$$\begin{aligned} P(e) &= P(e|\mathbf{s}_m)P(\mathbf{s}_m) \leq \\ &\leq \sum_{m=1}^M \sum_{i \neq m} Q\left(\frac{d_{im}}{\sqrt{2N_0}}\right) \end{aligned} \quad (6.62)$$

Questo risultato è già importante perché ci consente di trovare l'upper bound semplicemente calcolando le distanze d_{im} tra i segnali della costellazione.

Per semplificare ancora di più il calcolo possiamo far riferimento alla distanza minima

$$d_{min} = \min_{i,m} d_{im} \quad (6.63)$$

ed essendo Q decrescente otteniamo

$$\begin{aligned}
 P(e) &\leq P(\mathbf{s}_m) \sum_{m=1}^M \sum_{i \neq m} Q\left(\frac{d_{im}}{\sqrt{2N_0}}\right) \leq \\
 &\leq \sum_{m=1}^M P(\mathbf{s}_m) \sum_{i \neq m} Q\left(\frac{d_{im}}{\sqrt{2N_0}}\right) = \\
 &= (M-1)Q\left(\frac{d_{min}}{\sqrt{2N_0}}\right) \sum_{m=1}^M P(\mathbf{s}_m) = \\
 &= (M-1)Q\left(\frac{d_{min}}{\sqrt{2N_0}}\right)
 \end{aligned} \tag{6.64}$$

In conclusione abbiamo che l'espressione dell'upper bound è:

$$P(e) \leq (M-1)Q\left(\frac{d_{min}}{\sqrt{2N_0}}\right) \tag{6.65}$$

A prescindere dal tipo di segnalazione la $P(e)$ dipende dalla cardinalità del modulatore ($M - 1$) e dalla particolare costellazione (d_{min}). Questa $P(e)$ è uguale a quella di una segnalazione binaria con segnali a distanza d_{min} moltiplicata per $(M - 1)$, che è il numero di segnali diversi da quello trasmesso. È certamente il caso peggiore, perché stiamo dicendo che abbiamo $M - 1$ segnali e tutti a distanza minima e quindi non teniamo conto del fatto che ci sono dei segnali più lontani che danno minore contributo alla $P(e)$. Sommiamo tutte le $P(e)$ delle varie coppie di segnali considerati come se fossero tutti alla minima distanza.

Possiamo "ritoccare" in qualche modo questo limite per renderlo più realistico: al posto di $(M - 1)$ sostituiamo un parametro $(\bar{\nu})$, che rappresenta il numero medio di segnali a distanza minima (a seconda della costellazione e del particolare segnale nella costellazione il numero di segnali a distanza minima cambia e quindi consideriamo il numero medio).

Quindi otteniamo:

$$P(e) \leq \bar{\nu}Q\left(\frac{d_{min}}{\sqrt{2N_0}}\right) \tag{6.66}$$

Per valutare la $P(e)$ dobbiamo pertanto valutare due parametri: $\bar{\nu}$ e d_{min} . Il calcolo di $\bar{\nu}$ tipicamente non è difficile.

6.5.2.1 Esempio: PAM

Nel caso PAM i segnali a distanza minima possono essere due (quando il segnale è interno) oppure uno (quando il segnale è esterno). Quindi, pesando per le rispettive probabilità otteniamo:

$$\bar{\nu} = 1 \cdot \frac{1}{M} + 2 \cdot \frac{M-2}{M} = \frac{2}{M}(M-1) \tag{6.67}$$

che è esattamente lo stesso coefficiente della $P(e)$ nell'equazione (??) che abbiamo calcolato nel caso esatto.

d_{min} è la distanza tra due segnali adiacenti, che vale $2\sqrt{\mathcal{E}_g}$. Esprimendo \mathcal{E}_g in funzione di \mathcal{E}_{avb} . Abbiamo già trovato che $\mathcal{E}_{avb} = \frac{M^2-1}{3}\mathcal{E}_g$ quindi ricaviamo \mathcal{E}_g da questa formula e sostituendo si ha:

$$\begin{aligned} d_{min} &= 2\sqrt{\mathcal{E}_g} = \\ &= 2\sqrt{\frac{3}{M^2-1} \cdot \mathcal{E}_{avb}} = \\ &= 2\sqrt{\frac{3 \log_2 M \cdot (\mathcal{E}_b)_{av}}{M^2-1}} \end{aligned} \quad (6.68)$$

A questo punto, sostituendo $\bar{\nu}$ e d_{min} nell'upper bound otteniamo

$$P(e) \leq 2 \frac{M-1}{M} \cdot Q \left(\sqrt{\frac{6 \log_2 M}{M^2-1} \cdot \frac{(\mathcal{E}_b)_{av}}{N_0}} \right) \quad (6.69)$$

che è la stessa espressione che avevamo trovato con il calcolo esatto. Nel caso limite vale l'uguaglianza.

Osserviamo che nel caso PAM otteniamo proprio l'espressione esatta della $P(e)$. Infatti, nel valutare di volta in volta le coppie (quindi nel fare il confronto binario) si valutano effettivamente le probabilità di errore totali. Non otteniamo invece il valore esatto della $P(e)$ quando le regioni di errore si sovrappongono, perché in quel caso, facendo la somma di tutti i casi possibili, si ottengono delle regioni che vengono considerate più volte. questo nel caso PAM non accade mai perché le regioni di errore non si sovrappongono mai e quindi otteniamo il valore esatto.

Infine osserviamo che nel caso PAM avevamo già calcolato il valore esatto della $P(e)$ e quindi il bound è servito solo per confronto e verifica, ma in alcuni casi è più difficile il calcolo esatto della probabilità di errore ed è effettivamente utile valutare il bound.

6.6 PHASE SHIFT KEYING (PSK)

Per questo tipo di segnalazione l'informazione è associata alla fase del segnale:

$$s_m(t) = g_T(t) \cos \left[2\pi f_0 t + (m-1) \frac{2\pi}{M} \right] \quad (6.70)$$

A questo punto proviamo a separare il contributo che dipende dal tempo e quello che dipende da m :

$$\begin{aligned} s_m(t) &= g_T(t)\sqrt{2}\cos(2\pi f_0 t)\cos\left[(m-1)\frac{2\pi}{M}\right] + \\ &\quad -g_T(t)\sqrt{2}\sin(2\pi f_0 t)\sin\left[(m-1)\frac{2\pi}{M}\right] \end{aligned} \quad (6.71)$$

Abbiamo così espresso $s_m(t)$ come combinazione lineare di una funzione seno e una funzione coseno.

A questo punto dovremmo applicare la procedura di ortonormalizzazione di Gram-Schmidt per trovare i segnali di base. Osserviamo però che il generico segnale $s_m(t)$ è dato dalla somma di due termini, di cui uno è il prodotto tra un coefficiente dipendente da m e una funzione coseno e l'altro è il prodotto tra un coefficiente dipendente da m e una funzione seno. Pertanto, essendo le funzioni seno e coseno tra loro ortogonali, come segnali di base possiamo scegliere proprio $g_T(t)\cos(2\pi f_0 t)$ e $g_T(t)\sin(2\pi f_0 t)$, opportunamente normalizzati:

$$\psi_1(t) = \frac{\sqrt{2}g_T(t)\cos(2\pi f_0 t)}{\sqrt{\mathcal{E}_g}} \quad (6.72)$$

$$\psi_2(t) = \frac{-\sqrt{2}g_T(t)\sin(2\pi f_0 t)}{\sqrt{\mathcal{E}_g}} \quad (6.73)$$

I segnali del modulatore possono allora essere espressi come

$$s_m(t) = \sqrt{\mathcal{E}_g}\cos\left[(m-1)\frac{2\pi}{M}\right]\psi_1(t) + \sqrt{\mathcal{E}_g}\sin\left[(m-1)\frac{2\pi}{M}\right]\psi_2(t) \quad (6.74)$$

e il corrispondente vettore rappresentativo nello spazio dei segnali, che ha due componenti, è

$$\mathbf{s}_m = \left(\sqrt{\mathcal{E}_g}\cos\left[(m-1)\frac{2\pi}{M}\right], \sqrt{\mathcal{E}_g}\sin\left[(m-1)\frac{2\pi}{M}\right] \right) \quad (6.75)$$

Per calcolare l'energia possiamo sfruttare il vettore appena determinato. L'energia è infatti il prodotto scalare del vettore per se stesso, cioè:

$$\begin{aligned} \mathcal{E}_m &= \mathcal{E}_g \cos^2\left[(m-1)\frac{2\pi}{M}\right] + \mathcal{E}_g \sin^2\left[(m-1)\frac{2\pi}{M}\right] = \\ &= \mathcal{E}_g = \\ &= \mathcal{E}_s \end{aligned} \quad (6.76)$$

Come si può vedere, l'energia non dipende dal particolare m ma è la stessa per tutti i segnali, che sono quindi equienergetici di energia \mathcal{E}_g .

Il generico vettore può allora essere riscritto come

$$\mathbf{s}_m = \left(\sqrt{\mathcal{E}_s} \cos \left[(m-1) \frac{2\pi}{M} \right], \sqrt{\mathcal{E}_s} \sin \left[(m-1) \frac{2\pi}{M} \right] \right) \quad (6.77)$$

Quindi tutti i punti della costellazione stanno su una circonferenza di raggio $\sqrt{\mathcal{E}_s}$ e sono “distanziati” di $\frac{2\pi}{M}$. Per esempio, nel caso dell’8-PSK la costellazione sarà quella mostrata in figura 6.20a

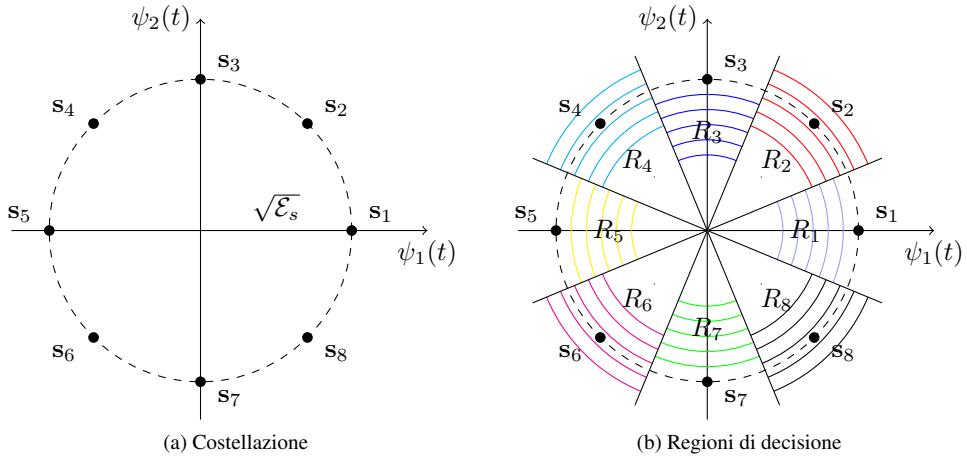


Figura 6.20: Costellazione 8-PSK

6.6.1 Calcolo della probabilità di errore

Per calcolare la $P(e)$ dobbiamo individuare la suddivisione in regioni di decisione. Dal momento che stiamo considerando segnali equiprobabili e canale AWGN, vale la regola a minima distanza.

Consideriamo, ad esempio, la costellazione 8-PSK in figura 6.20b. Per la geometria del problema, le regioni di decisione sono dei settori angolari. Calcolare la $P(e)$ in maniera esatta non è semplice perché dovremmo calcolare un integrale esteso al settore angolare. Sfruttiamo allora il bound, che è molto più semplice:

$$P(e) \leq \bar{\nu} Q \left(\frac{d_{min}}{\sqrt{2N_0}} \right) \quad (6.78)$$

Nel nostro caso $\bar{\nu} = 2$, perché i segnali a distanza minima sono sempre due. Calcoliamo invece d_{min} rifacendoci alla figura 6.21.

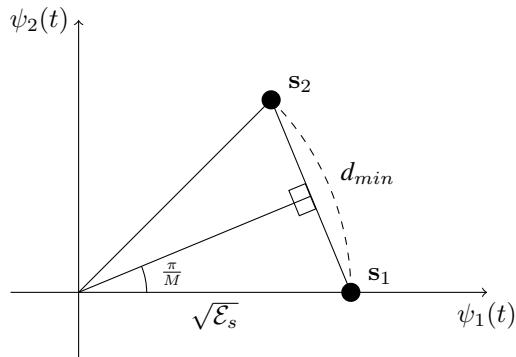


Figura 6.21: Costellazione 8-PSK. Calcolo della distanza minima tra due segnali

La distanza minima è la corda che congiunge s_1 e s_2 , che è data da:

$$\begin{aligned} d_{min} &= 2\sqrt{\mathcal{E}_s} \sin \frac{\pi}{M} = \\ &= 2\sqrt{\log_2 M \cdot \mathcal{E}_b} \sin \frac{\pi}{M} \end{aligned} \quad (6.79)$$

Quindi la $P(e)$ può essere riscritta come:

$$P(e) = 2Q\left(\sqrt{2\log_2 M \cdot \sin^2\left(\frac{\pi}{M}\right) \cdot \frac{\mathcal{E}_b}{N_0}}\right) \quad (6.80)$$

Per capire il comportamento di questa segnalazione dal punto di vista dell'efficienza in potenza consideriamo il coefficiente di efficienza in potenza γ_M , così definito:

$$\gamma_M = 2\log_2 M \cdot \sin^2\left(\frac{\pi}{M}\right) \quad (6.81)$$

La funzione sin non è una funzione monotona ma osserviamo che all'aumentare di M l'angolo $\frac{\pi}{M}$ diventa sempre più piccolo e quindi possiamo approssimare il valore della funzione con il suo argomento, cioè:

$$\gamma_M \approx 2\log_2 M \cdot \left(\frac{\pi}{M}\right)^2 \quad (6.82)$$

A questo punto è evidente che γ_M diminuisce all'aumentare di M , perché al numeratore abbiamo la funzione $\log_2 M$ che cresce più lentamente di M^2 che compare al denominatore. Di conseguenza la funzione Q aumenta.

Quindi la segnalazione PSK **non è efficiente in potenza**, come si può vedere dalla figura 6.22.

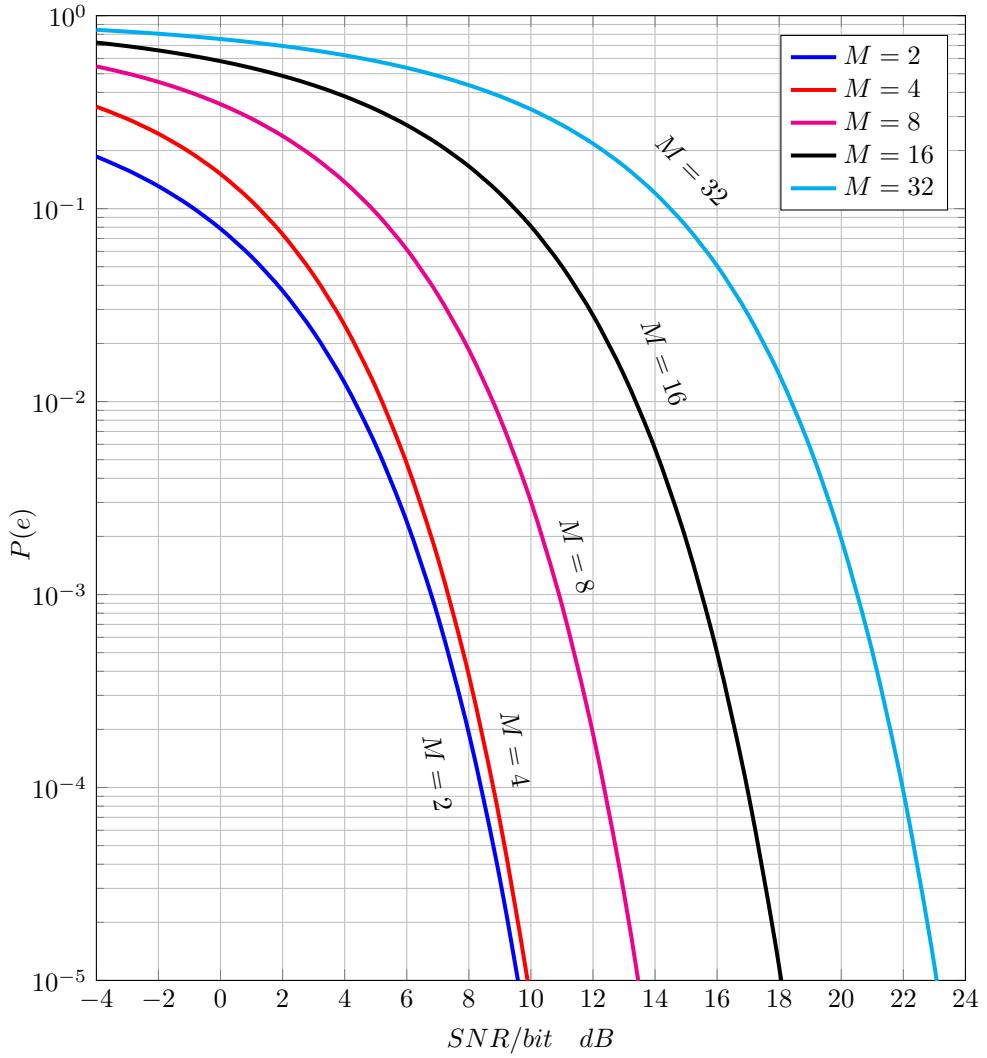


Figura 6.22: Probabilità di errore per simbolo per segnali PSK

6.7 SEGNALAZIONE ORTOGONALE

In questa sezione consideriamo la segnalazione ortogonale. I segnali del modulatore sono:

$$s_m(t) = \sqrt{\mathcal{E}_s} \psi_m(t), \quad m = 1, 2, \dots, M \quad (6.83)$$

e sono ortogonali ed equienergetici.

La dimensionalità dello spazio dei segnali è la massima possibile ($L = M$) e la costellazione è rappresentata da punti disposti ognuno su un asse e tutti a distanza $\sqrt{\mathcal{E}_s}$ dall'origine.

Lo studio dello spazio dei segnali è utile nella progettazione del proiettore, che avrà tanti rami quanti sono i segnali del modulatore, essendo questi segnali tutti ortogonali, come mostrato in figura 6.23.

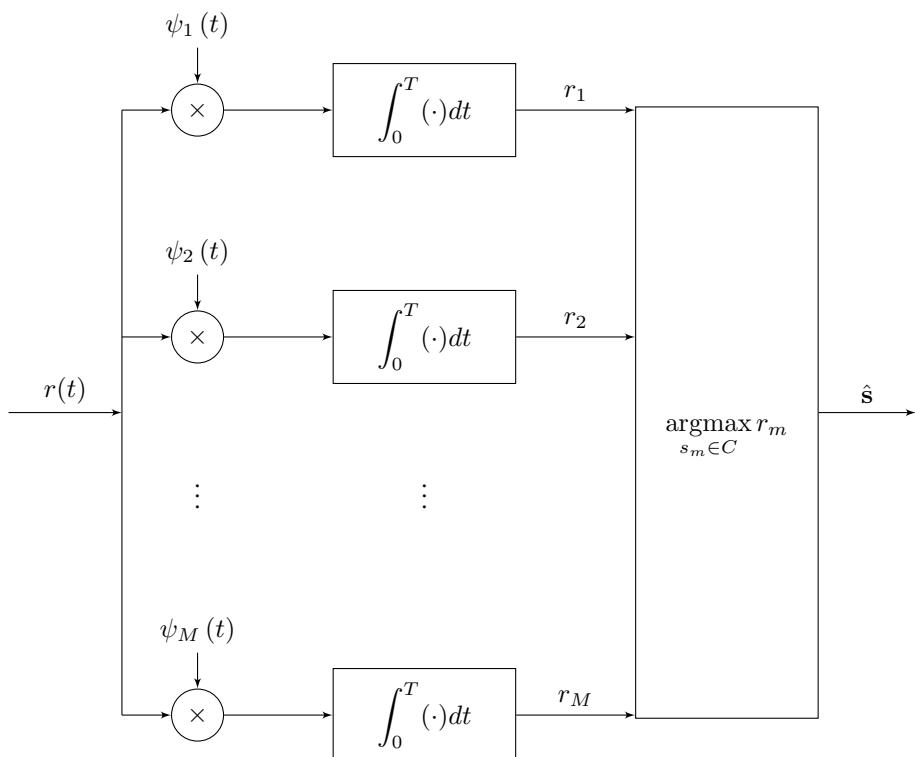


Figura 6.23: Ricevitore per segnalazione ortogonale

Dal proiettore avremo il vettore \mathbf{r} che costituisce la statistica sufficiente per stimare il segnale trasmesso.

A questo punto dobbiamo scegliere il criterio ottimo di decisione. In questo caso il criterio ottimo (quello cioè che garantisce la minima probabilità di errore) è quello a massima correlazione, perché i segnali sono equiprobabili, il canale è AWGN ed hanno tutta la stessa energia. È conveniente il criterio a massima correlazione per segnali ortogonali, cioè:

$$\hat{\mathbf{s}} = \underset{s_m \in C}{\operatorname{argmax}} \langle \mathbf{r}, \mathbf{s}_m \rangle \quad (6.84)$$

Nella segnalazione ortogonale ogni vettore rappresentativo dei segnali ha un'unica componente non nulla ed il suo valore è pari al coefficiente moltiplicativo del giusto versore nella combinazione lineare che esprime il segnale:

$$\mathbf{s}_m = (0, 0, \dots, \sqrt{\mathcal{E}_s}, 0, \dots, 0) \quad (6.85)$$

Il vettore \mathbf{r} , invece, nel caso generale ha tutte le componenti:

$$\mathbf{r}_m = (r_1, r_2, \dots, r_M) \quad (6.86)$$

Tuttavia nel prodotto scalare avremo la sommatoria dei prodotti delle componenti corrispondenti del segnale trasmesso e di quello ricevuto e quindi solo un termine della sommatoria sarà diverso da zero:

$$\begin{aligned} \hat{\mathbf{s}} &= \underset{s_m \in C}{\operatorname{argmax}} \langle \mathbf{r}, \mathbf{s}_m \rangle = \\ &= \underset{s_m \in C}{\operatorname{argmax}} r_m \sqrt{\mathcal{E}_s} = \\ &= \underset{s_m \in C}{\operatorname{argmax}} r_m \end{aligned}$$

essendo $\sqrt{\mathcal{E}_s}$ una costante.

Quindi dobbiamo semplicemente andare a valutare l'indice m per cui la componente di \mathbf{r} è massima, cioè la componente massima di \mathbf{r} . Quindi possiamo completare lo schema del ricevitore col decisore che sceglie $\hat{\mathbf{s}} = \mathbf{s}_m$ tale che r_m è la massima. Questa regola di decisione si spiega anche intuitivamente.

Ora andiamo a valutare il vettore \mathbf{r} dato che è stato trasmesso un particolare segnale, ad esempio \mathbf{s}_1 :

$$\begin{aligned} \mathbf{r} | \mathbf{s}_1 &= \mathbf{s}_1 + \mathbf{n} = \\ &= (\sqrt{\mathcal{E}_s} + n_1, n_2, \dots, n_M) \end{aligned}$$

Le componenti di rumore sono gaussiane:

$$n_i \sim \mathcal{N}\left(0, \frac{N_0}{2}\right) \quad (6.87)$$

Intuitivamente sceglieremo per il segnale corrispondente alla componente massima proprio perché immaginiamo che la componente massima è quella in cui c'è $\sqrt{\mathcal{E}_s}$ sommata al rumore, mentre le altre componenti sono costituite solo da rumore.

Possiamo inoltre calcolare in maniera esatta la $P(e)$, anche se il calcolo sarà complicato. Denotiamo con $P(c)$ la probabilità di corretta decisione, che sarà pari a

$$P(c) = 1 - P(e) \quad (6.88)$$

e consideriamo la legge della probabilità totale:

$$P(c) = \sum_{m=1}^M P(c|\mathbf{s}_m)P(\mathbf{s}_m) \quad (6.89)$$

Calcoliamo ora le probabilità condizionate. Consideriamo

$$P(c|\mathbf{s}_1) = P(\{r_2 < r_1\}, \{r_3 < r_1\}, \dots, \{r_M < r_1\} | \mathbf{s}_1) \quad (6.90)$$

cioè prendiamo la corretta decisione quando tutte le componenti di \mathbf{r} sono minori della componente r_1 . A questo punto si pone un problema, perché tutti i vari eventi che abbiamo considerato non sono indipendenti in quanto è vero che gli r_i sono indipendenti (perché sono solo componenti di rumore che essendo gaussiano bianco è incorrelato) ma sono tutti confrontati con r_1 , che è lo stesso per tutti gli eventi che quindi non sono indipendenti.

Gli eventi sarebbero indipendenti se r_1 fosse un numero, perché in quel caso avremmo, in ogni evento, la dipendenza solo da r_i .

Per fare questo dobbiamo introdurre un ulteriore condizionamento ad r_1 , cioè dobbiamo considerare r_1 fissato, pari ad un valore costante, e poi dobbiamo mediare su tutti i possibili valori del condizionamento. È una sorta di probabilità totale, cioè per passare dalla condizionata a quella generica dobbiamo mediare su tutti i possibili valori del condizionamento. Abbiamo così:

$$P(c|\mathbf{s}_1) = \mathbb{E}[P(\{r_2 < r_1\}, \{r_3 < r_1\}, \dots, \{r_M < r_1\} | \mathbf{s}_1, r_1)] \quad (6.91)$$

Calcoliamo ora la probabilità dei vari eventi congiunti che, dato r_1 , si fattorizza nel prodotto delle singole probabilità perché gli eventi sono statisticamente indipendenti:

$$P(\{r_2 < r_1\}, \{r_3 < r_1\}, \dots, \{r_M < r_1\} | \mathbf{s}_1, r_1) = \prod_{i=2}^M P(\{r_i < r_1\} | \mathbf{s}_1, r_1) \quad (6.92)$$

A questo punto osserviamo che le r_i sono componenti di rumore e quindi sono gaussiane e quindi la probabilità che r_i sia minore di r_1 è la probabilità che una gaussiana sia minore di un certo valore. Pertanto abbiamo

$$P(c|\mathbf{s}_1) = \mathbb{E}_{r_1} \left[\left(1 - Q \left(r_1 \sqrt{\frac{2}{N_0}} \right) \right)^{M-1} \right] \quad (6.93)$$

che, sfruttando il teorema fondamentale per il calcolo della media, diventa

$$P(c|\mathbf{s}_1) = \int_{-\infty}^{+\infty} \left(1 - Q \left(\sqrt{\frac{2}{N_0}} x \right) \right)^{M-1} f_{r_1|\mathbf{s}_1}(x|\mathbf{s}_1) dx \quad (6.94)$$

La pdf di $r_1|s_1$ la conosciamo, infatti

$$r_1|s_1 = \sqrt{\mathcal{E}_s} + n_1 \sim \mathcal{N}\left(\sqrt{\mathcal{E}_s}, \frac{N_0}{2}\right) \quad (6.95)$$

Quindi:

$$P(c|s_1) = \int_{-\infty}^{+\infty} \left(1 - Q\left(\sqrt{\frac{2}{N_0}}x\right)\right)^{M-1} \frac{1}{\sqrt{\pi N_0}} e^{-\frac{1}{N_0}(x-\sqrt{\log_2 M \cdot \mathcal{E}_b})^2} dx \quad (6.96)$$

Osserviamo che questa espressione di $P(c|s_1)$ non dipende da s_1 . Se avessimo fatto lo stesso calcolo per un altro segnale avremmo ottenuto la stessa cosa e quindi quella è proprio la probabilità di corretta decisione:

$$P(c) = \int_{-\infty}^{+\infty} \left(1 - Q\left(\sqrt{\frac{2}{N_0}}x\right)\right)^{M-1} \frac{1}{\sqrt{\pi N_0}} e^{-\frac{1}{N_0}(x-\sqrt{\log_2 M \cdot \mathcal{E}_b})^2} dx \quad (6.97)$$

L'espressione esatta di $P(c)$ è molto complicata. L'integrale non si risolve in forma chiusa e si può calcolare solo numericamente.

Il calcolo numerico ci consente di ricavare l'andamento di $P(e)$, che sarà del tipo mostrato in figura 6.24.

Osserviamo che, fissato un valore di SNR, all'aumentare di M la $P(e)$ diminuisce. Se fissiamo la $P(e)$, all'aumentare di M il contrasto di energia richiesto diminuisce. Quindi **la segnalazione ortogonale è efficiente in potenza**.

Inoltre, per valori piccoli di SNR le curve si incrociano, mentre nel caso PAM erano quasi "parallele". Quando le curve si incrociano si ribalta il comportamento, cioè non abbiamo più efficienza in potenza. Bisogna valutare un limite di SNR al di sopra del quale c'è efficienza in potenza. Tuttavia ciò accade per valori elevati della $P(e)$ e quindi sono effetti trascurabili perché non siamo interessati a quelle $P(e)$.

Nel grafico, sull'asse delle ordinate c'è la probabilità di errore per bit e non per simbolo. Le due probabilità d'errore sono però legate dalla seguente relazione:

$$P_b = \frac{2^{k-1}}{2^k - 1} P_M \approx \frac{P_M}{2} \quad (6.98)$$

dove con P_b denotiamo la probabilità di errore per bit e con P_M la probabilità di errore per simbolo. L'approssimazione vale per $k \gg 1$.

Cerchiamo di valutare la soglia di SNR al di sotto della quale perdiamo l'efficienza in potenza e lo facciamo usando il bound:

$$P(e) \leq \bar{\nu} Q\left(\frac{d_{min}}{\sqrt{2N_0}}\right) \quad (6.99)$$

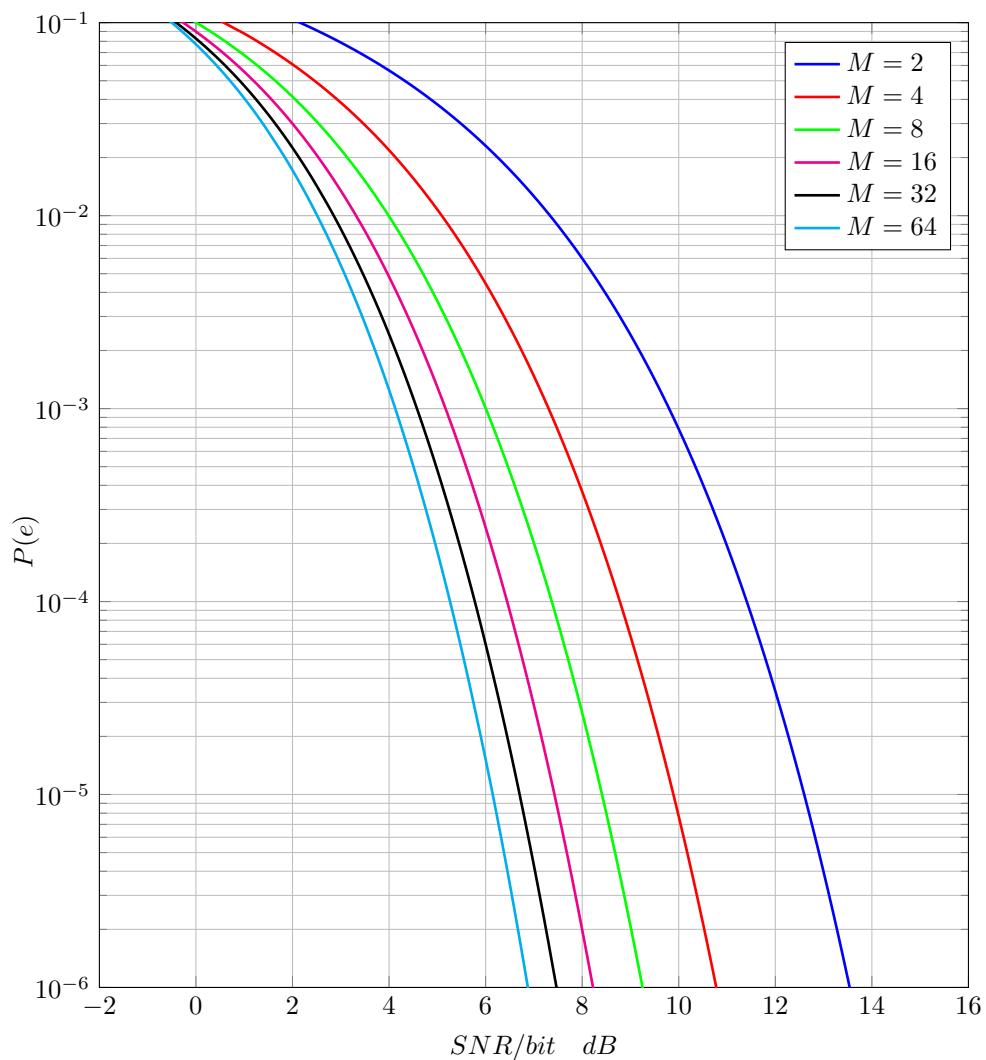


Figura 6.24: Probabilità di errore per rivelazione coerente di segnali ortogonali

Dobbiamo valutare i parametri $\bar{\nu}$ e d_{min} . Nella segnalazione ortogonale tutti i segnali sono alla stessa distanza, quindi per ogni segnale ci sono $M - 1$ segnali a distanza minima. Quindi

$$\bar{\nu} = M - 1 \quad (6.100)$$

La d_{min} è la distanza tra due segnali qualunque (essendo tutti alla stessa distanza):

$$d_{min} = \sqrt{2\mathcal{E}_s} = \sqrt{2\log_2 M \cdot \mathcal{E}_b} \quad (6.101)$$

Sostituendo i valori trovati, abbiamo che il bound si può esprimere come:

$$P(e) \leq (M - 1) Q \left(\sqrt{\log_2 M \cdot \frac{\mathcal{E}_b}{N_0}} \right) \quad (6.102)$$

Il calcolo approssimato è molto più semplice di quello esatto, ma nel caso ortogonale otteniamo effettivamente un valore che rappresenta un limite superiore della $P(e)$ e non otteniamo il valore esatto, come invece accadeva nella segnalazione PAM, perché in questo caso le regioni di errore si sovrappongono.

In figura 6.25 si riporta sia la $P(e)$ esatta che il relativo bound.

Per valutare l'efficienza in potenza analizziamo l'argomento della funzione Q . Per mantenere fissa la $P(e)$ deve rimanere fisso l'argomento della Q ed osserviamo che all'aumentare di M il coefficiente moltiplicativo del contrasto di energia ($\log_2 M$) aumenta, e quindi $\frac{\mathcal{E}_b}{N_0}$ deve diminuire. **Si ha allora efficienza in potenza.**

Questo discorso viene fatto trascurando il coefficiente $(M - 1)$. Tuttavia dobbiamo osservare che il coefficiente moltiplicativo di Q non è completamente trascurabile, perché è linearmente dipendente da M e diverge al crescere di M . In realtà, l'argomento della Q , se è abbastanza grande, prevale, altrimenti prevale il coefficiente $(M - 1)$.

Osserviamo inoltre che se i valori di $\frac{\mathcal{E}_b}{N_0}$ diventano troppo piccoli non possiamo più trascurare il coefficiente $(M - 1)$ perché l'argomento della Q è troppo piccolo e quindi la decrescita della Q non bilancia la crescita di $(M - 1)$, pertanto **non si ha efficienza in potenza**.

Per calcolare la soglia facciamo una ulteriore massimizzazione:

$$M - 1 < M \quad (6.103)$$

e

$$Q(x) \leq \frac{1}{2} e^{-\frac{x^2}{2}} \leq e^{-\frac{x^2}{2}} \quad (6.104)$$

Quindi otteniamo

$$\begin{aligned} P(e) &\leq (M - 1) Q \left(\sqrt{\log_2 M \cdot \frac{\mathcal{E}_b}{N_0}} \right) \leq \\ &\leq M e^{-\frac{\log_2 M \cdot \mathcal{E}_b}{2 \cdot N_0}} = \\ &= 2^k e^{-\frac{k}{2} \cdot \frac{\mathcal{E}_b}{N_0}} = \\ &= e^{k \ln 2} e^{-\frac{k}{2} \cdot \frac{\mathcal{E}_b}{N_0}} = \\ &= e^{-\frac{k}{2} \left(\frac{\mathcal{E}_b}{N_0} - 2 \ln 2 \right)} \end{aligned} \quad (6.105)$$

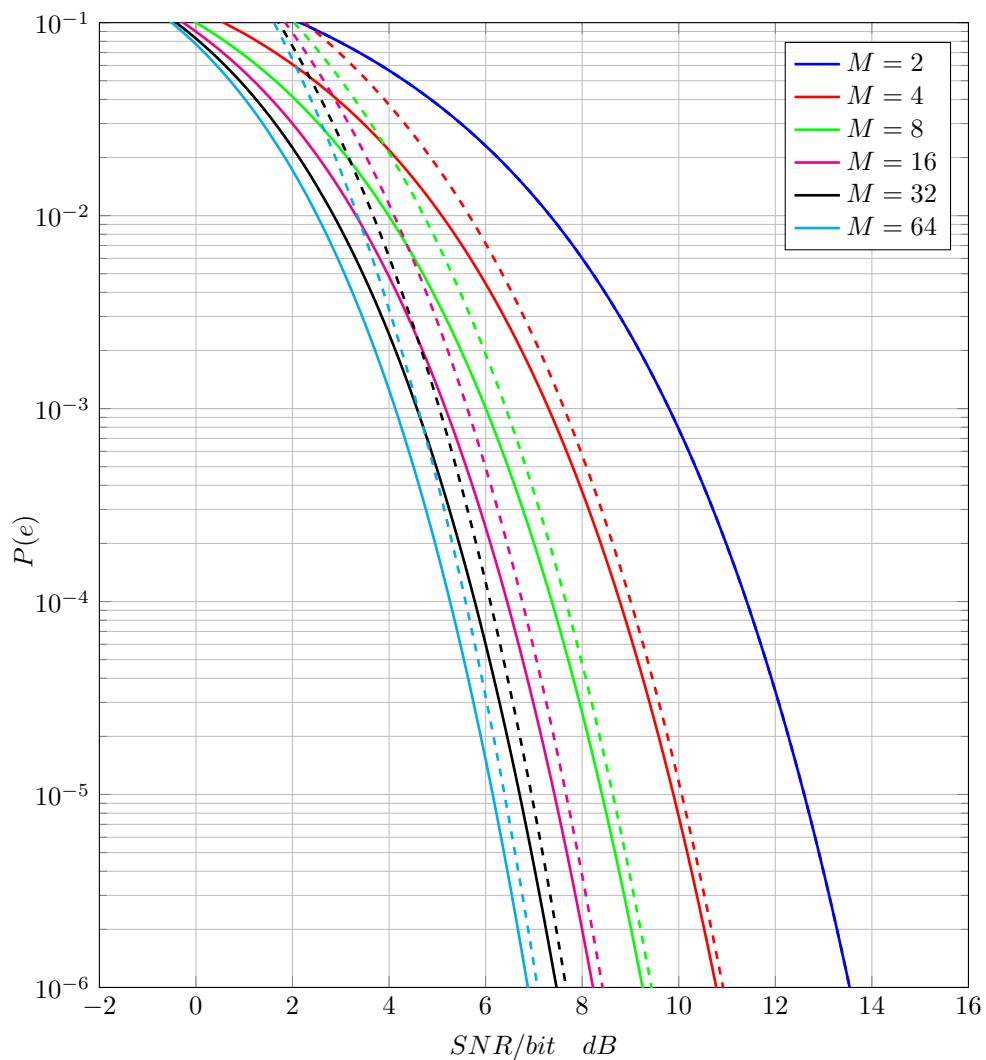


Figura 6.25: Probabilità di errore per rivelazione coerente di segnali ortogonali e relativi bound (curve tratteggiate).

k diverge al crescere di M e quindi andando a sostituire nell'espressione della $P(e)$ osserviamo che se il coefficiente $\left(\frac{\varepsilon_b}{N_0} - 2 \ln 2\right)$ è positivo allora $P(e)$ tende a zero, altrimenti diverge.

Quindi:

$$\lim_{k \rightarrow \infty} P(e) = 0 \quad \text{se} \quad \frac{\mathcal{E}_b}{N_0} > 2 \ln 2 \approx 1,4 \text{ dB} \quad (6.106)$$

In realtà, facendo un calcolo più preciso, ma complicato, si trova la vera soglia da rispettare, che è la metà di quella che abbiamo trovato:

$$\lim_{k \rightarrow \infty} P(e) = 0 \quad \text{se} \quad \frac{\mathcal{E}_b}{N_0} > \ln 2 \approx -1,6 \text{ dB} \quad (6.107)$$

In definitiva

$$0 \leq P(e) \leq e^{-\frac{k}{2} \left(\frac{\mathcal{E}_b}{N_0} - 2 \ln 2 \right)} \quad (6.108)$$

All'aumentare di M la $P(e)$ diminuisce. Ciò significa che se è rispettato il vincolo $\frac{\mathcal{E}_b}{N_0} > \ln 2$ possiamo rendere la $P(e)$ piccola a piacere semplicemente aumentando M .

Questa arbitrarietà sulla probabilità di errore ci ricorda il teorema di Shannon sulla codifica di canale, il quale afferma che se il bit rate è minore della capacità di canale allora esiste un codice che permette di rendere la $P(e)$ arbitrariamente bassa. Il problema è che questo teorema è solo un teorema di esistenza ma non ci dice qual è questo codice.

Osserviamo che il caso di segnalazione ortogonale è molto simile al teorema di Shannon, poiché se si verifica una certa condizione (che è $R_b \leq C$ per il teorema di Shannon e $\frac{\mathcal{E}_b}{N_0} > \ln 2$ per la segnalazione ortogonale) la $P(e)$ si può rendere piccola a piacere. In questo caso sappiamo pure come abbassare la $P(e)$.

Le due condizioni sembrano diverse ma in realtà si dimostra che nel caso di canale AWGN sono la stessa cosa e per questo la soglia $\ln 2$ è detta **limite di Shannon**.

La capacità di canale C per il canale AWGN a banda limitata è:

$$C = W \log_2 \left(1 + \frac{P}{N_0 W} \right) \quad (6.109)$$

dove W è la banda del canale e P è la potenza del segnale nell'intervallo T .

Da questa espressione dobbiamo valutare la capacità del canale AWGN (a banda infinita) e quindi dobbiamo far tendere W all'infinito:

$$C_{AWGN} = \lim_{W \rightarrow \infty} W \log_2 \left(1 + \frac{P}{N_0 W} \right) \quad (6.110)$$

Se moltiplichiamo la potenza del segnale per la durata del segnale stesso otteniamo l'energia media del segnale, cioè $PT = \mathcal{E}_s$, da cui ricaviamo:

$$P = \frac{\mathcal{E}_s}{T} = \frac{\log_2 M \cdot \mathcal{E}_b}{\log_2 M \cdot T_b} = \mathcal{E}_b R_b \quad (6.111)$$

e sostituendo nell'espressione della capacità otteniamo:

$$C_{AWGN} = \lim_{W \rightarrow \infty} W \log_2 \left(1 + \frac{\mathcal{E}_b}{N_0} \cdot \frac{R_b}{W} \right) \quad (6.112)$$

L'efficienza spettrale $\frac{R_b}{W}$ è un parametro significativo del canale che ci dice a che velocità possiamo trasmettere (R_b) su una certa banda (W). Il limite che stavamo valutando è il seguente:

$$\begin{aligned} C_{AWGN} &= \lim_{W \rightarrow \infty} W \log_2 \left(1 + \frac{\mathcal{E}_b}{N_0} \cdot \frac{R_b}{W} \right) = \\ &= \lim_{W \rightarrow \infty} \frac{\ln \left(1 + \frac{\mathcal{E}_b}{N_0} \cdot \frac{R_b}{W} \right)}{\ln 2 \cdot \frac{1}{W}} = \\ &= \frac{\mathcal{E}_b}{N_0} \cdot \frac{R_b}{\ln 2} \end{aligned} \quad (6.113)$$

Non c'è più la dipendenza dalla banda, come è giusto che sia, dal momento che la banda del canale AWGN è infinita.

Quindi il vincolo di Shannon, $R_b < C_{AWGN}$, diventa:

$$\boxed{\frac{\mathcal{E}_b}{N_0} > \ln 2} \quad (6.114)$$

Quindi non si può pensare di fare una trasmissione numerica non rispettando questo vincolo. Il valore $\ln 2$ equivale a circa -1.6 dB che è comunque un valore molto basso (in lineare è un valore minore di 1).

Se questo limite non è rispettato la probabilità di errore tende a 1. Quindi per avere una trasmissione affidabile, cioè con $P(e)$ arbitrariamente bassa, dobbiamo rispettare questo vincolo. Nel caso della segnalazione ortogonale ciò equivale a prendere M abbastanza grande.

6.8 PARAMETRI DI UN SISTEMA DI TRASMISSIONE

I parametri fondamentali per valutare un sistema di trasmissione numerica sono

- Probabilità di errore $P(e)$;
- Contrasto di energia $\frac{\mathcal{E}_b}{N_0}$;
- Ordine del modulatore M .

Vediamo a cosa è legato M . Se fissiamo l'intervallo di simbolo T e facciamo variare M , vuol dire che stiamo inviando nello stesso intervallo di tempo un pacchetto di bit più lungo (perché

per codificare più segnali servono più bit) e quindi aumenta il bit rate. Di conseguenza, M è legato al bit rate. Il symbol rate resta lo stesso ma aumentando i bit per simbolo aumenta il bit rate.

Il bit rate da solo non basta, bensì dobbiamo metterlo in relazione alla banda di trasmissione. Quindi introduciamo un altro parametro:

- Efficienza spettrale $\frac{R_b}{W}$.

dove R_b è il bit rate e W è la banda. Si misura in $\frac{\text{bit}}{\text{sec} \cdot \text{Hz}}$. Bit rate e symbol rate sono legati dalla relazione:

$$R_b = R_s \cdot \log_2 M \quad (6.115)$$

Per valutare l'efficienza spettrale dobbiamo quindi conoscere la banda W della segnalazione. A questo punto dovremmo andare a valutare la banda di ogni tipo di segnale trasmesso, ma questo richiede dei calcoli anche abbastanza complicati, per questo si preferisce far riferimento al **teorema di Landau-Pollack**:

$$2WT = L \quad (6.116)$$

Questo teorema fornisce il valore della banda in maniera semplice ed in relazione ad L .

Vediamo come si giustifica qualitativamente questo teorema.

Quando facciamo la rappresentazione nello spazio dei segnali associamo ogni segnale ad un vettore L -dimensionale. Queste L componenti ci permettono di ricostruire in maniera perfetta il segnale. Immaginiamo che L sia il numero di campioni che ci servono per ricostruire il segnale. Per campionare il segnale dobbiamo avere una frequenza di campionamento pari a quella di Nyquist, cioè $2W$, ed inoltre il segnale ha durata T , quindi il numero di campioni che prendiamo è $2WT$. Pertanto il numero di campioni che servono per ricostruire il segnale, cioè L , è proprio $2WT$.

Dal teorema ricaviamo W :

$$W = \frac{L}{2T} = \frac{L \cdot R_s}{2} = \frac{L \cdot R_b}{2 \log_2 M} \quad (6.117)$$

Quindi l'efficienza spettrale sarà:

$$\frac{R_b}{W} = \frac{2 \log_2 M}{L}$$

(6.118)

Quindi i parametri fondamentali per valutare un sistema di trasmissione numerica sono

- Probabilità di errore $P(e)$;
- Contrasto di energia $\frac{E_b}{N_0}$;

- Efficienza spettrale $\frac{R_b}{W}$.

Poiché la relazione tra $\frac{\mathcal{E}_b}{N_0}$ e $\frac{R_b}{W}$ è legata al teorema di Shannon sulla codifica di canale, il piano individuato da queste due grandezze è detto piano di Shannon $\left(\frac{\mathcal{E}_b}{N_0}, \frac{R_b}{W}\right)$ per un fissato valore di $P(e)$. Se cambiamo la $P(e)$ cambia la rappresentazione perché cambia $\frac{\mathcal{E}_b}{N_0}$.

Vediamo ora come il piano $\frac{\mathcal{E}_b}{N_0}$ e $\frac{R_b}{W}$ è messo in relazione al teorema di Shannon: se vogliamo una $P(e)$ bassa, per il teorema di Shannon sulla codifica di canale, dobbiamo avere il bit rate minore della capacità di canale, che è un parametro che cambia a seconda del canale che usiamo. Infatti

$$C = W \log_2 \left(1 + \frac{\mathcal{E}_b}{N_0} \cdot \frac{R_b}{W} \right) \quad (6.119)$$

Questo è il valore, calcolato precedentemente, della capacità del canale AWGN a banda limitata. Cioè, invece di considerare un canale AWGN consideriamo un canale con rumore che è a densità spettrale costante solo in una banda rigorosamente limitata W (infatti quello che accade fuori dalla banda non ci interessa).

Per il teorema di Shannon dobbiamo avere:

$$R_b \leq W \log_2 \left(1 + \frac{\mathcal{E}_b}{N_0} \cdot \frac{R_b}{W} \right) \quad (6.120)$$

che equivale a

$$2^{R_b} \leq 1 + \frac{\mathcal{E}_b}{N_0} \cdot \frac{R_b}{W} \quad (6.121)$$

e quindi

$$\frac{\mathcal{E}_b}{N_0} \geq \frac{2^{\frac{R_b}{W}} - 1}{\frac{R_b}{W}}$$

(6.122)

Rappresentiamo questa curva andando a valutare alcuni punti significativi. Se $\frac{R_b}{W} = 0$ risulta $\frac{\mathcal{E}_b}{N_0} = \ln 2 = -1.6$ dB (sfruttando il limite notevole $\lim_{x \rightarrow 0} \frac{a^x - 1}{0} = \ln a$), che è proprio il limite di Shannon che abbiamo visto per il canale AWGN. Per $\frac{R_b}{W} = 1$ risulta $\frac{\mathcal{E}_b}{N_0} = 1$, mentre per $\frac{R_b}{W} \rightarrow \infty$ si ha che $\frac{\mathcal{E}_b}{N_0} \rightarrow \infty$. In figura 6.26 è mostrato il piano di Shannon.

Per avere una trasmissione affidabile dobbiamo verificare la disuguaglianza vista, quindi la curva in figura 6.26 rappresenta il valore limite di una segnalazione affidabile. Le segnalazioni che ricadono al di sotto della curva sono affidabili, quelle al di sopra non lo sono. Tra le segnalazioni affidabili si fa una ulteriore classificazione:

- Se $\frac{R_b}{W} > 1$ le segnalazioni sono efficienti in banda. Infatti per $\frac{R_b}{W} > 1$ vuol dire che $R_b > W$ e quindi sfruttiamo meglio la banda. Tuttavia l'efficienza in banda implica in generale inefficienza in potenza;

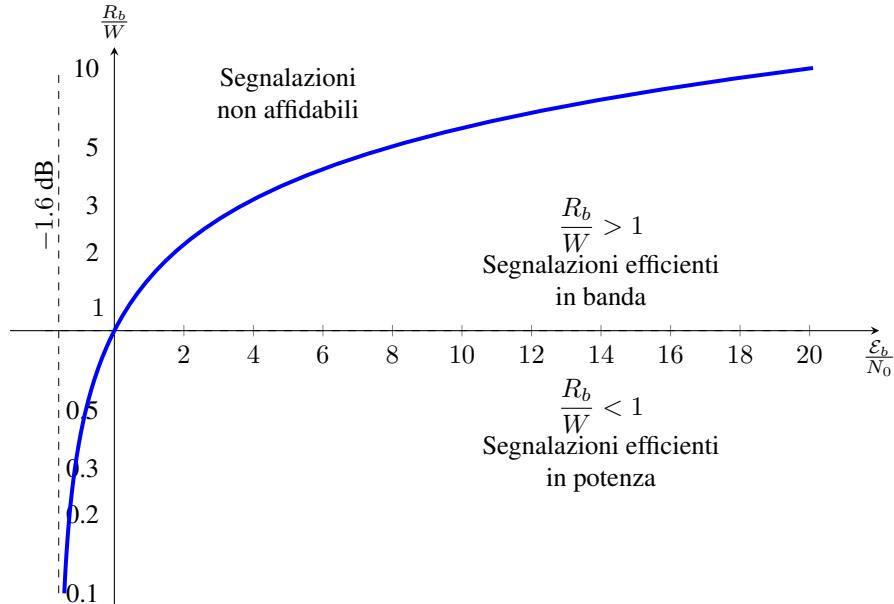


Figura 6.26: Piano di Shannon

- Se $\frac{R_b}{W} > 1$ le segnalazioni sono efficienti in potenza.

In generale possiamo scrivere la $P(e)$ di una segnalazione come

$$P(e) \cong \bar{\nu} Q \left(\sqrt{\gamma_M \frac{E_b}{N_0}} \right) \quad (6.123)$$

dove γ_M è un parametro che ci permette di stimare l'efficienza in potenza. Valutando invece l'efficienza spettrale $\frac{R_b}{W}$ riusciamo a stimare l'efficienza in banda.

Facciamo ora una tabella riassuntiva delle varie segnalazioni:

Vediamo ora, caso per caso, come possiamo definire le varie segnalazioni:

- **PAM:** all'aumentare di M , γ_M diminuisce e quindi la $P(e)$ aumenta (essendo Q una funzione decrescente, quando il suo argomento diminuisce la funzione cresce). Quindi, per mantenere $P(e) = \text{cost}$ all'aumentare di M bisogna aumentare il contrasto di energia $\frac{E_b}{N_0}$ in modo da bilanciare la diminuzione di γ_M . La segnalazione è quindi **inefficiente in potenza**. Tuttavia, all'aumentare di M , $\frac{R_b}{W}$ aumenta e quindi a parità di R_b riusciamo ad avere una banda più stretta. La segnalazione è quindi **efficiente in banda**.

	$\bar{\nu}$	γ_M	$\frac{R_b}{W}$
PAM	$\frac{2(M-1)}{M} \approx 2$	$\frac{6 \log_2 M}{M^2 - 1}$	$2 \log_2 M$
PSK	2	$2 \log_2 M \cdot \sin^2 \frac{\pi}{M}$	$\log_2 M$
ORTOGONALE	$M - 1$	$\log_2 M$	$\frac{2 \log_2 M}{M}$

Tabella 6.1

Tabella riassuntiva: confronto tra le tecniche di modulazione.

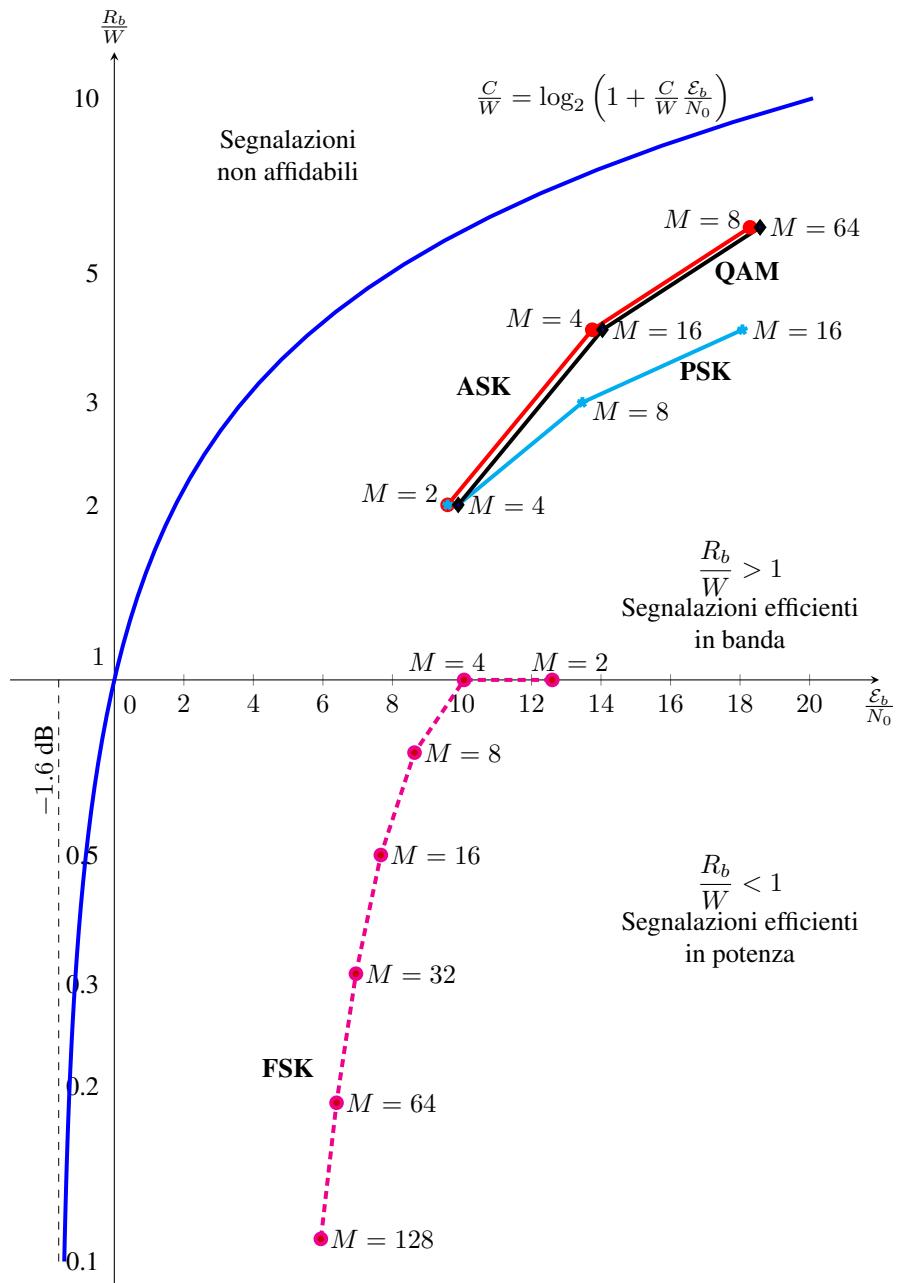
- **PSK:** il comportamento di questo tipo di modulazione è simile alle modulazioni di ampiezza, infatti γ_M diminuisce all'aumentare di M e quindi la segnalazione è **inefficiente in potenza**. D'altra parte, $\frac{R_b}{W}$ aumenta all'aumentare di M . Pertanto la segnalazione è **efficiente in banda**.
- **Ortogonale:** osserviamo un comportamento inverso rispetto a quello delle segnalazioni precedenti. Infatti, all'aumentare di M aumenta γ_M e quindi per mantenere costante la probabilità di errore dobbiamo diminuire il contrasto di energia. La segnalazione è quindi **efficiente in potenza**. Tuttavia paghiamo questa efficienza in potenza in termini di banda perché all'aumentare di M diminuisce $\frac{R_b}{W}$ e quindi la banda è sfruttata peggio. Ne consegue che la segnalazione ortogonale è **inefficiente in banda**.

In figura 6.27 si riportano le varie segnalazioni nel piano di Shannon.

Il grafico viene costruito per un valore costante di probabilità di errore. Una volta fissata la $P(e)$ (nel nostro caso $P(e) = 10^{-5}$) si ricava $\frac{E_b}{N_0}$ dalla formula della $P(e)$. Inoltre, precedentemente avevamo valutato le prestazioni delle varie segnalazioni al variare di M , quindi al crescere di M , dallo studio delle varie curve, sappiamo di quanto si deve incrementare o ridurre il contrasto di energia. Ad esempio, nel passaggio da 2-PAM al 4-PAM il contrasto di energia raddoppia.

Inoltre, come abbiamo visto in precedenza, in corrispondenza dei punti dell' M -PAM ci sono quelli del QAM con M^2 livelli. Infine osserviamo che la M ha un andamento logaritmico e varia per valori discreti, cioè vira su punti ben definiti.

Più le curve si avvicinano alla curva limite di Shannon più le segnalazioni funzionano meglio.

Figura 6.27: Piano di Shannon: confronto tra le tecniche di modulazione per $P(e) = 10^{-5}$.