

Probabilità processi e statistica

PRIMA di proseguire con lo studio dei segnali, affrontiamo la teoria delle probabilità, i cui risultati saranno utilizzati molto estesamente nei capitoli seguenti. Dopo aver brevemente riassunto assiomi, teoremi base e concetti di *calcolo* delle probabilità, viene definita la *variabile aleatoria*, caratterizzata mediante distribuzioni e momenti, e discusse le v.a. uniforme e gaussiana. Si passa quindi a descrivere i *processi aleatori*, le diverse modalità di definirne delle medie, e le proprietà di stazionarietà ed ergodicità. Viene poi sviluppata la teoria che consente di ottenere la descrizione probabilistica di una *trasformazione* di variabile aleatoria, ed introdotta la densità gaussiana multidimensionale con le sue proprietà. Infine si affrontano temi di statistica, come la teoria della decisione, la stima di parametro e di intervallo, i test di verifica di ipotesi.

6.1 Teoria delle probabilità

Tratta delle caratteristiche *regolari* di fenomeni *irregolari* o *casuali*. Una prima *definizione* di probabilità è quella fornita dalla teoria *frequentista*, che la associa al limite a cui tende il rapporto tra numero di casi favorevoli rispetto al numero di casi totali: se ripetendo N volte un esperimento la circostanza A si verifica per n_A volte la sua frequenza relativa vale n_A/N , da cui si deriva la probabilità di A come

$$Pr_A = \lim_{N \rightarrow \infty} \frac{n_A}{N}$$

In termini più astratti, l'insieme di tutte le circostanze possibili può essere pensato come un insieme algebrico, i cui elementi (o punti) sono appunto le diverse circostanze. I punti possono essere raggruppati in sottoinsiemi (eventualmente vuoti o di un solo punto) per i quali valgono le proprietà di unione, intersezione, complemento, inclusione...

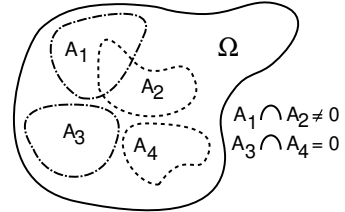
I fenomeni fisici sono posti in relazione con i punti degli insiemi suddetti mediante il concetto di *spazio campione* Ω , che è l'unione di tutti i possibili risultati di un *fenomeno aleatorio*, mentre i sottoinsiemi dello spazio campione sono detti *eventi*. L'intero spazio è l'*evento certo*, mentre l'insieme vuoto corrisponde all'evento impossibile ϕ (od evento *nullo*). Una *unione* \cup di eventi corrisponde all'evento che si verifica ogni qualvolta se ne verifichi *un suo componente*, mentre l'*intersezione* \cap è verificata se *tutti* i componenti lo

sono. Ad esempio, il lancio di un dado genera uno spazio con 6 punti (eventi) disgiunti. Uno spazio campione può avere un numero di punti finito, infinito numerabile, o infinito.

6.1.1 Assiomi delle probabilità

Costituiscono le basi da cui derivano i teoremi successivi, affermando che

- $0 \leq Pr(A) \leq 1$: la probabilità di un evento è compresa tra 0 ed 1;
- $Pr(\Omega) = 1$: la probabilità dell'evento certo è 1;
- Se $Pr(A_i \cap A_j) = \emptyset$ allora $Pr(\cup A_i) = \sum Pr(A_i)$: la probabilità dell'unione di eventi *disgiunti* è la somma delle singole probabilità.



6.1.2 Teoremi di base

- $Pr(\emptyset) = 0$: la probabilità dell'evento impossibile è nulla;
- $Pr(A \cap B) + Pr(A \cap \bar{B}) = Pr(A)$, e $Pr(B) + Pr(\bar{B}) = 1$: un evento ed il suo complemento riempiono lo spazio (detto anche teorema delle *probabilità totali*¹);
- $Pr(A \cup B) = Pr(A) + Pr(B) - Pr(A \cap B)$: la probabilità dell'evento intersezione si conta una volta sola. Esempio alla nota²;
- Se $B \subseteq A$ allora $Pr(B) \leq Pr(A)$: quando l'evento B è contenuto in A il verificarsi del primo implica il secondo.

6.1.3 Probabilità congiunta, condizionata e marginale

Può avvenire che il verificarsi di un evento *influenzi* il verificarsi o meno di un altro. Si dice allora che lo condiziona, ovvero che l'evento influenzato è *condizionato*. La probabilità che avvenga A , noto che B (evento *condizionante*) si sia verificato, si scrive $Pr(A/B)$, e si legge probabilità (condizionata) di A dato B , definita³ come

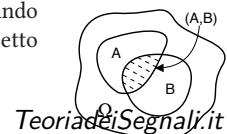
$$Pr(A/B) = \frac{Pr(A, B)}{Pr(B)} \tag{6.1}$$

in cui $Pr(A, B) = Pr(A \cap B)$ è la probabilità *congiunta* che A e B si verifichino entrambi, ed a patto che $Pr(B) \neq 0$ (altrimenti anche $Pr(A/B)$ è zero!). Viceversa, le probabilità dei singoli eventi $Pr(A)$ e $Pr(B)$ sono indicate come probabilità *marginali*.

¹Utile per scrivere la probabilità di un evento come “1 meno” quella dell'evento complementare.

²Lanciando un dado, la probabilità $Pr(\text{pari} \cup > 2)$ di ottenere un numero pari, *oppure* più grande di due, è la *somma* delle probabilità dei singoli eventi $Pr(\text{pari}) = \frac{3}{6}$ e $Pr(> 2) = \frac{4}{6}$, meno quella che si verifichino assieme $Pr(\text{pari} \cap > 2) = \frac{2}{6}$. Pertanto: $Pr(\text{pari} \cup > 2) = \frac{3}{6} + \frac{4}{6} - \frac{2}{6} = \frac{5}{6}$.

La relazione può essere verificata ricorrendo al diagramma in figura, ed interpretando $Pr(A/B)$ come il rapporto tra la misura di probabilità dell'evento congiunto, rispetto a quella dell'evento condizionante.



Esercizio Valutare la probabilità condizionata $Pr(A/B)$ che lanciando un dado si ottenga un numero pari (evento $A = (\text{pari})$), condizionatamente all'evento B che il numero sia > 2 .
Soluzione alla nota⁴.

Invertendo la definizione (6.1) la probabilità *congiunta* può essere ottenuta anche come $Pr(A, B) = Pr(A/B) Pr(B)$; inoltre, gli eventi condizionante e condizionato si possono invertire di ruolo, permettendo di scrivere anche: $Pr(A, B) = Pr(B/A) Pr(A)$. Eguagliando le due ultime espressioni per la probabilità congiunta si ottiene la via per calcolare una probabilità *condizionata* a partire dall'altra, qualora si conoscano entrambe le *marginali*:

$$Pr(A/B) = \frac{Pr(B/A) Pr(A)}{Pr(B)} \quad \text{ed anche} \quad Pr(B/A) = \frac{Pr(A/B) Pr(B)}{Pr(A)} \quad (6.2)$$

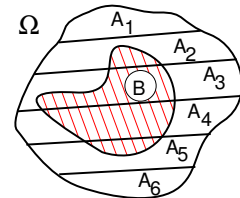
6.1.4 Probabilità a priori e a posteriori, teorema di Bayes

A volte un determinato evento A non può essere osservato direttamente, ma se A è in qualche modo legato ad un secondo evento B , che invece possiamo osservare, la probabilità condizionata $Pr(A/B)$ prende il nome di probabilità *a posteriori*, poiché indica un valore di probabilità valutato *dopo* la conoscenza di B . Viceversa, in tale contesto la probabilità marginale $Pr(A)$ viene ora indicata come *a priori*, ovvero *presunta* senza aver potuto osservare nulla.

In generale si conosce solamente $Pr(A)$ e $Pr(B/A)$, mentre per calcolare $Pr(A/B)$ occorre conoscere anche $Pr(B)$, vedi (6.2). Quest'ultima quantità si determina *saturando* la probabilità congiunta $Pr(A, B)$ rispetto a tutti i possibili eventi marginali A_i :

$$Pr(B) = \sum_i Pr(B, A_i) = \sum_i Pr(B/A_i) Pr(A_i)$$

a patto che risulti $Pr(A_i, A_j) = 0$ e $\cup A_i = \Omega$, ovvero che gli eventi A_i siano disgiunti e che il loro insieme $\{A_i\}$ costituisca una partizione dello spazio degli eventi Ω , come rappresentato in figura.



L'ultima relazione ci permette di enunciare il *teorema di Bayes*, che mostra come ottenere le probabilità *a posteriori* a partire da quelle *a priori* e da quelle condizionate *in senso opposto*:

$$Pr(A_i/B_j) = \frac{Pr(B_j/A_i) Pr(A_i)}{\sum_k Pr(B_j/A_k) Pr(A_k)}$$

⁴Il risultato è pari alla probabilità $Pr(A, B) = Pr(\text{pari}, > 2)$ che i due eventi si verifichino contemporaneamente, divisa per la probabilità $Pr(B) = Pr(> 2)$ che il numero sia > 2 .

Si rifletta sulla circostanza che la probabilità del pari $Pr(A) = \frac{1}{2}$, quella $Pr(B) = \frac{4}{6}$, o quella congiunta di entrambi $Pr(A, B) = \frac{2}{6}$, sono tutte riferite ad *un qualunque* lancio di dado, mentre $Pr(\text{pari} / > 2)$ è relativa ad un numero ridotto di lanci, solo quelli che determinano un risultato > 2 . Pertanto, essendo $Pr(B) \leq 1$, si ottiene $Pr(A/B) \geq Pr(A, B)$; infatti per l'esempio del dado si ottiene $Pr(\text{pari} / > 2) = Pr(\text{pari}, > 2) / Pr(> 2) = \frac{2/6}{4/6} = \frac{1}{2}$, che è maggiore di $Pr(\text{pari}, > 2) = \frac{1}{3}$.

Si ottiene invece $Pr(A/B) = Pr(A, B)$ solo se $Pr(B) = 1$, ossia se B corrisponde all'unione di tutti gli eventi possibili.

Al § ?? è mostrata l'applicazione di queste considerazioni ad un problema di decisione statistica tipico delle telecomunicazioni, relativo alla ricezione binaria. Di seguito, invece, è illustrato un esempio più diretto di applicazione del teorema di Bayes.

Esempio Un sistema di comunicazione radio è affetto da attenuazioni supplementari causate da pioggia. Indicando con FS l'evento che il sistema vada fuori servizio, e conoscendo le probabilità condizionate $Pr(FS/piove) = 0.5$, $Pr(FS/non\ piove) = 0.05$ e la probabilità marginale $Pr(piove) = 0.03$, determinare:

1. La probabilità di fuori servizio $Pr(FS)$, indipendentemente dal verificarsi o meno dell'evento piovoso;
2. La probabilità che stia piovendo, sapendo che il sistema è fuori servizio.

Risposte alla nota⁵).

6.1.5 Indipendenza statistica

Si verifica quando

$$Pr(A/B) = Pr(A)$$

in quanto il verificarsi di B non influenza A . Come conseguenza, per due eventi statisticamente indipendenti avviene che

$$Pr(A, B) = Pr(A) Pr(B) \quad (6.3)$$

Esempi

- Quale è la probabilità che, lanciando 3 volte un dado, esca 3 volte 1? Risultato (⁶).
- Un'urna contiene 2 biglie bianche e 3 nere. Qual è la probabilità che su 2 estrazioni consecutive *senza reinserimento*, escano le 2 biglie bianche? Risultato (⁷).
- Qual è la probabilità che 2 carte, estratte a caso da un mazzo da bridge da 52, siano K e Q? Risultato (⁸).

6.2 Variabili aleatorie

Finora si è parlato di *eventi* in modo astratto, mentre spesso ci si trova ad associare ad ogni punto dello spazio campione un valore numerico: lo spazio campione Ω diventa

⁵La probabilità marginale di fuori servizio si calcola applicando il teorema delle probabilità totali

$$Pr(FS) = Pr(FS/piove) \cdot Pr(piove) + Pr(FS/non\ p.) \cdot Pr(non\ p.) = .5 \cdot .03 + .05 \cdot .97 = .0635 = 6.35\%$$

dato che $Pr(non\ piove) = 1 - Pr(piove) = .97$. Applicando il teorema di Bayes si trova quindi

$$Pr(piove/FS) = \frac{Pr(FS/piove) \cdot Pr(piove)}{Pr(FS)} = \frac{.5 \cdot .03}{.0635} = .236 = 23.6\%$$

Si noti come la probabilità *a priori* che piova (3%) venga rimpiazzata dal suo valore *a posteriori* (23.6%) grazie alla nuova informazione di cui disponiamo (collegamento fuori servizio). Per una definizione più precisa delle probabilità a priori ed a posteriori si veda il § ??.

⁶E' pari al prodotto delle probabilità marginali, essendo i lanci statisticamente indipendenti, visto che il dado è "senza memoria". Pertanto il risultato è $(\frac{1}{6})^3 = \frac{1}{216} \approx 4.6296 \cdot 10^{-3}$.

⁷Anche l'urna è senza memoria, ma non l'esperimento aleatorio, visto che dopo la prima estrazione le biglie restano in 4! Pertanto ora il prodotto delle probabilità marginali risulta $\frac{2}{5} \cdot \frac{1}{4} = \frac{1}{10}$.

⁸ $Pr(K, Q) = Pr(K\ prima, Q\ seconda) + Pr(Q\ prima, K\ seconda) = Pr(K\ prima) \cdot Pr(Q\ seconda/K\ prima) + Pr(Q\ prima) \cdot Pr(K\ seconda/Q\ prima) = 2 \left(\frac{4}{52} \frac{4}{51} \right) = \frac{8}{663} \approx 1.2 \cdot 10^{-2}$

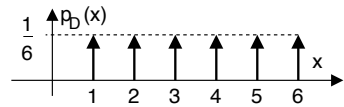
allora l'insieme dei numeri e prende il nome di *variabile aleatoria*, d'ora in poi spesso abbreviato in *v.a.* Il verificarsi di un evento corrisponde ora all'assegnazione di un valore (tra i possibili) alla v.a.; tale valore "prescelto" prende dunque il nome di *realizzazione* della v.a. Distinguiamo poi tra variabili aleatorie *discrete* e *continue*, a seconda se la grandezza che descrivono abbia valori numerabili o continui⁹. La caratterizzazione della variabile aleatoria in termini probabilistici si ottiene indicando come la "massa di probabilità" si distribuisce sull'insieme di valori che essa può assumere, per mezzo delle due funzioni (di v.a.) seguenti.

6.2.1 Densità di probabilità e funzione di distribuzione

Come la massa di un oggetto *non omogeneo* è distribuita in modo più o meno denso in regioni differenti del suo volume complessivo, così la *densità di probabilità* (o *d.d.p.*) indica su quali valori della variabile aleatoria si concentra la probabilità. Ad esempio, la densità della v.a. discreta associata al lancio di un dado può essere scritta:

$$p_D(x) = \sum_{n=1}^6 \frac{1}{6} \delta(x-n) \quad (6.4)$$

il cui significato discutiamo subito, con l'aiuto del grafico a lato, in cui D indica la v.a. (il numero che uscirà), e x una sua realizzazione (una delle 6 facce). I 6 impulsi centrati in $x = n$ rappresentano una *concentrazione* di probabilità nei sei possibili valori, e l'area di tali impulsi è esattamente pari alla probabilità di ognuno dei sei risultati. È facile verificare che



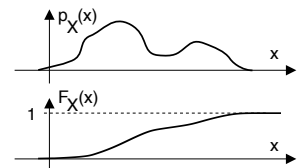
$$\int_{-\infty}^{\infty} p_D(x) dx = 1 \quad \text{e che risulta} \quad \int_a^b p_D(x) dx = \Pr\{a < D \leq b\}$$

ovvero pari alla probabilità che la v.a. D assuma un valore tra a e b . In particolare, non potendosi verificare una probabilità negativa, si ha $p_D(x) \geq 0$ con $\forall x$.

Una funzione di v.a. strettamente collegata alla densità è la funzione di *ripartizione* o di *distribuzione*¹⁰, definita come

$$F_X(x) = \int_{-\infty}^x p_X(\xi) d\xi = \Pr\{X \leq x\} \quad (6.5)$$

e che risulta una funzione non decrescente di x , limitata ad un valore massimo di 1, ed il cui andamento mostriamo a lato sotto a quello di $p_X(x)$, per la quale ovviamente risulta $p_X(x) = \frac{d}{dx} F_X(x)$; nel caso invece della v.a. discreta D , la



⁹Un esempio classico di v.a. discreta è quello del lancio di un dado, un altro sono i numeri del lotto. Una v.a. continua può essere ad esempio un valore di pressione atmosferica in un luogo, oppure l'attenuazione di una trasmissione radio dovuta a fenomeni atmosferici.

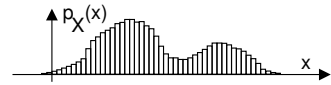
¹⁰In realtà, l'ordine storico è quello di definire prima $F_X(x)$ come la probabilità che X sia non superiore ad un valore x , ovvero $F_X(x) = \Pr\{X \leq x\}$, e quindi $p_X(x) = \frac{dF_X(x)}{dx}$. Il motivo di tale "priorità" risiede nel fatto che $F_X(x)$ presenta minori "difficoltà analitiche" di definizione (ad esempio presenta solo discontinuità di prima specie, anche con v.a. discrete).

relativa funzione di distribuzione è discontinua¹¹.

Ora è ancora più evidente la circostanza che $p_X(x)$ è una *densità*, e diviene una probabilità solo quando moltiplicata per un intervallo di x ⁽¹²⁾.

Istogramma Qualora non si disponga di una espressione analitica idonea a rappresentare il modo con cui si distribuiscono i valori di una v.a., può essere utile svolgerne una *stima* mediante un *istogramma*. Questo assume l'aspetto di una versione per così dire *quantizzata* della d.d.p. incognita, e si ottiene a partire da una serie di realizzazioni¹³ della v.a., suddividendo il campo di variabilità della grandezza X in sotto-intervalli, e disegnandovi rettangoli verticali, ognuno di altezza pari al numero di volte che (nell'ambito del campione statistico a disposizione) X assume un valore in quell'intervallo, come rappresentato in figura.

Dividendo l'altezza di ogni rettangolo per il numero di osservazioni N , si ottiene una approssimazione di $p_X(x)$, via via più precisa con $N \rightarrow \infty$, e con una contemporanea riduzione dell'estensione degli intervalli.



6.2.2 Valore atteso, momento e momento centrato

Si tratta di grandezze per così dire *riassuntive* del modo con cui si distribuiscono i valori di una v.a., e sono definite a partire da una *generica funzione di variabile aleatoria*¹⁴ che indichiamo con $g(x)$.

Valore atteso Si definisce *valore atteso* (o *media di insieme*¹⁵) di $g(x)$ rispetto alla variabile aleatoria X la quantità

$$E_X \{g(x)\} = \int_{-\infty}^{\infty} g(x) p_X(x) dx \quad (6.6)$$

che corrisponde ad una *media pesata*, in cui i valori assunti da $g(x)$ in corrispondenza ad un certo x sono *pesati* mediante il corrispettivo valore di probabilità $p_X(x) dx$; tale operazione di media integrale è indicata con la notazione $E_X \{.\}$ ¹⁶, mediante la quale

A fianco è mostrata la $F_D(x)$ relativa al lancio di un dado: ricordiamo
¹¹infatti che la derivata di un gradino è un impulso di area pari al dislivello, e dunque applicando la (6.5) alla (6.4) si ottiene il risultato illustrato.



¹²Infatti la probabilità che X cada tra x_0 e $x_0 + \Delta x$ vale $\int_{x_0}^{x_0+\Delta x} p_X(x) dx \approx p_X(x_0) \Delta x$.

¹³Ricavate ad esempio da basi di dati anagrafici, sanitari, meteorologici o quant'altro, oppure effettuando una apposita *campagna di misura* basata su di un *campione statistico* di adeguata numerosità (vedi anche § 6.6).

¹⁴Un esempio di funzione di v.a. potrebbe essere il valore della vincita associata ai 13 in schedina, che dipende dalla v.a. rappresentata dai risultati delle partite, una volta noto il montepremi e le giocate. Infatti, per ogni possibile vettore di risultati, si determina un diverso numero di giocate vincenti, e quindi un diverso modo di suddividere il montepremi. Essendo i risultati improbabili giocati da un ridotto numero di schedine, a queste compete un valore maggiore in caso di vincita, ben superiore al suo *valore atteso*, indicativo invece della vincita *media*.

¹⁵Per *insieme* ci si riferisce allo *spazio campione* Ω , costituito dai possibili valori assunti dalla v.a. X .

¹⁶In effetti, la E simboleggia la parola *Expectation*, che è il termine inglese usato per indicare il valore atteso.

si indica a pedice la v.a. (X) rispetto a cui eseguire la pesatura.

Nel caso di una funzione di più v.a. $g(x, y)$ il relativo valore atteso è calcolato in base alla d.d.p. *congiunta*, ovvero

$$E_{X,Y} \{g(x, y)\} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) p_{XY}(x, y) dx dy$$

in cui $p_{XY}(x, y)$ si ottiene a partire dalle d.d.p. condizionate e marginali, come esposto al § 6.1.3. Infine, la (6.6) può essere calcolata utilizzando una d.p.p. condizionata $p_{X/Y}(x/y)$, ed in tal caso anche il valore atteso $E_{X/Y} \{g(x)\}$ è detto *condizionato*, risultando funzione di y .

Momento Qualora si ponga $g(x) = x^n$, ovvero pari alla n -esima potenza della v.a., il valore atteso prende il nome di *momento di ordine n* , e si indica come

$$m_X^{(n)} = E \{x^n\} = \int_{-\infty}^{\infty} x^n p_X(x) dx \quad (6.7)$$

Nel caso di variabili aleatorie discrete, i momenti sono definiti come $m_X^{(n)} = \sum_i x_i^n p_i$, in cui $p_i = Pr \{x = x_i\}$, pesando quindi le possibili realizzazioni x_i con le rispettive probabilità. Notiamo subito che $m_X^{(0)} = \int_{-\infty}^{\infty} p_X(x) dx = 1$. Ragioniamo ora su due importanti momenti.

Valor medio e media quadratica Il momento di *primo* ordine

$$m_X = m_X^{(1)} = \int_{-\infty}^{\infty} x p_X(x) dx \quad (6.8)$$

prende il nome di *valor medio* della v.a., a volte denominato *centroide*, e coincide con la *media aritmetica* ottenibile a partire dalla conoscenza delle realizzazioni della v.a. ottenute ripetendo all'infinito l'esperimento aleatorio. Viceversa il momento di secondo ordine

$$m_X^{(2)} = \int_{-\infty}^{\infty} x^2 p_X(x) dx$$

viene indicato come *media quadratica*.

Esempio Supponiamo che la v.a. X rappresenti l'altezza degli individui: l'altezza *media* m_X può essere *stimata* come media aritmetica delle relative misurazioni

$$\hat{m}_X = \frac{\overbrace{x_1 + x_1 + \dots}^{N_1 \text{ volte}} + \overbrace{x_2 + x_2 + \dots}^{N_2 \text{ volte}} + \dots + \overbrace{x_n + x_n + \dots}^{N_n \text{ volte}}}{N} = \frac{x_1 N_1 + x_2 N_2 + \dots + x_n N_n}{N}$$

Al tendere di $N = \sum_{i=1}^n N_i$ ad ∞ , la stima \hat{m}_X viene a coincidere con il risultato m_X fornito dalla (6.8) qualora al posto delle probabilità $p_X(x) dx$ si sostituiscano i valori $Pr(x_i)$ ottenuti tramite un istogramma $Pr(x_i) = \frac{N(x_i < x \leq x_i + \Delta x)}{N} = \frac{N_i}{N}$, tramutando così l'integrale in una sommatoria, ovvero $\int_{-\infty}^{\infty} x p_X(x) dx \Rightarrow \sum_i x_i Pr(x_i)$. Tale punto di vista motiva il senso del concetto di *pesatura* dei possibili valori di x con le rispettive frequenze.

Momento centrato Nel caso in cui $g(x) = (x - m_X)^n$ il relativo valore atteso è chiamato *momento centrato* di ordine n , ed indicato come

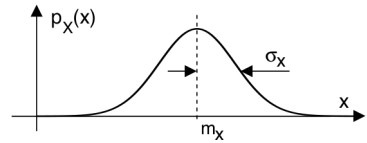
$$\mu_X^{(n)} = E \{ (x - m_X)^n \} = \int_{-\infty}^{\infty} (x - m_X)^n p_X(x) dx$$

E' immediato constatare che $\mu_X^{(0)} = 1$ e che $\mu_X^{(1)} = 0$.

Varianza E' il nome dato al momento centrato del 2° ordine, corrispondente a

$$\sigma_X^2 = \mu_X^{(2)} = E \{ (x - m_X)^2 \} = \int_{-\infty}^{\infty} (x - m_X)^2 p_X(x) dx$$

La radice quadrata σ_X della varianza σ_X^2 prende il nome di deviazione standard, e mentre la media m_X indica dove si colloca il "centro statistico" della densità di probabilità, σ_X indica quanto le singole determinazioni della v.a. siano disperse attorno ad m_x .

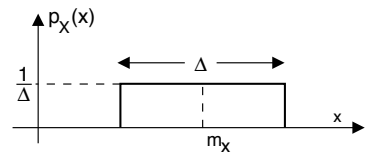


Una relazione notevole che lega i primi due momenti (centrati e non) è ⁽¹⁷⁾:

$$\sigma_X^2 = m_X^{(2)} - (m_X)^2 \tag{6.9}$$

6.2.3 Variabile aleatoria uniforme

E' caratterizzata dal presentare uno stesso valore di probabilità per tutto l'intervallo delle possibili realizzazioni, comprese tra un valore minimo ed uno massimo, come rappresentato in figura; pertanto la densità di probabilità è esprimibile mediante una funzione rettangolare



$$p_X(x) = \frac{1}{\Delta} \text{rect}_{\Delta}(x - m_X)$$

in cui Δ rappresenta l'estensione dell'intervallo di esistenza della variabile aleatoria, mentre il parametro m_X , che indica l'ascissa a cui è centrato il rettangolo, corrisponde esattamente al momento di primo ordine di X . Il calcolo della varianza¹⁸ invece fornisce: $\sigma_X^2 = \frac{\Delta^2}{12}$.

¹⁷Infatti risulta

$$\begin{aligned} \sigma_X^2 &= E \{ (x - m_X)^2 \} = E \{ x^2 + (m_X)^2 - 2xm_X \} = E \{ x^2 \} + (m_X)^2 - 2m_X E \{ x \} = \\ &= m_X^{(2)} + (m_X)^2 - 2(m_X)^2 = m_X^{(2)} - (m_X)^2 \end{aligned}$$

Si è preferito usare la notazione $E \{ x \}$, più compatta rispetto all'indicazione degli integrali coinvolti; i passaggi svolti si giustificano ricordando la proprietà distributiva degli integrali (appunto), ed osservando che il valore atteso di una costante è la costante stessa.

¹⁸Anziché calcolare σ_X^2 per la $p_X(x)$ data, calchiamo $m_X^{(2)}$ per una v.a. uniforme a media nulla, ovvero con $m_X = 0$, sfruttando il fatto che in base alla (6.9) in tal caso risulta $m_X^{(2)} = \sigma_X^2$. Si ottiene:

$$m_X^{(2)} = \int_{-\frac{\Delta}{2}}^{\frac{\Delta}{2}} x^2 \frac{1}{\Delta} dx = \frac{x^3}{3\Delta} \Big|_{-\frac{\Delta}{2}}^{\frac{\Delta}{2}} = \frac{1}{3\Delta} \left(\frac{\Delta^3}{8} + \frac{\Delta^3}{8} \right) = \frac{1}{3\Delta} 2 \frac{\Delta^3}{8} = \frac{\Delta^2}{12}$$

6.2.4 Variabile aleatoria gaussiana

A differenza del caso uniforme, la v.a. gaussiana presenta valori più probabili in prossimità del valor medio m_x , in accordo alla d.d.p. con espressione

$$p_X(x) = \frac{1}{\sqrt{2\pi}\sigma_x} \exp\left\{-\frac{(x - m_x)^2}{2\sigma_x^2}\right\} \quad (6.10)$$

ed il cui grafico dalla caratteristica forma *a campana* è mostrato a lato per diversi valori dei parametri m_x e σ_x che compaiono nella (6.10), pari rispettivamente a media e deviazione standard della v.a. (vedi § 6.7.1), e che descrivono completamente la d.d.p. dal punto di vista analitico: pertanto la stima di m_x e σ_x (a partire da un buon numero di realizzazioni¹⁹) è sufficiente per descrivere completamente il fenomeno aleatorio. La v.a. gaussiana emerge in molti fenomeni naturali, ed è dimostrabile analiticamente che la sua densità è tipica²⁰ per grandezze ottenute dalla somma di un numero molto elevato di cause aleatorie, tutte statisticamente indipendenti e con la medesima d.d.p.²¹ (*teorema centrale del limite*²²).

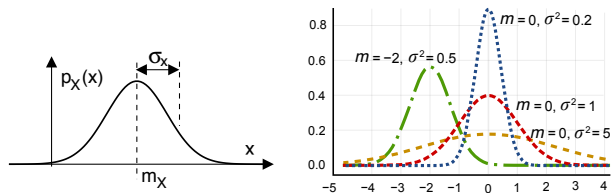
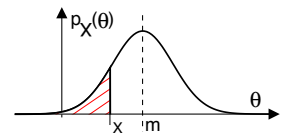


Figura 6.1: Grafico di una d.d.p. gaussiana

La v.a. gaussiana emerge in molti fenomeni naturali, ed è dimostrabile analiticamente che la sua densità è tipica²⁰ per grandezze ottenute dalla somma di un numero molto elevato di cause aleatorie, tutte statisticamente indipendenti e con la medesima d.d.p.²¹ (*teorema centrale del limite*²²).

6.2.4.1 Probabilità di un evento gaussiano

Accade che il valore dell'integrale $\int_{-\infty}^x \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(\theta-m)^2}{2\sigma^2}} d\theta$ mostrato in figura e che corrisponde alla funzione di distribuzione $F_X(x)$ della v.a. gaussiana x non sia esprimibile in forma chiusa, e dunque per $F_X(x)$ non esiste una formula precisa. Al contrario, il suo valore viene *calcolato* per via numerica²³, e reso disponibile mediante tabelle e grafici. Per evitare di dover ripetere il calcolo per ogni possibile valore di media e varianza, l'estensione dell'area tratteggiata viene valutata per una v.a. gaussiana *normalizzata* Z a media nulla e varianza $\frac{1}{2}$, ed espressa nei termini della



¹⁹Disponendo di un insieme $\{x_n\}$ di N realizzazioni di una variabile aleatoria X , possiamo effettuare le stime $\widehat{m}_x = \frac{1}{N} \sum_{n=1}^N x_n$ e $\widehat{m}_x^{(2)} = \frac{1}{N} \sum_{n=1}^N x_n^2$, il cui valore tende *asintoticamente* a quello delle rispettive medie di insieme, come N (la dimensione del campione statistico) tende a ∞ . Al proposito, vedi § 6.6.3.1.

²⁰Tanto che la (6.10) è anche detta *Normale*, e per questo è indicata anche come $N(m, \sigma^2)$.

²¹Questa condizione è anche detta di v.a. *indipendenti e identicamente distribuite*, ovvero *i.i.d.*

²²Il teorema viene dimostrato al § 6.7.2, ma può essere divertente ed utile sperimentarne la validità ricorrendo alla applet presente presso

<http://www.randomservices.org/random/apps/DiceExperiment.html>

Inoltre, considerando che al § 6.2.5 si mostra come la d.d.p. di una somma di v.a. indipendenti sia pari alla convoluzione tra le rispettive d.d.p., osserviamo che la convoluzione ripetuta di una stessa d.d.p. con se stessa, la *gaussianizza*.

²³Ovvero mediante del software che implementa uno dei metodi descritti ad es. presso

https://en.wikipedia.org/wiki/Numerical_integration

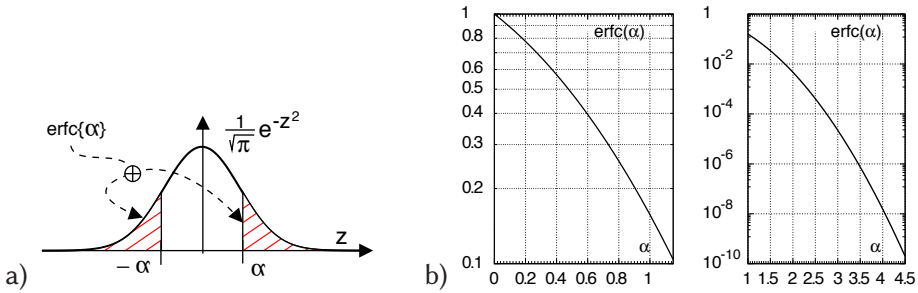


Figura 6.2: Valore di $\text{erfc} \{ \alpha \}$ per una gaussiana a media nulla e varianza $1/2$

funzione $\text{erfc} \{ \alpha \}$ ²⁴ definita come rappresentato in fig. 6.2-a), ovvero

$$\text{erfc} \{ \alpha \} = Pr \{ |z| > \alpha \} = 2 \int_{\alpha}^{\infty} \frac{1}{\sqrt{\pi}} e^{-z^2} dz$$

il cui andamento è graficato in fig. 6.2-b) per i diversi valori dell'argomento $\alpha \geq 0$. In questi termini, la funzione di distribuzione di z si ottiene come

$$F_Z(z) = Pr \{ Z \leq z \} = \begin{cases} 1 - \frac{1}{2} \text{erfc} \{ z \} & z > 0 \\ \frac{1}{2} \text{erfc} \{ -z \} & z \leq 0 \end{cases}$$

Tale risultato può quindi essere usato per calcolare il valore di probabilità $Pr \{ X > \beta \}$ con cui una v.a. X con media m e varianza σ^2 supera una soglia $\beta \geq m$, applicando il cambio di variabile $z = \frac{x-m}{\sqrt{2\sigma}}$, che fornisce

$$Pr \{ X > \beta \} = \frac{1}{2} \text{erfc} \left\{ \frac{\beta - m}{\sqrt{2\sigma}} \right\} \tag{6.11}$$

Esempio Valutare la probabilità che una v.a. gaussiana X con $m = 2$ e $\sigma^2 = 4$ superi il valore $x = 2.5$. Il cambio di variabile $\frac{x-m}{\sqrt{2\sigma}}$ determina per la v.a. normalizzata il nuovo valore di soglia $\frac{2.5-2}{1.41 \cdot 2} \simeq 0.17$, e dalla fig. 6.2-b) si ottiene $Pr \{ X > \beta \} = \frac{1}{2} \text{erfc} \{ 0.17 \} \simeq \frac{1}{2} 0.82 = 0.41$.

Ma proviamo a svolgere i calcoli: il valore di probabilità richiesto dall'esercizio corrisponde a $Pr \{ X > \beta \} = \int_x^{\infty} \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(\theta-m)^2}{2\sigma^2}} d\theta$; ponendo $\frac{\theta-m}{\sqrt{2\sigma}} = \eta$ risulta $d\theta = \sqrt{2\sigma} \cdot d\eta$ mentre l'estremo inferiore di integrazione diviene $\eta = \frac{x-m}{\sqrt{2\sigma}}$, ottenendo così

$$Pr \{ X > \beta \} = \int_{\frac{x-m}{\sqrt{2\sigma}}}^{\infty} \frac{1}{\sqrt{2\pi\sigma}} e^{-\eta^2} \sqrt{2\sigma} \cdot d\eta = \int_{\frac{x-m}{\sqrt{2\sigma}}}^{\infty} \frac{1}{\sqrt{\pi}} e^{-\eta^2} d\eta = \frac{1}{2} \text{erfc} \left\{ \frac{x - m}{\sqrt{2\sigma}} \right\}$$

Questo risultato tornerà utile al § 15.4, quando dovremo valutare la probabilità di errore nelle trasmissioni numeriche.

La funzione Q Alcuni esprimono la probabilità di evento gaussiano come $Q \{ x \} = Pr \{ X > x \} = \int_x^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{\theta^2}{2}} d\theta$, riferita dunque ad una sola coda di una v.a. gaussiana a

²⁴Il termine *erfc* sta per *funzione di errore complementare*, e trae origine dai risultati della misura di grandezze fisiche, in cui l'errore di misura, dipendente da cause molteplici, si assume appunto gaussiano. Vedi anche https://it.wikipedia.org/wiki/Funzione_degli_errori.

media nulla e varianza *unitaria*. Tra le due notazioni sussiste pertanto la relazione $Q\{x\} = \frac{1}{2}\operatorname{erfc}\left\{\frac{x}{\sqrt{2}}\right\}$.

Valori notevoli Per avere una idea della rapidità di azzeramento della campana gaussiana, può essere utile tenere conto che in un intervallo di estensione 2σ centrato attorno alla media si trova il 68,3% della probabilità, che sale al 95,5% per un intervallo che si estende per $\pm 2\sigma$ attorno alla media, ed arriva al 99,7% per un intervallo $\pm 3\sigma$. Vedi anche la tabella a pag. 176.

Mentre ai §§ 6.5 e 6.7.1 sono approfonditi ulteriori aspetti della v.a. gaussiana, altri tipi di v.a. sono descritti assieme ai rispettivi casi di utilizzo, come nel caso della v.a. esponenziale § 22.2.1, di Bernoulli § 22.1, Poisson § 22.2, Rayleigh e Rice pag. 423. Qui invece il capitolo prosegue introducendo una descrizione di v.a. *complementare* a quella fornita da d.d.p. e distribuzione, mediante la quale dimostrare il comportamento di una somma di v.a. indipendenti.

6.2.5 Funzione caratteristica

La funzione caratteristica $\Phi_X(\omega)$ di una v.a. X è definita come l'antitrasformata di Fourier della sua densità di probabilità, ovvero (equivalentemente) come il valore atteso di $e^{j\omega x}$:

$$\Phi_X(\omega) = \mathcal{F}^{-1}\{p_X(x)\} = E_X\{e^{j\omega x}\} = \int p_X(x) e^{j\omega x} dx \quad (6.12)$$

Intuitivamente, possiamo pensare che si sia scelta l'anti-trasformata anziché la trasformata in quanto una d.d.p. è una *densità* (di probabilità), similmente ad una densità *spettrale*. Tra una d.d.p. $p_X(x)$ e la relativa $\Phi_X(\omega)$ intercorre una relazione *biunivoca*, nel senso che se due d.d.p. hanno la stessa $\Phi_X(\omega)$, esse coincidono. Affrontiamo subito due importanti applicazioni di questo nuovo strumento.

6.2.5.1 Densità di probabilità della somma di v.a. indipendenti

Osserviamo che, se $z = x + y$ è la somma di v.a. *indipendenti*, per la sua funzione caratteristica si ottiene

$$\begin{aligned} \Phi_z(t) &= E_Z\{e^{j(x+y)t}\} = E_Z\{e^{jxt} e^{jyt}\} = E_X\{e^{jxt}\} E_Y\{e^{jyt}\} \\ &= \Phi_x(t) \Phi_y(t) \end{aligned}$$

in quanto sotto tale ipotesi la d.d.p. congiunta $p_{XY}(x, y)$ si fattorizza nel prodotto delle d.d.p. marginali $p_X(x)$ e $p_Y(y)$ (vedi § 6.1.5), ed il valore atteso si scompone nel prodotto di due integrali. Pertanto, la funzione caratteristica di una somma di v.a. indipendenti è pari al prodotto delle funzioni caratteristiche.

Effettuando ora l'operazione inversa (trasformata di Fourier della funzione caratteristica della somma), e ricordando che ad un prodotto in un dominio corrisponde una convoluzione nell'altro, si ottiene il risultato $p_Z(z) = \mathcal{F}\{\Phi_z(t)\} = \mathcal{F}\{\Phi_x(t) \Phi_y(t)\} = p_X(x) * p_Y(y)$ che ci permette di enunciare:

La densità di probabilità della somma di v.a. indipendenti è pari alla convoluzione tra le rispettive densità di probabilità marginali.

Esempio Se x ed y sono due v.a. a distribuzione uniforme tra $\pm\Delta$, la loro somma ha densità di probabilità triangolare con base 2Δ . Pertanto, nel lancio di 2 dadi il risultato più probabile è 7. Infatti può essere ottenuto come $6+1, 5+2, 4+3, 3+4, 2+5, 1+6$, ovvero in 6 modi diversi, ognuno con probabilità $\frac{1}{6} \cdot \frac{1}{6} = \frac{1}{36}$ e dunque $Pr\{7\} = 6 \frac{1}{36} = \frac{1}{6}$.

6.2.5.2 Funzione caratteristica di una v.a. gaussiana

Consideriamo il caso di una v.a. gaussiana a valor medio nullo e varianza σ^2 : si tratta di eseguire il calcolo $\Phi_X(\omega) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{x^2}{2\sigma^2}} e^{j\omega x} dx$, il cui svolgimento²⁵ porta al risultato

$$\Phi_X(\omega) = e^{-\frac{1}{2}\omega^2\sigma^2} \tag{6.13}$$

ovvero ancora un andamento gaussiano, con dispersione (varianza) inversamente proporzionale a quella della gaussiana di partenza. Qualora la v.a. abbia invece valore medio $m_x \neq 0$ la proprietà di traslazione della \mathcal{F} -trasformata fornisce $\Phi_X(\omega) = e^{-\frac{1}{2}\omega^2\sigma^2} \cdot e^{j\omega m_x}$.

Ma per non farci mancare nulla, citiamo ulteriori proprietà della funzione caratteristica:

Valore massimo Osserviamo che

$$|\Phi_X(\omega)| = |E_X\{e^{j\omega x}\}| \leq E_X\{|e^{j\omega x}|\} = \int p_X(x) dx = \Phi_X(\omega)|_{\omega=0} = 1$$

dunque $\Phi_X(\omega)$ ha un massimo nell'origine.

Relazione con i momenti Se $\Phi_X(\omega)$ è derivabile k volte, dalla (6.12) si ottiene $\frac{d^k \Phi_X(\omega)}{d\omega^k} = E\{(jx)^k e^{j\omega x}\}$ che calcolata per $\omega = 0$ fornisce

$$\left. \frac{d^k \Phi_X(\omega)}{d\omega^k} \right|_{\omega=0} = j^k E\{x^k\} = j^k m_x^{(k)}$$

Pertanto conoscendo i primi n momenti $m_x^{(1)}, m_x^{(2)}, \dots, m_x^{(n)}$ della v.a. x è possibile ottenere una approssimazione $\hat{\Phi}_X(\omega)$ della relativa funzione caratteristica $\Phi_X(\omega)$ nella forma di una espansione in serie di potenze, ovvero

$$\Phi_X(\omega) \simeq \sum_{k=0}^n \left. \frac{d^k \Phi_X(\omega)}{d\omega^k} \right|_{\omega=0} \frac{\omega^k}{k!} = 1 + \sum_{k=1}^n m_x^{(k)} \frac{(j\omega)^k}{k!}$$

²⁵Ricordando che $e^a e^b = e^{a+b}$ possiamo scrivere $\Phi_X(\omega) = \frac{1}{\sqrt{2\pi\sigma}} \int_{-\infty}^{\infty} e^{j\omega x - \frac{x^2}{2\sigma^2}} dx$; riformuliamo quindi l'esponente $j\omega x - \frac{x^2}{2\sigma^2}$ come $-\frac{x^2}{2\sigma^2} + j\omega x = -\frac{(j\omega\sigma)^2}{2} + \frac{(j\omega\sigma)^2}{2} - \frac{1}{2} \left(\frac{x^2}{\sigma^2} - 2j\omega x + (j\omega\sigma)^2 \right) + \frac{(j\omega\sigma)^2}{2} = -\frac{1}{2} \left(\frac{x}{\sigma} - j\omega\sigma \right)^2 + \frac{(j\omega\sigma)^2}{2}$ in modo da ottenere

$$\Phi_X(\omega) = e^{\frac{1}{2}(j\omega\sigma)^2} \frac{1}{\sqrt{2\pi\sigma}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(\frac{x}{\sigma} - j\omega\sigma)^2} dx = e^{-\frac{1}{2}(\omega\sigma)^2} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(y)^2} dy = e^{-\frac{1}{2}\omega^2\sigma^2}$$

avendo effettuato il cambio di variabile $y = \frac{x}{\sigma} - j\omega\sigma$ che da luogo agli stessi estremi di integrazione, mentre $dx = \sigma dy$, ed avendo notato come l'integrale ora calcoli l'area di una gaussiana con varianza unitaria, pari a ad uno.

Inversione Conoscendo una stima $\hat{\Phi}_X(\omega)$ della f.c. della v.a. x si può ottenere una approssimazione $\hat{p}_X(x)$ della relativa d.d.p. calcolandone la trasformata di Fourier, ovvero

$$\hat{p}_X(x) = \frac{1}{2\pi} \int \hat{\Phi}_X(\omega) e^{-j\omega x} d\omega$$

Estendiamo ora i concetti fin qui esposti al caso di v.a. vettoriali o multivariate, le cui realizzazioni corrispondono ad una n -upla di valori.

6.2.6 Variabile aleatoria multivariata

In questo caso la v.a. rappresenta congiuntamente un intero *vettore* \mathbf{x} di variabili aleatorie monodimensionali, ossia una loro collezione ordinata, in numero finito (ad es. N), in relazione o meno tra loro in base a legami di tipo probabilistico.

Densità di probabilità Indicando con \mathbf{X} la v.a. vettoriale, e con \mathbf{x} una sua realizzazione costituita dalle N componenti x_1, x_2, \dots, x_N , la v.a. multivariata è descritta per mezzo della d.d.p. $p_X(\mathbf{x}) = p_X(x_1, x_2, \dots, x_N)$ funzione di N variabili, per la quale deve risultare

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} p_X(x_1, x_2, \dots, x_N) dx_1 dx_2 \dots dx_N = 1$$

Distribuzione Anche nel caso multivariato può essere definita una funzione di distribuzione $F_X(\mathbf{x})$, anch'essa N -dimensionale, il cui valore $F_X(\bar{\mathbf{x}}) = Pr\{\mathbf{x} \leq \bar{\mathbf{x}}\}$ nel punto $\bar{\mathbf{x}} = (\bar{x}_1, \bar{x}_2, \dots, \bar{x}_N)$ si calcola come

$$F_X(\bar{\mathbf{x}}) = \int_{-\infty}^{\bar{x}_1} \int_{-\infty}^{\bar{x}_2} \dots \int_{-\infty}^{\bar{x}_N} p_X(x_1, x_2, \dots, x_N) dx_1 dx_2 \dots dx_N$$

Densità di probabilità marginale La d.d.p. *marginale* $p_{X_i}(x_i)$ della singola v.a. *monodimensionale* x_i che prende parte al sistema di coordinate su cui \mathbf{X} è definita, può essere calcolata a partire dalla d.d.p. *congiunta* $p_X(\mathbf{x})$ mediante *saturazione* delle altre v.a., ovvero

$$p_{X_i}(x_i) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} p_X(x_1, x_2, \dots, x_N) dx_1 \dots dx_{i-1} dx_{i+1} \dots dx_N \quad (6.14)$$

$N-1$ integrali

Densità di probabilità condizionata La d.d.p. di un sotto-gruppo di v.a. $\mathbf{x}_a = (x_1, x_2, \dots, x_a)$, qualora il valore delle restanti coordinate $\mathbf{x}_b = (x_{a+1}, x_{a+2}, \dots, x_N)$ di \mathbf{x} sia da ritenersi noto, si ottiene dividendo la d.d.p. congiunta $p_X(\mathbf{x})$ per quella marginale $p_X(\mathbf{x}_b)$ che descrive gli eventi condizionanti, ovvero

$$p_X(\mathbf{x}_a/\mathbf{x}_b) = \frac{p_X(\mathbf{x})}{p_X(\mathbf{x}_b)}$$

in cui $p_X(\mathbf{x}_b)$ è ottenuta per saturazione (6.14). La separazione *ordinale* tra i due gruppi di variabili ha lo scopo di semplificare la notazione di questa definizione; in realtà, le v.a. dei due gruppi possono essere prese con un ordine qualsiasi.

Valore atteso e momenti Nel caso in cui si tratti del valore atteso di una funzione di una sola v.a. marginale, si utilizza ancora la (6.6) in cui la d.d.p. è quella marginale $p_{X_i}(x_i)$ relativa alla v.a. rispetto alla quale si sta eseguendo la media di insieme. Per questa via è possibile ottenere un vettore $m_X = (m_{x_1}, m_{x_2}, \dots, m_{x_N})$ che rappresenta il valor medio della v.a. multivariata \mathbf{X} , le cui componenti m_{x_i} sono i momenti di primo ordine delle v.a. marginali, ovvero

$$m_{x_i} = E_X \{x_i\} = \int x_i p_{X_i}(x_i) dx_i$$

Notiamo che sebbene la notazione $E_X \{x_i\}$ indichi un valore atteso calcolato rispetto alla variabilità di tutte le componenti della v.a. multivariata \mathbf{X} , il calcolo è svolto ricorrendo alla d.d.p. marginale, in quanto le altre v.a. x_j con $j \neq i$ saturano (6.14).

D'altra parte, è possibile ora valutare anche i cosiddetti momenti *misti*, in cui cioè la media di insieme considera tutti i possibili valori di due o più componenti di \mathbf{X} , pesando ognuno di questi con il relativo valore di probabilità. Ad esempio, un momento misto di ordine (n, m) è definito come

$$m_{x_i x_j}^{(n, m)} = E_X \{x_i^n x_j^m\} = \int \int x_i^n x_j^m p_{X_i X_j}(x_i, x_j) dx_i dx_j$$

ed un momento misto *centrato* di ordine (n, m) come

$$\mu_{x_i x_j}^{(n, m)} = E_X \{(x_i - m_{x_i})^n (x_j - m_{x_j})^m\} = \int \int (x_i - m_{x_i})^n (x_j - m_{x_j})^m p_{X_i X_j}(x_i, x_j) dx_i dx_j$$

in cui la $p_{X_i X_j}(x_i, x_j)$ bidimensionale è ottenuta saturando la $p_X(\mathbf{x})$ sulle dimensioni diverse da i e j .

E' ora il turno di estendere i concetti probabilistici allo spazio dei segnali.

6.3 Processi stazionari ed ergodici

Dopo aver descritto come caratterizzare statisticamente i valori di v.a. singole o vettoriali, occupiamoci del caso in cui si voglia descrivere da un punto di vista probabilistico *un intero segnale*, la cui reale identità non sia nota a priori²⁶.

Un segnale siffatto viene detto *membro* (o realizzazione) di un *processo aleatorio*, e può essere indicato come $x(t, \theta)$, mediante una descrizione formale che prevede una coppia di insiemi: il primo di questi è l'insieme \mathcal{T} degli istanti temporali (tipicamente entro un intervallo) su cui sono definiti i membri del processo, mentre il secondo è relativo ad una variabile aleatoria Θ , i cui valori θ identificano ognuno una particolare realizzazione del processo. Pertanto, una specifica realizzazione θ_i della v.a. Θ *indicizza* il processo, i cui membri $x(t, \theta_i)$, con $t \in \mathcal{T}$, sono noti solo dopo la conoscenza di $\theta_i \in \Theta$ (²⁷). Il processo aleatorio è quindi definito come l'insieme dei segnali $\{x(t, \theta)\}$,

²⁶Chiaramente, la maggioranza dei segnali trasmessi da apparati di TLC sono di questo tipo.

²⁷Per fissare le idee, conduciamo parallelamente al testo un esempio "reale" in cui il processo aleatorio è costituito da.... la selezione musicale svolta da un DJ. L'insieme \mathcal{T} sarà allora costituito dall'orario di apertura delle discoteche (dalle 22 all'alba ?), mentre in θ faremo ricadere tutte le caratteristiche di variabilità (umore del dj, i dischi che ha in valigia, la discoteca in cui ci troviamo, il giorno della

con $t \in \mathcal{T}$ e $\theta \in \Theta$.

Se viceversa fissiamo un particolare istante temporale t_j , il valore $x(t_j, \theta)$ è una variabile aleatoria, la cui realizzazione dipende da quella di $\theta \in \Theta$; pertanto, è definita la densità $p_X(x(t_j))$ (indipendente da θ), che possiamo disegnare in corrispondenza dell'istante t_j in cui è prelevato il campione²⁸; a tale riguardo, si faccia riferimento alla figura 6.3, che mostra le densità di probabilità definite e riferite a membri di un processo.

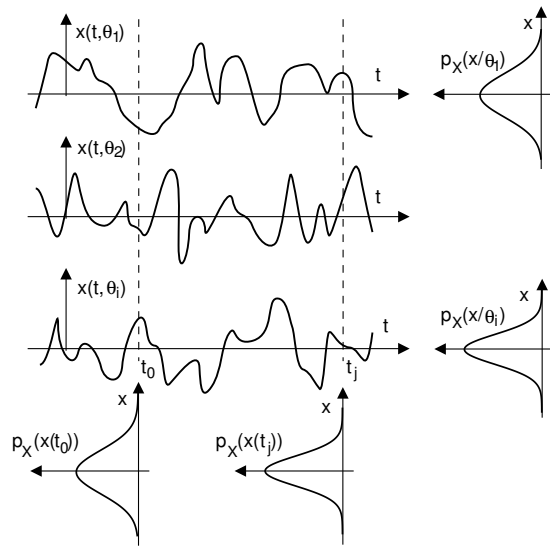


Figura 6.3: Un processo non ergodico

6.3.1 Momento come media di insieme

Consiste nel *valore atteso* di una *potenza n-esima* dei valori del segnale, eseguito rispetto alla variabilità dovuta a Θ , ed è pertanto calcolata come

$$m_X^{(n)}(t_j) = E_{\Theta} \{x^n(t_j, \theta)\} = \int_{-\infty}^{\infty} x^n(t_j, \theta) p_{\Theta}(\theta) d\theta = \int_{-\infty}^{\infty} x^n p_X(x(t_j)) dx$$

in cui l'ultima eguaglianza indica come la variabilità statistica di x^n sia completamente descritta dalla d.d.p. $p_X(x(t_j))$ di $x(t_j, \theta)$ al variare di $\theta \in \Theta$, mostrata in basso in fig. 6.3. Notiamo che secondo questo approccio, la media di insieme dipende dall'istante t_j in cui è prelevato un valore²⁹.

6.3.2 Media temporale

In alternativa, possiamo fissare una particolare realizzazione θ_i di Θ , e quindi concentrare l'attenzione su di un singolo membro $x(t, \theta_i)$, che è ora un segnale *certo*³⁰: per esso possono quindi essere calcolate le *medie temporali*, indicate con una linea sopra alla quantità di cui si calcola la media $\overline{(\cdot)}$:

$$\overline{x^n(t, \theta_i)} = \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} x^n(t, \theta_i) dt$$

settimana...).

²⁸Nell'esempio, $x(t_0, \theta)$ è il valore di pressione sonora rilevabile ad un determinato istante (es. le 23.30) al variare di θ (qualunque dj, discoteca, giorno...).

²⁹Ad esempio, se in tutte le serate il volume aumenta progressivamente nel tempo, la $p_X(x(t_j))$ si allargherà per t_j crescenti.

³⁰ $x(t, \theta_i)$ rappresenta, nel nostro esempio, l'intera selezione musicale (detta *serata*) proposta da un ben preciso dj, in un preciso locale, un giorno ben preciso.

In particolare, troviamo il *valore medio* (pag. 14)

$$\overline{x(t, \theta_i)} = \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} x(t, \theta_i) dt$$

e la *potenza*³¹ (eq. (1.1)) (o *media quadratica*)

$$\overline{x^2(t, \theta_i)} = \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} x^2(t, \theta_i) dt$$

Notiamo che una generica media temporale:

- non dipende dal tempo;
- è una variabile aleatoria, in quanto dipende dalla realizzazione di Θ .

6.3.3 Media temporale calcolata come media di insieme

L'estrazione da $x(t, \theta_i)$ di un valore ad un istante casuale $t \in \mathcal{T}$ definisce una ulteriore variabile aleatoria, descritta dalla densità di probabilità (condizionata) $p_X(x/\theta_i)$, che disegniamo a fianco dei singoli membri mostrati in fig. 6.3. Qualora la $p_X(x/\theta_i)$ sia nota, le medie temporali di ordine n possono essere calcolate (per quel membro) come i rispettivi momenti:

$$\overline{x^n(t, \theta_i)} = \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} x^n(t, \theta_i) dt = \int_{-\infty}^{\infty} x^n p_X(x/\theta_i) dx = E_{X/\Theta=\theta_i} \{x^n\} = m_X^{(n)}(\theta_i)$$

Ciò equivale infatti ad effettuare una media ponderata, in cui ogni possibile valore di x è pesato per la sua probabilità $p_X(x/\theta_i) dx$ (vedi l'esempio a pag. 151).

6.3.4 Processo stazionario

Qualora $p_X(x(t_j))$ non dipenda da t_j , ma risulti $p_X(x(t_j)) = p_X^{\mathcal{T}}(x)$ per qualsiasi $t_j \in \mathcal{T}$, il processo $\{x(t, \theta)\}$ è detto stazionario³² *in senso stretto*. In tal caso tutte le medie di insieme non dipendono più dal tempo, ossia $m_X^{(n)}(t) = m_X^{(n)}$ per $\forall t \in \mathcal{T}$, e le $p_X(x(t_j))$ in basso in fig. 6.3 sono tutte uguali.

Se invece sono solamente le prime due medie di insieme $m_X(t)$ e $m_X^{(2)}(t)$ a non dipendere da t , il processo $\{x(t, \theta)\}$ è detto stazionario *in media* ed *in media quadratica*, od anche stazionario *in senso lato*³³. Nel caso di un processo gaussiano (§ 6.5.3), la stazionarietà in senso lato implica quella in senso stretto³⁴.

³¹ $m_X^{(2)}(\theta_i)$ in questo caso rappresenta la potenza media con cui è suonata la musica nella particolare serata θ_i .

³²La "serata in discoteca" stazionaria si verifica pertanto se non mutano nel tempo il genere di musica, il volume dell'amplificazione... o meglio se eventuali variazioni in alcune particolari discoteche-realizzazioni sono compensate da variazioni opposte in altrettanti differenti membri del processo.

³³In questo caso la $p_X(x(t))$ non è nota, oppure non è stazionaria, ma le maggiori applicazioni della proprietà di stazionarietà dipendono solo da $m_X(t)$ e $m_X^{(2)}(t)$, che possono essere misurati (o per meglio dire *stimati*, vedi § 6.6.3.1), e risultare stazionari anche se $p_X(x(t))$ non lo è.

³⁴Infatti la d.d.p. gaussiana è completamente definita qualora siano noti i valori di media e (co)varianza, vedi §§ 6.2.4 e 6.5.

Supponiamo ora di suddividere il membro $x(t, \theta_i)$ in più intervalli temporali, e di calcolare per ciascuno di essi le medie temporali, limitatamente al relativo intervallo. Nel caso in cui queste risultino uguali tra loro, e di conseguenza uguali alla media temporale $m_X^{(n)}(\theta_i)$, il membro è (individualmente) stazionario³⁵. Ovviamente, se tutti i membri sono individualmente stazionari, lo è anche il processo a cui appartengono.

6.3.5 Processo stazionario ed ergodico

Questa importante sottoclasse di processi stazionari identifica la circostanza che *ogni membro del processo è statisticamente rappresentativo di tutti gli altri*. Ciò si verifica quando la densità di probabilità (a destra in fig. 6.3) dei valori estratti da un singolo membro $p_X(x/\theta_i)$ è sempre la stessa, indipendentemente dal particolare θ_i , ottenendo in definitiva $p_X(x/\theta_i) = p_X^\ominus(x)$ indipendentemente dalla realizzazione e, per la stazionarietà, anche $p_X(x/t_j) = p_X^\top(x)$, e dunque $p_X^\ominus(x) = p_X^\top(x) = p_X(x)$. In questo caso le medie temporali $m_X^{(n)}(\theta_i)$, calcolabili come momenti sulla singola realizzazione come illustrato al § 6.3.3, sono identiche per tutti i membri³⁶ θ_i , ed identiche anche alle medie di insieme $m_X^{(n)}(t_j)$ calcolate per un qualunque istante. Enunciamo pertanto la definizione:

Un processo stazionario è ergodico se la media temporale calcolata su di una qualunque realizzazione del processo, coincide con la media di insieme relativa ad una variabile aleatoria estratta ad un istante qualsiasi (per la stazionarietà) dall'insieme dei suoi membri.

Esempio: la potenza di segnale Mostriamo come il calcolo della potenza di un membro di un processo ergodico sia equivalente a quello del momento di 2° ordine del processo:

$$\begin{aligned} \mathcal{P}_X(\theta) &= \overline{x^2(\theta)} = \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} x^2(t, \theta) dt = \int_{-\infty}^{\infty} x^2 p_X(x/\theta) dx = \\ &= \int_{-\infty}^{\infty} x^2 p_X(x) dx = m_X^{(2)} = E\{x^2\} = \mathcal{P}_X \end{aligned}$$

Questo risultato mostra come sia possibile calcolare la potenza di una realizzazione di un processo, senza conoscere la forma d'onda dei suoi membri.

Esempio: il valore medio A pag. 14 è stato definito come $\bar{x} = \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} x(t) dt$, ovvero come una media *temporale* del primo ordine. Qualora $x(t)$ sia membro di un processo ergodico, tale valore può esser calcolato anche come *valore atteso* di $x(t)$, ovvero momento di *primo ordine* m_X della v.a. x estratta dal processo:

$$\begin{aligned} \bar{x}(\theta) &= \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} x(t, \theta) dt = \int_{-\infty}^{\infty} x p_X(x/\theta) dx = \\ &= \int_{-\infty}^{\infty} x p_X(x) dx = E\{x\} = m_X \end{aligned}$$

³⁵Questo accade se la selezione musicale di una particolare serata si mantiene costante (es. solo raggamuffin) oppure variata ma in modo omogeneo (es. senza tre "lenti" di fila).

³⁶Volendo pertanto giungere alla definizione di una serata *ergodica* in discoteca, dovremmo eliminare quei casi che, anche se individualmente stazionari, sono decisamente "fuori standard" (tutto metal, solo liscio...).

Potenza, varianza, media quadratica e valore efficace In particolare osserviamo che in base alla (6.9) possiamo scrivere

$$\mathcal{P}_X = m_X^{(2)} = \sigma_x^2 + (m_x)^2 \quad (6.15)$$

e per i segnali a media nulla ($m_x = 0$) si ottiene $\mathcal{P}_X = \sigma_x^2$; in tal caso il valore efficace (pag. 15) $\sqrt{\mathcal{P}_X}$ coincide con la deviazione standard σ_x . La radice della potenza è inoltre spesso indicata come *valore RMS* (ROOT MEAN SQUARE), definito come $x_{RMS} = \sqrt{\overline{x^2(t)}}$, ovvero la radice della *media quadratica* (nel tempo). Se il segnale è a media nulla, x_{RMS} coincide quindi con il valore efficace; se $x(t)$ è membro di un processo ergodico a media nulla, x_{RMS} coincide con la deviazione standard.

6.3.6 Riassumendo

- Se un processo è ergodico, è anche stazionario, ma non il viceversa. Esempio: se $x(t, \theta) = C_\theta$ pari ad una costante (aleatoria), allora è senz'altro stazionario, ma $p_X(x/\theta) = \delta(x - C_\theta)$, e quindi non ergodico.
- Se un processo è ergodico è possibile:
 - calcolare le medie di insieme in forma di medie temporali a partire da una singola realizzazione *oppure*
 - ottenere le medie temporali di una qualunque realizzazione a partire dalle medie di insieme, disponendo della statistica $p_X(x)$, e *anche*
 - stimare la d.d.p. a partire dall'istogramma dei valori estratti da un qualunque membro.
- Se l'eguaglianza tra medie di insieme e temporali sussiste solo fino ad un determinato ordine e non oltre, il processo *non* è ergodico *in senso stretto*. Per ciò che concerne le telecomunicazioni, è spesso sufficiente la proprietà di ergodicità *in senso lato*, ovvero limitata al 2° ordine, che garantisce $\overline{x(t)} = E\{x\} = m_x$; $\overline{x^2(t)} = E\{x^2\} = m_x^{(2)}$.

6.3.7 Processo ad aleatorietà parametrica

A volte può convenire pensare un segnale certo come rappresentante di una intera classe di segnali che definiscono un processo ergodico, in modo da poter calcolare le medie temporali che lo riguardano mediante delle medie di insieme. In tal caso si rientra nella categoria di processi *parametrici* $\{x(t, \theta)\}$ in cui la v.a. θ compare in modo *esplicito* nella espressione analitica dei segnali membri. Ad esempio, il segnale periodico

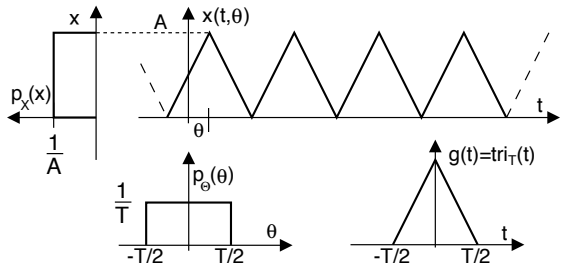
$$x(t, \theta) = \sum_{n=-\infty}^{\infty} A \cdot g_T(t - \theta - nT) \quad (6.16)$$

rappresentato alla figura seguente, ha come parametro aleatorio un ritardo θ , che ne rende imprecisata la fase iniziale. Se θ è (come in figura) una v.a. a distribuzione uniforme tra $-T/2$ e $T/2$ (ovvero $p_\Theta(\theta) = \frac{1}{T} \text{rect}_T(\theta)$), allora il processo (6.16) risulta

stazionario ed ergodico. Infatti scegliendo ad esempio una $g(t) = \text{tri}_T(t)$ la d.d.p. per una v.a. estratta dal processo ad un istante qualunque diviene pari a³⁷

$$p_X(x) = \frac{1}{A} \text{rect}_A(x - \frac{A}{2}) \quad (6.17)$$

uguale cioè alla d.d.p. ottenibile estraendo una v.a. da un membro qualunque.



Esercizio Possiamo verificare la coincidenza tra medie temporali e di insieme, osservando che il valor medio $m_X = E\{x\}$ di un qualunque membro di (6.16) è pari alla media temporale $\frac{A}{2}$, la varianza è pari a quella della d.d.p. uniforme $\sigma_X^2 = \frac{A^2}{12}$ (§ 6.2.3), e la potenza vale³⁸

$$\mathcal{P}_X = \sigma_X^2 + m_X^2 = \frac{A^2}{12} + \frac{A^2}{4} = \frac{4A^2}{12} = \frac{A^2}{3} \quad (6.18)$$

Se la $p_\theta(\theta)$ fosse stata diversa, il processo avrebbe perso stazionarietà e quindi ergodicità. Infatti, ponendo ad esempio $p_\theta(\theta) = \frac{2}{T} \text{rect}_{T/2}(\theta)$ e volendo ottenere una media di insieme considerando i possibili membri del processo nell'intervallo temporale $-T/4 < t < T/4$, tutte le realizzazioni avrebbero valori maggiori del valor medio $\frac{A}{2}$.

Processo armonico Si tratta di un processo ad aleatorietà parametrica, i cui membri hanno espressione

$$x(t, \theta) = A \cos(2\pi f_0 t + \theta)$$

dove θ è una v.a. uniforme con d.d.p. $p_\theta(\theta) = \frac{1}{2\pi} \text{rect}_{2\pi}(\theta)$. In tal caso il processo è stazionario ed ergodico, ed a pag. 166 si dimostra che un valore estratto a caso da un membro qualsiasi è una v.a. con d.d.p.

$$p_X(x) = \frac{1}{\pi \sqrt{A^2 - x^2}} \quad (6.19)$$

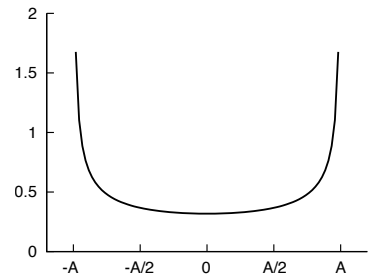


Figura 6.4: Densità di prob. per un processo armonico

mostrata in figura 6.4, la cui sagoma è detta *a vasca da bagno* (BATH TUBE).

³⁷La (6.17) non è frutto di un calcolo, bensì di un ragionamento: l'impulso $g_T(t)$ triangolare non "passa più tempo" su di un valore o su di un altro, ma passa *lo stesso* tempo su un qualunque valore tra 0 ed A. Pertanto i diversi membri del processo, ognuno relativo ad un diverso θ , qualora valutati ad un medesimo istante t , assumono uno qualsiasi dei valori tra 0 ed A con d.d.p. uniforme.

³⁸Verifichiamo per esercizio che il valore (6.18) corrisponda a quello calcolato come media temporale. Calcoliamo innanzitutto l'energia \mathcal{E}_g di $g(t)$:

$$\begin{aligned} \mathcal{E}_g &= 2 \int_0^{-T/2} [g(t)]^2 dt = 2 \int_0^{-T/2} \left[1 - \frac{2t}{T}\right]^2 dt = 2 \int_0^{-T/2} \left[1 + \frac{4t^2}{T^2} - \frac{4t}{T}\right]^2 dt = \\ &= 2 \left(\frac{T}{2} + \frac{4}{T^2} \frac{t^3}{3} \Big|_0^{-T/2} - \frac{4}{T} \frac{t^2}{2} \Big|_0^{-T/2} \right) = T + \frac{8}{T^2} \frac{T^3}{3 \cdot 8} - \frac{8}{T} \frac{T^2}{2 \cdot 4} = T + \frac{T}{3} - T = \frac{T}{3} \end{aligned}$$

da cui la potenza di $x(t)$ si ottiene come $\mathcal{P}_X = A^2 \mathcal{E}_g / T = \frac{A^2}{3}$.

Segnale dati Anticipiamo l'espressione (15.1) a cui aggiungiamo un elemento di indeterminazione per quanto riguarda la relazione temporale tra l'origine dei tempi e gli istanti caratteristici, scrivendo

$$x(t, \theta) = \sum_{n=-\infty}^{\infty} a_n g(t - nT + \theta) \quad (6.20)$$

con θ v.a. a distribuzione uniforme tra $\pm \frac{T}{2}$, in modo da rendere il processo ergodico³⁹. Mentre il calcolo della sua densità di potenza sarà affrontato al § 7.2.5, qui ci limitiamo ad osservare che, considerando i valori a_n come determinazioni di v.a. indipendenti ed identicamente distribuite, la densità di probabilità di $x(t)$ può euristicamente essere desunta dalla analisi del corrispondente diagramma ad occhio (§ 15.1.2.3). Ad esempio, nel caso di $g(t)$ rettangolare e a_n a due livelli equiprobabili (vedi fig. 15.3 a pag. 440) la $p_X(x)$ sarà costituita da due impulsi di area $1/2$, mentre nei casi di limitazione in banda e/o adozione di un impulso con caratteristica a coseno rialzato, la stessa assumerà un andamento continuo⁴⁰.

6.4 Trasformazione di v.a. e cambio di variabili

Quando più v.a. si combinano con leggi diverse dalla somma, il risultato del § 6.2.5 non è più sufficiente a fornire una espressione per la d.d.p. risultante. Illustriamo quindi il procedimento analitico generale, necessario ad ottenere una espressione per la d.d.p. di una generica funzione di v.a.

6.4.1 Caso unidimensionale

Consideriamo una prima v.a. X , ed una seconda Y da essa derivata per mezzo della relazione $y = f(x)$, che si applica alle determinazioni x di X . Nel caso in cui $f(x)$ sia monotona non decrescente (vedi fig. 6.5-a), e indicando con $x = g(y)$ la corrispondente funzione inversa, la caratterizzazione probabilistica di Y nei termini della sua d.d.p. $p_Y(y)$ può essere ottenuta a partire da quella di X nei termini della funzione di distribuzione di Y , come

$$F_Y(y) = Pr\{Y \leq y\} = Pr\{X \leq g(y)\} \quad (6.21)$$

e calcolando poi $p_Y(y) = \frac{dF_Y(y)}{dy}$. D'altra parte, qualora la trasformazione $f(x)$ non sia monotona come nel caso mostrato in fig. 6.5-b), la (6.21) non è più usabile, in quanto i valori $y \leq \tilde{y}$ hanno origine da due diversi intervalli di X , in corrispondenza dei quali l'area sottesa dalla $p_X(x)$ individua la probabilità cercata.

Procedendo con ordine, trattiamo prima il caso di $f(x)$ monotona crescente come in fig. 6.5-a), in cui per ogni valore di \tilde{y} esiste un solo intervallo di $\tilde{X} \subset X$ tale che

³⁹In assenza del parametro θ , e considerando la sequenza aleatoria degli a_n stazionaria ed ergodica, $x(t, \theta = 0)$ costituisce un processo *ciclostazionario* in senso stretto (vedi https://en.wikipedia.org/wiki/Cyclostationary_process), ossia per il quale le medie di insieme di qualsiasi ordine sono periodiche di periodo T . La presenza della v.a. uniforme θ rende $x(t, \theta)$ un processo stazionario, ed anche ergodico.

⁴⁰In una futura edizione, potrei calcolare le ddp corrispondenti ai diagrammi ad occhio di fig. 15.8

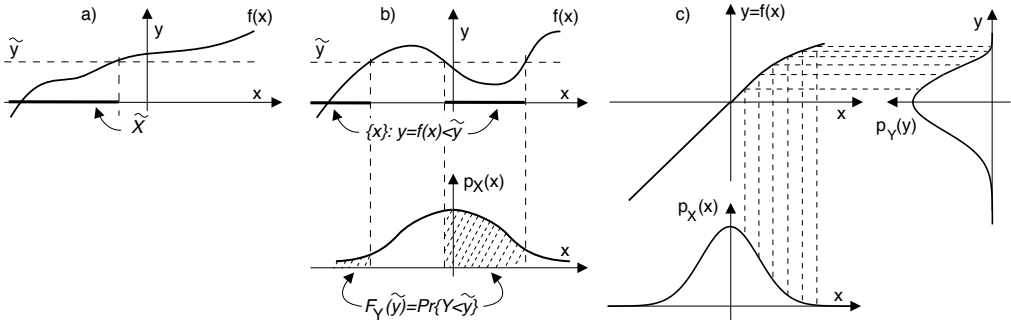


Figura 6.5: Trasformazioni tra variabili aleatorie

$y = f(x)|_{x \in \tilde{X} \leq \tilde{y}}$, e la (6.21) può essere riscritta come

$$F_Y(y) = Pr\{X \leq g(y)\} = F_X(x = g(y))$$

che, derivata, permette di giungere alla espressione che consente il calcolo della $p_Y(y)$:

$$p_Y(y) = \frac{dF_Y(y)}{dy} = \frac{dF_X(x)}{dx} \Big|_{x=g(y)} \frac{dg(y)}{dy} = p_X(g(y)) \frac{dg(y)}{dy} \quad (6.22)$$

La (6.22) indica che la nuova v.a. $y = f(x)$ possiede una d.d.p. pari a quella di x , calcolata con argomento pari alla funzione inversa $x = g(y)$, moltiplicata per la derivata di $g(y)$. La d.d.p. della v.a. risultante si presta anche ad un processo di costruzione grafica, come esemplificato in fig. 6.5-c).

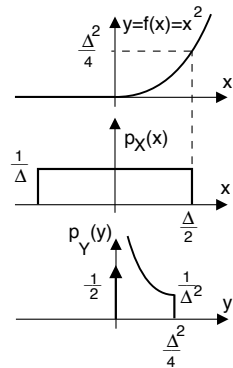
Esempio Determinare $p_Y(y)$, qualora risulti $y = f(x) = \begin{cases} 0 & \text{con } x \leq 0 \\ x^2 & \text{con } x > 0 \end{cases}$, nel caso in cui

$p_X(x) = \frac{1}{\Delta} \text{rect}_{\Delta}(x)$. Osserviamo innanzitutto che tutte le determinazioni

$x \leq 0$ danno luogo ad un unico valore $y = 0$; pertanto si ottiene $p_Y(0) = \frac{1}{2} \delta(y)$. Per $0 < y \leq \frac{\Delta^2}{4}$ (corrispondente ad $0 < x \leq \frac{\Delta}{2}$) si applica la teoria svolta, ottenendo $F_Y(y) = Pr\{x \leq \sqrt{y}\} = F_X(\sqrt{y})$, e dunque

$$p_Y(y) = \frac{dF_Y(y)}{dy} = \frac{dF_X(x)}{dx} \Big|_{x=\sqrt{y}} \frac{d(x = \sqrt{y})}{dy} = \frac{1}{\Delta} \frac{1}{2\sqrt{y}}$$

in cui l'ultima eguaglianza tiene conto che $\frac{dF_X(x)}{dx} = p_X(x)$, che vale $\frac{1}{\Delta}$ per tutti gli x nell'intervallo in considerazione. L'ultima curva mostra la d.d.p risultante per questo esempio.



Se invece la $f(x)$ è monotona *ma decrescente*, consideriamo semplicemente che le probabilità $Pr\{x \leq X \leq x + dx\} = p_X(x) dx$ e $Pr\{y \leq Y \leq y + dy\}|_{y=f(x)} = p_Y(y) dy$ devono essere uguali, ma dato che con $f(x)$ decrescente ad un dx positivo corrisponde un dy negativo, prendiamo il valore assoluto di entrambi: $p_X(x) |dx| = p_Y(y) |dy|$; sostituendo quindi x con la sua funzione inversa $x = g(y)$ e riarrangiando si ottiene

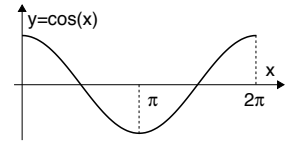
$$p_Y(y) = p_X(g(y)) \left| \frac{dg(y)}{dy} \right| \quad (6.23)$$

che è la versione più generale del risultato (6.22).

Esempio Qualora $f(x)$ sia una relazione lineare $y = ax + b$ possiamo scrivere $x = g(y) = \frac{y-b}{a}$ e $\frac{d}{dy}g(y) = \frac{1}{a}$; pertanto la (6.23) si traduce in $p_Y(y) = \frac{1}{|a|}p_X\left(\frac{y-b}{a}\right)$, ovvero la nuova v.a. Y possiede una d.d.p. con lo stesso andamento di $p_X(x)$, ma traslata di b e compressa o espansa di a .

Trasformazione non monotona In questa circostanza due o più valori di X producono lo stesso valore di Y (vedi fig. 6.5-b)), e non esiste una funzione inversa $x = g(y)$ univoca. In tal caso si suddivide la variabilità di X in più intervalli i , in modo che per ciascuno di essi possa definirsi una $f_i(x)$ monotona: tali intervalli individuano eventi mutuamente esclusivi, e dunque si può calcolare il lato destro di (6.23) per ogni funzione inversa $g_i(y) = f_i^{-1}(x)$, e quindi sommare i risultati per ottenere $p_Y(y)$.

Esempio Consideriamo la funzione $y = f(x) = \cos(x)$ in cui x è una v.a. con d.d.p. uniforme $p_X(x) = \frac{1}{2\pi}rect_{2\pi}(x - \pi)$. Dato che per $0 \leq x \leq \pi$ il coseno è decrescente, mentre per $\pi \leq x \leq 2\pi$ è crescente, applichiamo la (6.23) su questi due intervalli. Per il primo si ha $x = g_1(y) = \arccos(y)$, la cui derivata vale $\frac{dg_1(y)}{dy} = \frac{-1}{\sqrt{1-y^2}}$, mentre $p_X(x)$ è costante e pari a $\frac{1}{2\pi}$ indipendentemente da x , dunque $p_X(g_1(y)) = \frac{1}{2\pi}$. Per il secondo intervallo la funzione inversa vale ancora $x = g_2(y) = \arccos(y)$, così come medesime sono le altre considerazioni. Pertanto si ottiene



$$\begin{aligned}
 p_Y(y) &= p_X(g_1(y)) \left| \frac{dg_1(y)}{dy} \right| + p_X(g_2(y)) \left| \frac{dg_2(y)}{dy} \right| = \\
 &= 2 \cdot \frac{1}{2\pi} \cdot \left| \frac{-1}{\sqrt{1-y^2}} \right| = \begin{cases} \frac{1}{\pi\sqrt{1-y^2}} & -1 \leq y \leq 1 \\ 0 & \text{altrove} \end{cases}
 \end{aligned}$$

6.4.2 Caso multidimensionale

Descriviamo questo caso per mezzo del vettore di v.a. $\mathbf{X} = (x_1, x_2, \dots, x_n)$, a cui è associata una d.d.p. congiunta $p_X(x_1, x_2, \dots, x_n)$, e di un secondo vettore aleatorio \mathbf{Y} dipendente dal primo mediante la trasformazione $\mathbf{Y} = \mathbf{F}(\mathbf{X})$, ovvero

$$\begin{cases} y_1 = f_1(x_1, x_2, \dots, x_n) \\ y_2 = f_2(x_1, x_2, \dots, x_n) \\ \vdots \\ y_n = f_n(x_1, x_2, \dots, x_n) \end{cases} \tag{6.24}$$

Se esiste la relazione inversa $\mathbf{X} = \mathbf{F}^{-1}(\mathbf{Y}) = \mathbf{G}(\mathbf{Y})$ univoca, composta dall'insieme di funzioni $x_i = g_i(y_1, y_2, \dots, y_n)$ per $i = 1, 2, \dots, n$, allora per la d.d.p di \mathbf{Y} sussiste⁴¹ un risultato formalmente molto simile a quello valido nel caso monodimensionale, e cioè

$$p_Y(y_1, y_2, \dots, y_n) = p_X(\mathbf{X} = \mathbf{G}(\mathbf{Y})) \cdot |\det(\mathbf{J}(\mathbf{X}/\mathbf{Y}))| \tag{6.25}$$

⁴¹La dimostrazione segue le medesime linee guida del caso precedente, ed è impostata sulla base della considerazione che la funzione di distribuzione di \mathbf{Y} , calcolata in un generico punto $\tilde{\mathbf{y}} = (\tilde{y}_1, \tilde{y}_2, \dots, \tilde{y}_n)$, rappresenta la probabilità che \mathbf{Y} appartenga alla regione (dominio) delimitata dal punto $\tilde{\mathbf{y}}$, indicata con $\mathcal{D}_{\tilde{\mathbf{y}}}$:

$$F_Y(\tilde{\mathbf{y}}) = Pr\{\mathbf{Y} \leq \tilde{\mathbf{y}}\} = Pr\{\mathbf{Y} \in \mathcal{D}_{\tilde{\mathbf{y}}}\}$$

in cui $p_X(\mathbf{x} = \mathbf{G}(\mathbf{Y}))$ è la d.d.p. di \mathbf{X} calcolata con argomento dipendente da \mathbf{Y} , e $|\det(\mathbf{J}(\mathbf{X}/\mathbf{Y}))|$ è il modulo del *jacobiano* della trasformazione inversa \mathbf{G} , ossia del determinante della matrice costituita da tutte le derivate parziali di \mathbf{G} , detta *jacobiana*⁴²:

$$\mathbf{J}(\mathbf{X}/\mathbf{Y}) = \left[\frac{\partial x_i}{\partial y_j} \right] = \begin{bmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} & \dots & \frac{\partial x_1}{\partial y_n} \\ \frac{\partial x_2}{\partial y_1} & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ \frac{\partial x_n}{\partial y_1} & \dots & \dots & \frac{\partial x_n}{\partial y_n} \end{bmatrix}$$

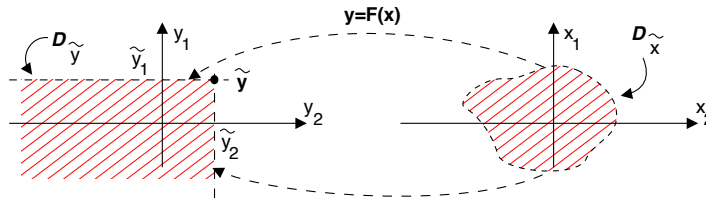
Un esempio di applicazione della teoria appena discussa viene svolta al § 14.4, allo scopo di descrivere in termini probabilistici il problema della detezione di una sinusoide immersa nel rumore; tale descrizione è quindi usata al § 6.6.1 per impostare il problema della decisione statistica. Un altro caso applicativo si riferisce alla d.d.p. del prodotto tra v.a. (pag. 211).

6.5 Gaussiana multidimensionale

Questo termine individua una v.a. vettoriale \mathbf{X} ottenuta a partire da n v.a. marginali x_i , $i = 1, 2, \dots, n$ tutte gaussiane. La d.d.p. *congiunta* in questo caso è espressa in modo formalmente simile a quello del caso unidimensionale, come

$$p_X(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^n \det(\Sigma_x)}} \exp \left\{ -\frac{1}{2} (\mathbf{x} - \mathbf{m}_x) \Sigma_x^{-1} (\mathbf{x} - \mathbf{m}_x)^\top \right\} \quad (6.26)$$

in cui $\mathbf{x} = [x_1, x_2, \dots, x_n]$ è il vettore riga che rappresenta le n v.a. marginali, \mathbf{m}_x è il vettore dei rispettivi valori medi, Σ_x è la *matrice di covarianza* (vedi § 7.1.2) i cui $n \times n$ elementi risultano pari a $\sigma_{x_i, x_j} = E\{(x_i - m_{x_i})(x_j - m_{x_j})\}$, e $^\top$ rappresenta l'operatore di



Alla stessa regione $\mathcal{D}_{\tilde{\mathbf{y}}}$, ne corrisponde una diversa $\mathcal{D}_{\tilde{\mathbf{x}}}$ nello spazio \mathbf{X} , tale che per ogni valore $\mathbf{x}^\circ \in \mathcal{D}_{\tilde{\mathbf{x}}}$ risulti $\mathbf{y}^\circ = \mathbf{F}(\mathbf{x}^\circ) \in \mathcal{D}_{\tilde{\mathbf{y}}}$. Con queste posizioni, la $F_Y(\tilde{\mathbf{y}}) = Pr\{\mathbf{Y} \in \mathcal{D}_{\tilde{\mathbf{y}}}\}$ si calcola a partire dalla d.d.p. $p_X(\mathbf{x})$, integrata sul dominio $\mathcal{D}_{\tilde{\mathbf{x}}}$:

$$F_Y(\tilde{\mathbf{y}}) = Pr\{\mathbf{X} \in \mathcal{D}_{\tilde{\mathbf{x}}}\} = \int_{\mathcal{D}_{\tilde{\mathbf{x}}}} p_X(\mathbf{x}) d\mathbf{x}$$

Infine, osservando che

$$p_Y(y_1, y_2, \dots, y_n) = \frac{\partial^n F_Y(y_1, y_2, \dots, y_n)}{\partial y_1 \partial y_2 \dots \partial y_n}$$

si ottiene il risultato mostrato.

⁴² $\mathbf{J}(\mathbf{X}/\mathbf{Y})$ è indicata come *matrice jacobiana*, ed il suo determinante come *jacobiano*, vedi https://it.wikipedia.org/wiki/Matrice_jacobiana

trasposizione. In tal caso le v.a. marginali x_i vengono dette *congiuntamente gaussiane*, e la conoscenza di \mathbf{m}_x e Σ_x ne definisce in modo *completo* la densità di probabilità. Per un approfondimento delle proprietà analitiche di Σ_x , si veda il § 6.7.3.

Notiamo che il termine $1/\sqrt{(2\pi)^n \det(\Sigma_x)}$ rappresenta l'altezza di $p_X(\mathbf{x})$ per $\mathbf{x} = \mathbf{m}_x$, in cui l'esponente di (6.26) si annulla. Per $\mathbf{x} \neq \mathbf{m}_x$ lo stesso esponente è una *forma quadratica* (pag. 186) sempre positiva, e che cresce all'aumentare di $|\mathbf{x} - \mathbf{m}_x|$.

Esempio In figura 6.6-a) è rappresentato il grafico 3D di una d.d.p. gaussiana bidimensionale

$p_{XY}(x, y)$, con $\mathbf{m} = (0, 1)$ e $\Sigma = \begin{pmatrix} 1 & 0.1 \\ 0.1 & 0.5 \end{pmatrix}$: si può notare l'asimmetria legata ai valori

$\sigma_x^2 \neq \sigma_y^2$, e la centratura dovuta ad $m_y \neq 0$. La fig. 6.6-b) mostra la stessa d.d.p. da un punto di vista corrispondente agli assi, mentre la fig. 6.6-c) ne riporta le *curve di livello*, che mostrano come la forma quadratica dell'esponente determini per la superficie di $p_{XY}(x, y)$ dei contorni *ellittici*, in cui la lunghezza degli assi delle ellissi è legata a σ_x e σ_y , mentre l'inclinazione dipende dalla covarianza σ_{xy} .

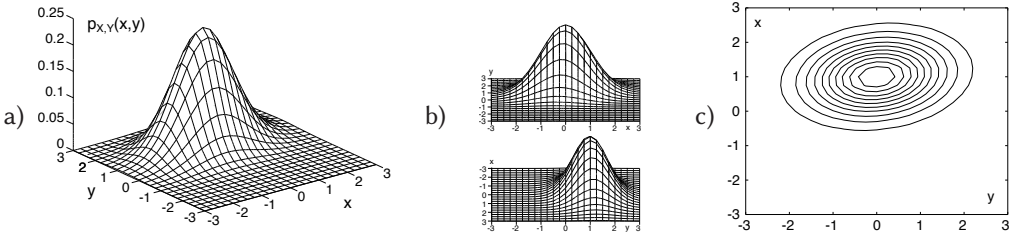


Figura 6.6: a) - d.d.p. gaussiana bidimensionale; b) - vista frontale e laterale; c) - curve di livello

6.5.1 Indipendenza statistica per v.a. gaussiane incorrelate

Affrontiamo la dimostrazione di quanto affermato in fondo al § 7.1.2, ovvero che, *unicamente nel caso di v.a. congiuntamente gaussiane*, il sussistere di incorrelazione tra le stesse ne implica l'indipendenza statistica. Osserviamo infatti che nel caso in cui le v.a. marginali siano incorrelate, ossia $\sigma_{x_i x_j} = 0$ con $i \neq j$, la matrice di covarianza Σ_x risulta essere *diagonale*, così come la sua inversa, i cui elementi risultano in tal caso essere pari a $1/\sigma_{x_i}^2$; inoltre, si ottiene che $\det(\Sigma_x) = \prod_{i=1}^n \sigma_{x_i}^2$. Pertanto in questo caso la (6.26) si esprime come

$$p(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^n \prod_{i=1}^n \sigma_{x_i}^2}} \exp \left\{ -\frac{1}{2} \left[\sum_{i=1}^n \frac{(x_i - m_{x_i})^2}{\sigma_{x_i}^2} \right] \right\} \quad (6.27)$$

che evidentemente equivale al *prodotto* delle singole d.d.p. marginali⁴³

$$p(x_i) = \frac{1}{\sqrt{2\pi}\sigma_{x_i}} \exp \left\{ -\frac{1}{2} \frac{(x_i - m_{x_i})^2}{\sigma_{x_i}^2} \right\}$$

Ma dato che questo risultato è proprio la definizione di indipendenza statistica (§ 7.1.2) tra le v.a. marginali, abbiamo ottenuto la dimostrazione cercata.

⁴³Si verifichi per esercizio che nel caso di una coppia di v.a. congiuntamente gaussiane, a media nulla ed uguale varianza, si ottiene l'espressione (14.11) di pag. 423.

Infine, osserviamo che nel caso in cui le v.a. marginali siano indipendenti, eguagliando l'esponente della (6.27) ad una costante si ottiene l'equazione di una ellisse riferita agli assi principali, ossia le curve di livello di fig. 6.6-c) si dispongono con gli assi paralleli a quelli del dominio descritto dal vettore aleatorio \mathbf{x} .

6.5.2 Trasformazione lineare di v.a. gaussiane

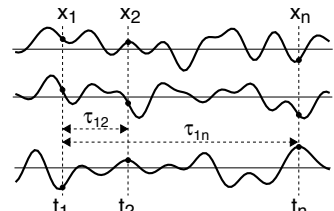
Un'altra importante proprietà di questo tipo di v.a. è la sua *invarianza* rispetto alle operazioni di combinazione lineare. Se infatti indichiamo con \mathbf{X} una v.a. gaussiana multivariata, e con $\mathbf{Y} = \mathbf{X}\mathbf{A}$ un secondo vettore aleatorio ottenuto mediante moltiplicazione di \mathbf{X} per una matrice \mathbf{A} , fornendo $y_j = \sum_{i=1}^n a_{ij}x_i$, possiamo mostrare che anche \mathbf{Y} risulta descrivere una v.a. gaussiana. In accordo con la trattazione svolta al § 6.4.2, scriviamo la trasformazione inversa come $\mathbf{X} = \mathbf{Y}\mathbf{B}$ in cui $\mathbf{B} = \mathbf{A}^{-1}$, mentre per la d.d.p. della nuova v.a. \mathbf{Y} , in base alla (6.25) otteniamo $p_Y(\mathbf{y}) = p_X(\mathbf{x} = \mathbf{y}\mathbf{B}) \cdot \det(\mathbf{B})$, in quanto la matrice jacobiana $\mathbf{J}(\mathbf{X}/\mathbf{Y})$ corrisponde alla trasposta della matrice \mathbf{B} stessa⁴⁴; inoltre, risulta che $\mathbf{m}_x = \mathbf{m}_y\mathbf{B}$. Sostituendo questi risultati nella (6.26) si ottiene pertanto

$$\begin{aligned} p_Y(\mathbf{y}) &= \frac{\det(\mathbf{B})}{\sqrt{(2\pi)^n \det(\Sigma_x)}} \exp \left\{ -\frac{1}{2} (\mathbf{y}\mathbf{B} - \mathbf{m}_y\mathbf{B}) \Sigma_x^{-1} (\mathbf{y}\mathbf{B} - \mathbf{m}_y\mathbf{B})^\top \right\} \\ &= \frac{\det(\mathbf{B})}{\sqrt{(2\pi)^n \det(\Sigma_x)}} \exp \left\{ -\frac{1}{2} (\mathbf{y} - \mathbf{m}_y) \mathbf{B} \Sigma_x^{-1} \mathbf{B}^\top (\mathbf{y} - \mathbf{m}_y)^\top \right\} \end{aligned}$$

che è nuovamente l'espressione di una d.d.p. gaussiana multivariata \mathbf{y} , con media $\mathbf{m}_y = \mathbf{m}_x\mathbf{B}^{-1} = \mathbf{m}_x\mathbf{A}$ e covarianza $\Sigma_y = \mathbf{A}^\top \Sigma_x \mathbf{A}$ ⁴⁵.

6.5.3 Processo gaussiano

Una importante classe di segnali aleatori è costituita da un processo stazionario in senso lato, la cui d.d.p. di primo ordine è gaussiana, e dai cui membri è possibile estrarre ad istanti diversi una o più v.a. gaussiane, che indichiamo collettivamente con il vettore aleatorio \mathbf{x} , descritto dalla d.d.p. multivariata (6.26).



La stazionarietà garantisce che il corrispondente vettore dei valori medi \mathbf{m}_x presenti tutti gli elementi uguali e pari a $m_x = E\{x(t)\}$, e che la matrice di covarianza Σ_x presenti elementi ottenuti valutando la covarianza $\sigma_x(\tau) = E\{(x(t) - m_x)(x(t + \tau) - m_x)\}$ del processo (vedi eq. 7.3) in corrispondenza degli intervalli temporali τ_{ij} tra gli istanti di campionamento nei quali sono estratte le coppie di v.a. marginali x_i e x_j descritte dalla gaussiana multivariata. In altre parole, i valori σ_{ij} che compaiono in Σ_x sono ottenuti come $\sigma_{ij} = \sigma_x(\tau_{ij})$, mentre su tutta la diagonale compare la varianza $\sigma_x^2 = \sigma_x(0)$.

Essendo il processo gaussiano, le due grandezze \mathbf{m}_x e Σ_x lo descrivono completamente, e se si verifica anche l'ipotesi di ergodicità, possono essere stimate a partire da una qualunque realizzazione, vedi eq. (6.28) a pag. 173.

⁴⁴Infatti, potendo scrivere $x_i = \sum_{j=1}^n b_{ji}y_j$, l'elemento i, j della matrice \mathbf{J} risulta pari a $j_{ij} = \frac{\partial x_i}{\partial y_j} = b_{ji}$.

⁴⁵Infatti risulta $(\mathbf{B}\Sigma_x^{-1}\mathbf{B}^\top)^{-1} = (\mathbf{B}^\top)^{-1}\Sigma_x\mathbf{B}^{-1}$ che, essendo $\mathbf{B}^{-1} = \mathbf{A}$, fornisce il risultato per Σ_y .

6.6 Elementi di statistica

La statistica si occupa dell'utilizzo numerico di osservazioni e misure, il cui esito viene pensato come il risultato di un fenomeno aleatorio, al fine ad es. di prendere delle decisioni, o *inferire*⁴⁶ dai dati la plausibilità di ipotesi. Ma le cose *si rendono interessanti* per il fatto che

- potendo ripetere la misura e/o l'osservazione più volte, è possibile ridurre il margine di errore, e migliorare l'affidabilità della decisione;
- non è detto che la d.d.p. del fenomeno in esame sia completamente definita.

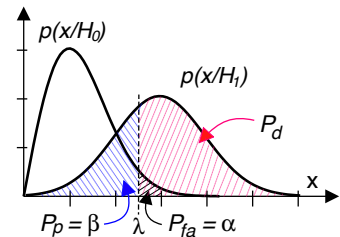
Ad esempio, la d.d.p. che produce i valori di osservazione può essere nota, a meno del valore dei suoi parametri media e varianza: compito della statistica è allora anche quello di *stimare* grandezze altrimenti ignote. Senza alcuna ambizione di affrontare compiutamente l'argomento, limitiamoci qui ad esporre brevemente questi particolari campi applicativi.

6.6.1 Test di verifica di ipotesi

Si tratta di stabilire se sia da ritenere valida una ipotesi H_1 , oppure se al contrario non si debba ritenere valida l'ipotesi *alternativa* H_0 (o ipotesi *nulla*); la decisione viene presa confrontando un grandezza di osservazione x (che *incarna* una realizzazione di v.a. X) con un valore-soglia (*di decisione*) λ : se $x > \lambda$ si decide per H_1 , mentre viceversa se $x < \lambda$ si decide per H_0 . Allo stesso tempo, si desiderano valutare i margini probabilistici che indicano il grado di affidabilità della decisione.⁴⁷

Sviluppiamo il ragionamento basandoci sui risultati dell'analisi svolta al § 14.4.2 e riassunti dalla figura a lato, che illustra come nei due casi H_0 ed H_1 la v.a. x presenti due diverse d.d.p. condizionate $p(x/H_0)$ e $p(x/H_1)$. Sempre in figura sono indicate le aree di valore P_d e $P_p = 1 - P_d$ rispettivamente pari alla probabilità (di *detezione*) che x superi la soglia λ nel caso in cui sia vera H_1 , e (di *perdita*) che non la superi. La P_{fa} è invece detta di prob. di *falso allarme* in quanto rappresenta la prob. che $x > \lambda$ nel caso in cui sia vera H_0 , ovvero di decidere per H_1 quando è vera H_0 , evento detto anche *errore di primo tipo* o *livello di significatività* del test, e indicata con α . La prob. di *detezione* P_d è invece detta *potenza* del test ed indicata come $1 - \beta$, dove β evidentemente è la probabilità di *perdita* P_p o *errore di secondo tipo*.

Conoscendo la d.d.p. $p(x/H_0)$ della grandezza di osservazione quando è vera H_0 , si pone la soglia λ ad un valore tale che $Pr(x > \lambda/H_0) = \int_{\lambda}^{\infty} p(x/H_0) dx = \alpha$, tale cioè da rendere $P_{fa} = \alpha$, in genere posto pari a 0.05 (ovvero il 5%). Qualora l'osservazione



⁴⁶Vedi ad es. http://it.wikipedia.org/wiki/Inferenza_statistica

⁴⁷Un modello del genere si applica tanto al caso di detezione di un bersaglio radar, che può essere presente o meno, quanto ai casi di una diagnosi medica a partire dai risultati degli esami clinici, a quello di attuare o meno un investimento finanziario a partire dall'andamento delle borse, a quello se prendere o meno l'ombrello prima di uscire di casa a partire dallo scrutare il cielo...

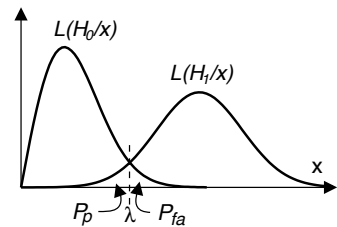
di $\check{x} > \lambda$ faccia decidere per H_1 , spesso al risultato si aggiunge anche la probabilità $\int_{\check{x}}^{\infty} p(x/H_0) dx$ detta *p-value*⁴⁸, che descrive quanto sia *poco probabile* osservare un valore di x che si discosti ancor più da H_0 di quanto non sia avvenuto per la “nostra” osservazione \check{x} , e che perciò indica meglio di α la significatività del test, ovvero la *discrepanza* tra i dati osservati e l’ipotesi nulla H_0 . Naturalmente, per tutte le possibili osservazioni $x > \lambda$ risulta un *p-value* $< \alpha$.

6.6.2 Funzione di verosimiglianza

Viene definita a partire da una probabilità condizionata, qualora nella stessa espressione siano *invertiti* i ruoli di evento condizionato e condizionante, nel senso che ora illustriamo. Rimanendo nel campo della verifica di ipotesi, il valore assunto dalle d.d.p. condizionate $p_X(\check{x}/H_0)$ e $p_X(\check{x}/H_1)$ qualora valutate in $x = \check{x}$ (l’osservazione) può essere interpretato come la *verosimiglianza* di H_0 ed H_1 una volta noto che $x = \check{x}$, ribaltando il ruolo delle ipotesi da quello di evento condizionante a quello di variabile indipendente. Ma attenzione, una verosimiglianza *non* è una probabilità, in quanto $p_X(\check{x}/H_0) + p_X(\check{x}/H_1) \neq 1$! La funzione di verosimiglianza (o *likelihood*) è spesso indicata con la lettera \mathcal{L} , in modo da poter scrivere $\mathcal{L}(H_{0,1}/\check{x}) = p_X(\check{x}/H_{0,1})$.

6.6.2.1 Decisione di massima verosimiglianza

Riprendendo il tema della verifica di ipotesi § 6.6.1, osserviamo che qualora la soglia di decisione λ *non sia* assegnata in base al livello di significatività $P_{fa} = \alpha$ desiderato, ma in corrispondenza della intersezione tra $p_X(x/H_0)$ e $p_X(x/H_1)$ come mostrato alla figura che segue, la decisione stessa è detta di *massima verosimiglianza*, dato che sceglie l’ipotesi H_i per la quale la funzione di verosimiglianza $\mathcal{L}(H_i/\check{x})$ è massima. Notiamo che un tale criterio di decisione comporta $P_p = P_{fa}$ solo nel caso in cui la d.d.p. sia la stessa per entrambe le ipotesi (vedi ad es. § 15.4.2); inoltre, anche in questo caso alla decisione effettuata si può *allegare* il relativo *p-value*. A volte, anziché confrontare tra di loro i valori di \mathcal{L} , se ne esegue *il rapporto*, decidendo quindi in base all’esito maggiore o minore di uno (vedi ad es. pag. 552).

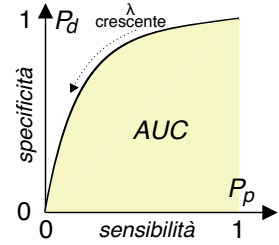


6.6.2.2 Curva operativa

Per quanto fino ad ora esposto, all’aumentare della soglia λ si assiste ad un aumento della P_{fa} a cui corrisponde una diminuzione della P_d , da cui la possibilità di tracciare una curva parametrica sui cui assi sono riportati entrambi i valori, al variare appunto di λ . Tale rappresentazione fu ideata ai fini della detezione radar e per tale motivo denominata *ROC* o *Receiver Operating Characteristic*, e da allora viene adottata anche in altri ambiti come ad esempio quelli clinici, psicologici, o meteorologici. In tali contesti la prob. di detezione P_d viene anche indicata come *specificità* del test o percentuale di *veri positivi*, mentre la prob. di perdita $P_p = 1 - P_{fa}$ è detta *sensibilità* o % di *falsi positivi*.

⁴⁸Vedi ad es. <http://en.wikipedia.org/wiki/P-value>

Ovviamente l'andamento esatto della curva, esemplificato a lato, dipende dalla conoscenza delle due probabilità condizionate $p_X(x/H_0)$ e $p_X(x/H_1)$. Il classificatore che si basa sulla osservazione x viene quindi giudicato come *migliore* o *peggiore* rispetto ad un altro basato su di un'altro parametro y in base al computo dell'area compresa tra la curva e l'asse delle ascisse (AUC, *Area Under the Curve*), pari⁴⁹ alla probabilità che per un evento (a caso) della classe H_1 si osservi un valore x maggiore di quello per un evento (a caso) appartenente alla classe H_0 .



6.6.3 Stima di parametro

Nel caso in cui si conosca il tipo di v.a. che caratterizza il fenomeno aleatorio in esame, ma non i suoi parametri (es. media m_x e varianza σ_x^2), una sequenza di valori osservati $\mathbf{x} = (x_1, x_2, \dots, x_N)$ può essere usata per effettuare la *stima*, ovvero la valutazione sperimentale (ma imprecisa) del valore di \hat{m}_x e $\hat{\sigma}_x^2$. Indicando con θ il vero (incognito) valore di un parametro da stimare, il suo *stimatore* $\hat{\theta}(\mathbf{x})$ dipende dalla variabilità di \mathbf{x} e quindi è esso stesso una v.a., che può essere caratterizzata nei termini della relativa d.d.p. $p_{\hat{\theta}}(\hat{\theta})$, nonché di media $m_{\hat{\theta}}$ e varianza $\sigma_{\hat{\theta}}^2$.

Anche se per uno specifico vettore di osservazione \mathbf{x} lo stimatore commette un errore

$$\varepsilon(\mathbf{x}) = \hat{\theta}(\mathbf{x}) - \theta$$

desidereremmo che il valore atteso di tale errore fosse nullo, ovvero che

$$m_{\hat{\theta}} = E\{\hat{\theta}(\mathbf{x})\} = \theta$$

Qualora ciò si verifichi lo stimatore è detto *non polarizzato*⁵⁰, o *centrato*. Per quanto riguarda la varianza della stima

$$\sigma_{\hat{\theta}}^2 = E\{(\hat{\theta}(\mathbf{x}) - m_{\hat{\theta}})^2\}$$

vorremmo che fosse la più piccola possibile. Uno stimatore non polarizzato e che consegue la minima varianza tra i possibili stimatori è detto *efficiente* ed eguaglia il limite di Cramér-Rao⁵¹; viene inoltre detto *consistente* se $\lim_{N \rightarrow \infty} \Pr\{\hat{\theta}(\mathbf{x}) = \theta\} = 1$, ovvero se all'aumentare della dimensione N del campione, *sicuramente* $\hat{\theta}(\mathbf{x}) \rightarrow \theta$. Infine, uno stimatore consistente è anche *asintoticamente normale* se all'aumentare di N la sua d.d.p. tende ad una gaussiana⁵² con valor medio θ , e varianza $\sigma_{\hat{\theta}}^2$ che si riduce come $1/N$.

Prendiamo dunque in esame una tra le diverse tecniche⁵³ che consentono di arrivare

⁴⁹Vedi ad es. https://en.wikipedia.org/wiki/Receiver_operating_characteristic

⁵⁰Dall'inglese *unbiased*, ove con *bias* si intende una forma di *errore sistematico*. Diversi testi usano il termine *non distorto*, che qui non è adottato onde evitare confusioni concettuali con il cap. 8.

⁵¹Vedi ad es. http://it.wikipedia.org/wiki/Disuguaglianza_di_Cramér-Rao

⁵²Ad esempio, il teorema centrale del limite (§ 6.7.2) fa sì che la *media campionaria* (6.28) $\hat{m}_x = \frac{1}{N} \sum_{i=1}^N x_i$, in quanto somma di v.a. indipendenti e identicamente distribuite, tenda ad una v.a. gaussiana per $N \rightarrow \infty$.

⁵³Vedi ad es. http://en.wikipedia.org/wiki/Point_estimation

ad un procedimento di stima.

6.6.3.1 Stima di massima verosimiglianza

In questo caso il valore stimato è quello che massimizza la funzione di verosimiglianza (§ 6.6.2) calcolata in corrispondenza ai dati campionari \check{x} , ovvero $\hat{\theta} = \operatorname{argmax}_{\theta} \{ \mathcal{L}(\theta/\check{x}) \}$. Ad esempio, nel caso in cui i valori osservati $\check{x} = (x_1, x_2, \dots, x_N)$ siano ottenuti da un fenomeno aleatorio caratterizzato da una d.d.p. gaussiana con media m_x e varianza σ_x^2 (incognite), e gli esperimenti siano statisticamente indipendenti, la funzione di verosimiglianza risulta⁵⁴

$$\mathcal{L}(m_x, \sigma_x^2/\check{x}) = p_X(\check{x}/m_x, \sigma_x^2) = \prod_{i=1}^N \frac{1}{\sqrt{2\pi\sigma_x^2}} \exp\left\{-\frac{1}{2}\left(\frac{x_i - m_x}{\sigma_x}\right)^2\right\}$$

ma a questa espressione si preferisce il suo logaritmo naturale $L = \ln \mathcal{L}$, che essendo *monotono* non altera la massimizzazione cercata:

$$L(m_x, \sigma_x^2|\check{x}) = \ln \mathcal{L}(m_x, \sigma_x^2|\check{x}) = -\frac{N}{2} \ln(2\pi\sigma_x^2) - \frac{1}{2} \sum_{i=1}^n \left(\frac{x_i - m_x}{\sigma_x}\right)^2$$

Per trovare le stime \hat{m}_x e $\hat{\sigma}_x^2$ di massima (log)verosimiglianza, si eguagliano a zero le derivate parziali di L

$$\begin{aligned} \frac{\partial L}{\partial m_x} &= \frac{1}{\sigma_x^2} \sum_{i=1}^N (x_i - m_x) = 0 \\ \frac{\partial L}{\partial \sigma_x^2} &= -\frac{N}{2} \frac{1}{\sigma_x^2} + \frac{1}{2\sigma_x^4} \sum_{i=1}^N (x_i - m_x)^2 = 0 \end{aligned}$$

ottenendo gli stimatori (o *statistiche*)⁵⁵

$$\begin{cases} \hat{m}_x = \frac{1}{N} \sum_{i=1}^N x_i & \text{media campionaria} \\ \hat{\sigma}_x^2 = \frac{1}{N} \sum_{i=1}^N (x_i - \hat{m}_x)^2 & \text{varianza campionaria} \end{cases} \quad (6.28)$$

di cui ora discutiamo uno alla volta.

Stima della media La media campionaria \hat{m}_x risulta *non polarizzata*⁵⁶ e la sua varianza decresce⁵⁷ con N , ovvero

⁵⁴Vedi § 6.5.1 per l'espressione di una gaussiana multidimensionale.

⁵⁵Per quanto riguarda \hat{m}_x , imponendo $\sum_i (x_i - \hat{m}_x) = 0$ si perviene facilmente al risultato, mentre per $\hat{\sigma}_x^2$ l'eguaglianza a zero di $\frac{\partial L}{\partial \sigma_x^2}$ produce

$$\frac{1}{2\sigma_x^4} \sum_i (x_i - \hat{m}_x)^2 = \frac{N}{2} \frac{1}{\sigma_x^2} \quad \text{ovvero} \quad \frac{1}{\sigma_x^2} \sum_i (x_i - \hat{m}_x)^2 = N$$

e dunque il risultato (6.28).

⁵⁶Infatti

$$E\{\hat{m}_x\} = E\left\{\frac{1}{N} \sum_{i=1}^N x_i\right\} = \frac{1}{N} \sum_{i=1}^N E\{x_i\} = \frac{1}{N} N m_x = m_x$$

⁵⁷In questo caso riscriviamo \hat{m}_x come $\hat{m}_x = \sum_{i=1}^N \frac{x_i}{N}$, consideriamo che la varianza di una somma di v.a. i.i.d. è la somma delle varianze (vedi § 7.5.2), e che $\sigma_{aX}^2 = a^2 \sigma_X^2$: pertanto si ottiene $\sigma_{\hat{m}_x}^2 = \sum_{i=1}^N \frac{\sigma_x^2}{N^2} = \frac{\sigma_x^2}{N}$.

$$E \{ \hat{m}_x \} = m_x; \quad \sigma_{\hat{m}_x}^2 = \frac{\sigma_x^2}{N} \quad (6.29)$$

risulta pertanto essere uno stimatore *consistente e asintoticamente normale*. In particolare

- la proprietà $\sigma_{\hat{m}_x}^2 = \sigma_x^2/N$ ci permette di ridurre a piacimento la variabilità della stima, ripetendo le osservazioni per un adeguato numero di volte;
- all'aumentare di N ed in presenza di indipendenza statistica tra i campioni, il teorema centrale del limite (§ 6.7.2) stabilisce che \hat{m}_x è una v.a. gaussiana, indipendentemente dalla natura della v.a. x da cui è estratto il campione.

Esercizio Si desidera stimare il valor medio \hat{m}_x di una v.a. x di cui si conosce il valore della relativa varianza $\sigma_x^2 = 4$; determinare la dimensione N del campione statistico affinché la varianza della stima sia pari a 0.1. Dalla seconda delle (6.29) risulta $N = \frac{\sigma_x^2}{\sigma_{\hat{m}_x}^2} = \frac{4}{0.1} = 40$ campioni della v.a.

Stima della varianza I calcoli per determinare il valore atteso $m_{\hat{\sigma}_x^2} = E \left\{ \frac{1}{N} \sum_{i=1}^N (x_i - \hat{m}_x)^2 \right\}$ della varianza campionaria $\hat{\sigma}_x^2$ passano prima per la dimostrazione⁵⁸ che $\sum_i (x_i - \hat{m}_x)^2 = \sum_i (x_i - m_x)^2 - N (\hat{m}_x - m_x)^2$, in modo da ottenere

$$\begin{aligned} m_{\hat{\sigma}_x^2} &= \frac{1}{N} \sum_{i=1}^N E \{ (x_i - m_x)^2 \} - \frac{N}{N} E \{ (\hat{m}_x - m_x)^2 \} = \\ &= \frac{1}{N} N \sigma_x^2 - \frac{\sigma_x^2}{N} = \sigma_x^2 \frac{N-1}{N} \end{aligned}$$

dato che il primo termine corrisponde all'espressione della varianza di X , ed il secondo alla varianza di \hat{m}_x , calcolata prima. Il risultato ottenuto indica come la varianza campionaria si riveli uno stimatore *polarizzato*. Per questo, la formula dello stimatore (di massima verosimiglianza) della varianza viene ordinariamente sostituita dall'espressione

$$\hat{\sigma}_x^2 = \frac{1}{N-1} \sum_{i=1}^N (x_i - \hat{m}_x)^2 \quad (6.30)$$

attuando così la *correzione di Bessel*⁵⁹.

6.6.4 Stima di intervallo

L'esercizio precedente conferma la proprietà di *consistenza* per la media campionaria, ma mette anche in luce come per dimensioni N del campione finite, la sua accuratezza

⁵⁸Occorre innanzitutto riscrivere $x_i - \hat{m}_x$ come $x_i - m_x + m_x - \hat{m}_x = (x_i - m_x) - (\hat{m}_x - m_x)$, in modo da ottenere $(x_i - \hat{m}_x)^2 = (x_i - m_x)^2 - 2(x_i - m_x)(\hat{m}_x - m_x) + (\hat{m}_x - m_x)^2$. Eseguendo ora la sommatoria su i si ottiene

$$\begin{aligned} \sum_i (x_i - \hat{m}_x)^2 &= \sum_i (x_i - m_x)^2 - 2(\hat{m}_x - m_x) \sum_i (x_i - m_x) + \sum_i (\hat{m}_x - m_x)^2 \\ &= \sum_i (x_i - m_x)^2 - 2N(\hat{m}_x - m_x)^2 + N(\hat{m}_x - m_x)^2 = \sum_i (x_i - m_x)^2 - N(\hat{m}_x - m_x)^2 \end{aligned}$$

in quanto $\sum_i (x_i - m_x) = \sum_i x_i - \sum_i m_x = N\hat{m}_x - Nm_x = N(\hat{m}_x - m_x)$.

⁵⁹Vedi ad es. https://it.wikipedia.org/wiki/Correzione_di_Bessel

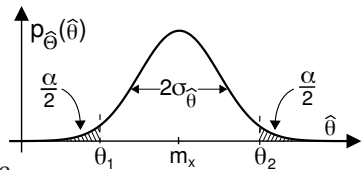
sarà necessariamente limitata. Allora, dato che la stima del parametro θ sarà comunque erronea, può invece avere interesse stimare entro quale intervallo (detto *intervallo di confidenza*) di valori (θ_1, θ_2) possa ricadere θ con sufficiente certezza, espressa quest'ultima dalla probabilità $Pr\{\theta_1 \leq \theta \leq \theta_2\} = 1 - \alpha$, detta *livello di confidenza*, mentre α è la probabilità *di errore*.

Intervallo per la media con varianza nota Nel caso in cui lo stimatore $\hat{\theta}$ sia espresso dalla media campionaria

$$\hat{\theta} = \hat{m}_x = \frac{1}{N} \sum_{i=1}^N x_i$$

a pag. 174 si è osservato che per N sufficientemente elevato $\hat{\theta}$ è una v.a. gaussiana⁶⁰ con media $m_{\hat{\theta}} = m_x$ e deviazione standard $\sigma_{\hat{\theta}} = \frac{\sigma_x}{\sqrt{N}}$, come mostrato alla figura seguente; consideriamo quindi σ_x nota a priori.

Adottando il cambio di variabile $z = \frac{\hat{\theta} - m_x}{\sigma_{\hat{\theta}}}$ ci conduciamo a risultati ottenibili per una v.a. gaussiana normalizzata⁶¹, in modo che la condizione $Pr\{\theta_1 \leq m_x \leq \theta_2\} = 1 - \alpha$ possa essere riscritta⁶² come



$$Pr\left\{-z_{\alpha/2} \leq z = \frac{\hat{\theta} - m_x}{\sigma_{\hat{\theta}}} \leq z_{\alpha/2}\right\} = 1 - \alpha \tag{6.31}$$

in cui $z_{\alpha/2}$ è il *percentile*⁶³ $\alpha/2$ per z , e rappresenta il valore che l'errore *normalizzato* di stima $\frac{|\hat{\theta} - m_x|}{\sigma_{\hat{\theta}}}$ supera con probabilità α , vedi fig. 6.7-a). Una serie di riscritture⁶⁴ della (6.31) porta al risultato

$$Pr\left\{\hat{m}_x - z_{\alpha/2} \cdot \frac{\sigma_x}{\sqrt{N}} < m_x < \hat{m}_x + z_{\alpha/2} \cdot \frac{\sigma_x}{\sqrt{N}}\right\} = 1 - \alpha \tag{6.32}$$

in cui a $\sigma_{\hat{\theta}}$ si è sostituito il suo valore $\frac{\sigma_x}{\sqrt{N}}$, che dipende dalla varianza della popolazione e dalla dimensione del campione. Il valore di $z_{\alpha/2}$ relativo al livello di confidenza α si ottiene tramite apposite tabelle⁶⁵ come quella di fig. 6.7-b).

⁶⁰Indipendentemente alla natura di x

⁶¹Ossia a media nulla e varianza unitaria come a pag. 154, mentre la normalizzazione del § 6.2.4.1 prevede $\sigma^2 = 1/2$.

⁶²Anche grazie a fatto che la gaussiana è simmetrica, dando luogo ad intervalli centrati rispetto a $\hat{\theta}$.

⁶³Il percentile η per una v.a. *gaussiana normalizzata* z è definito come il valore z_η tale che $Pr\{z \leq z_\eta\} = \eta$ e quindi corrisponde alla inversa $z_\eta = F_Z^{-1}(\eta)$ della funzione di distribuzione della v.a. $F_Z(z_\eta) = \int_{-\infty}^{\eta} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz = \eta$. Alternativamente, è definito come $1 - Q\{z_\eta\} = \eta$ (vedi pag. 154). Il termine *percentile* scaturisce dall'essere $\eta \cdot 100\%$ pari alla percentuale delle volte che una determinazione della v.a. z risulta inferiore a z_η .

⁶⁴

$$-z_{\alpha/2} \leq \frac{\hat{\theta} - m_x}{\sigma_{\hat{\theta}}} \leq z_{\alpha/2}$$

$$-z_{\alpha/2} \cdot \sigma_{\hat{\theta}} \leq \hat{\theta} - m_x \leq z_{\alpha/2} \cdot \sigma_{\hat{\theta}}$$

$$-\hat{\theta} - z_{\alpha/2} \cdot \sigma_{\hat{\theta}} \leq -m_x \leq -\hat{\theta} + z_{\alpha/2} \cdot \sigma_{\hat{\theta}}$$

$$\hat{\theta} - z_{\alpha/2} \cdot \sigma_{\hat{\theta}} \leq m_x \leq \hat{\theta} + z_{\alpha/2} \cdot \sigma_{\hat{\theta}}$$

⁶⁵https://it.wikipedia.org/wiki/Funzione_di_ripartizione_della_variabile_casuale_normale

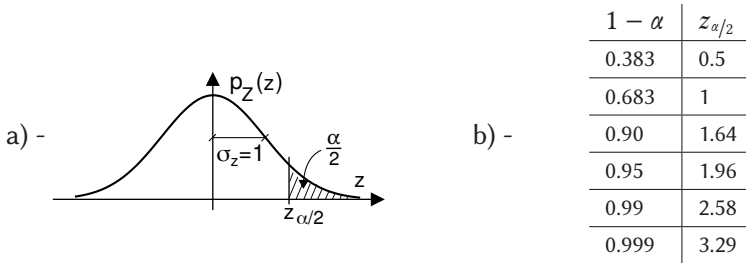


Figura 6.7: Gaussiana normalizzata: a) - percentile; b) - livello di confidenza

Esempio Avendo deciso il livello di confidenza $1 - \alpha$ desiderato, dalla tabella mostrata in fig. 6.7-b) si determina il corrispondente valore di $z_{\alpha/2}$. Ad un valore normalizzato $z_{\alpha/2} = 1$ corrisponde un errore massimo di stima pari a $\sigma_{\hat{\theta}}$, che comporta un livello di confidenza $1 - \alpha = 68.3\%$; aumentando il livello fino al 95% si ottiene un valore di percentile $z_{\alpha/2}$ pari 1.96, mentre ad un livello del 99% corrisponde un percentile di 2.58, ossia un intervallo complessivo esteso per $2 \cdot 2.58 = 5.16$ volte la $\sigma_{\hat{\theta}}$.

Notiamo che per ridurre l'ampiezza dell'intervallo di confidenza dell'errore di stima non normalizzato $|\hat{m}_x - m_x|$ esistono due vie. La prima consiste nell'accettare un minore livello di confidenza $1 - \alpha$, ma ciò equivale ad aumentare la prob. α che l'errore *effettivo* sia maggiore di $z_{\alpha/2} \cdot \sigma_{\hat{\theta}}$. La seconda via (se praticabile) è migliore e consiste nell'aumentare la dimensionalità N del campione statistico, sfruttando così la riduzione con legge $1/\sqrt{N}$ della deviazione standard della media campionaria $\sigma_{\hat{\theta}} = \frac{\sigma_x}{\sqrt{N}}$.

Per approfondire il significato del risultato ottenuto alla (6.32) consideriamo che per ogni campione \mathbf{x} di N elementi si ottiene una *diversa* stima \hat{m}_x , a cui corrisponde un intervallo (θ_1, θ_2) centrato in \hat{m}_x della stessa ampiezza ma *ogni volta differente*. Ciò significa che essendo per ogni campione \mathbf{x} la probabilità che il valore vero della media m_x cada all'interno di (θ_1, θ_2) pari ad $1 - \alpha$, esiste una probabilità α che un diverso campione \mathbf{x}' produca un intervallo (θ_1, θ_2) che *NON* contiene la vera media m_x .

Esercizio Si voglia stimare l'età media m degli utenti di una biblioteca civica. A questo scopo si intervista un campione casuale composto da $N = 100$ persone, da cui si ottiene una media campionaria pari ad $\hat{m} = 29$ anni. Adottando un valore di deviazione standard per l'età pari a $\sigma = 8$ anni, valutare gli intervalli di confidenza per m , con un livello pari al 95% ed al 99%. **Risposta** Essendo N sufficientemente grande, assumiamo vera l'ipotesi di gaussianità per la v.a. \hat{m} e dunque applicabile la teoria esposta. Dunque il percentile $z_{\alpha/2}$ nei due casi $1 - \alpha = 95$ e 99% vale rispettivamente $z_{0.025} = 1.96$ e $z_{0.005} = 2.58$, mentre la varianza della media campionaria risulta $\sigma_{\hat{m}} = \frac{\sigma}{\sqrt{N}} = 0.8$ fornendo in definitiva $z_{0.025} \cdot \frac{\sigma_x}{\sqrt{N}} = 1.96 \cdot 0.8 = 1.57$ e $z_{0.005} \cdot \frac{\sigma_x}{\sqrt{N}} = 2.58 \cdot 0.8 = 2.06$ nei due casi; pertanto si ottiene

$$\begin{cases} 27.43 < m < 30.57 & \text{con confidenza 95\%} \\ 26.94 < m < 31.06 & \text{con confidenza 99\%} \end{cases}$$

ovvero la fascia di età entro cui collocare la *vera età media* m , in base al campione a disposizione, al diminuire della probabilità che m sia in realtà *esterno* all'intervallo. **Nel caso** in cui il campione venga ridotto a sole $N = 30$ persone, ammettendo gli stessi valori di $\hat{m} = 29$ anni e $\sigma = 8$ anni, dato che $\sqrt{30} = 5.48$, ora risulta $z_{0.025} \cdot \sigma_{\hat{m}} = 2.86$

e $z_{0.005} \cdot \sigma_{\hat{m}} = 3.77$, a cui corrispondono intervalli (26.14 - 31.86) e (25.23 - 32.77) con confidenza 95 e 99% rispettivamente. **Poniamo infine** che un *diverso* campione sempre di 30 persone produca una stima $\hat{m} = 27$ anni anziché 29. In tal caso l'intervallo di confidenza al 95% diviene (24.14 - 29.86).

Caso di varianza incognita Qualora la varianza σ_x della v.a. x di cui vogliamo stimare la media m_x non sia nota, ad essa viene sostituito il valore della varianza campionaria non polarizzata $\hat{\sigma}_x^2 = \frac{1}{N-1} \sum_{i=1}^N (x_i - \hat{m}_x)^2$. Una prima conseguenza è che ora $\hat{\sigma}_x^2$ è anch'essa una v.a., ed il suo valore cambia per ogni diversa scelta del campione N , producendo intervalli di confidenza di *estensione variabile*. Essendo $\hat{\sigma}_x^2$ uno stimatore non polarizzato e consistente, per N sufficientemente elevato l'intervallo può ancora essere ottenuto come indicato dalla (6.32).

Accade inoltre che al diminuire di N la variabilità di $\hat{\sigma}_x^2$ causa (a parità di livello di confidenza) un aumento dell'estensione dell'intervallo, in quanto ora il cambio di variabile che porta ad una v.a. normalizzata⁶⁶ $t = \frac{\hat{\theta} - m_x}{\sigma_{\hat{\theta}}}$ dà luogo ad una v.a. non più gaussiana bensì di Student⁶⁷ con $N - 1$ gradi di libertà, rappresentata in fig. 6.8. Per $N \rightarrow \infty$ la v.a. $t^{(N)}$ di Student tende ad una gaussiana, mentre per N più piccoli la sua velocità di azzeramento è ridotta, o come si dice presenta *code più pesanti*. Ciò comporta che il percentile da utilizzare nella (6.32) non è più lo $z_{\alpha/2}$ relativo alla gaussiana normalizzata, bensì il valore $t_{\alpha/2}^{(N-1)}$ relativo alla nuova d.d.p., e tabulato in fig. 6.8-b).

Come è possibile notare confrontando con la tabella in fig. fig. 6.7-b), a parità di livello di confidenza $1 - \alpha$ i valori di $t_{\alpha/2}^{(N-1)}$ sono sempre maggiori di quelli per $z_{\alpha/2}$, tanto più per bassi N , in quanto in tal caso la d.d.p. di Student è più *bassa e larga*, e per catturare la stessa area α occorre *allontanarsi di più* dal valor medio. Nel caso di σ_x^2 incognita l'espressione dell'intervallo di confidenza risulta pertanto

$$Pr \left\{ \hat{m}_x - t_{\alpha/2}^{(N-1)} \cdot \frac{\hat{\sigma}_x}{\sqrt{N}} < m_x < \hat{m}_x + t_{\alpha/2}^{(N-1)} \cdot \frac{\hat{\sigma}_x}{\sqrt{N}} \right\} = 1 - \alpha \quad (6.33)$$

⁶⁶Ossia con media nulla e varianza unitaria.

⁶⁷Vedi https://it.wikipedia.org/wiki/Distribuzione_t_di_Student. La v.a. $t = \frac{\hat{\theta} - m_x}{\sigma_{\hat{\theta}}}$ è definita come il rapporto di due v.a.: il numeratore $\hat{\theta} - m_x$ si comporta come una gaussiana centrata in quanto la media di insieme è una somma di v.a. gaussiane, mentre il denominatore $\sigma_{\hat{\theta}} = \sqrt{\hat{\sigma}_x^2/N}$ dipende da $\hat{\sigma}_x^2$ che è una somma di quadrati di gaussiane, e dunque assume d.d.p. *chi quadro* o χ^2 , vedi § 6.6.5.

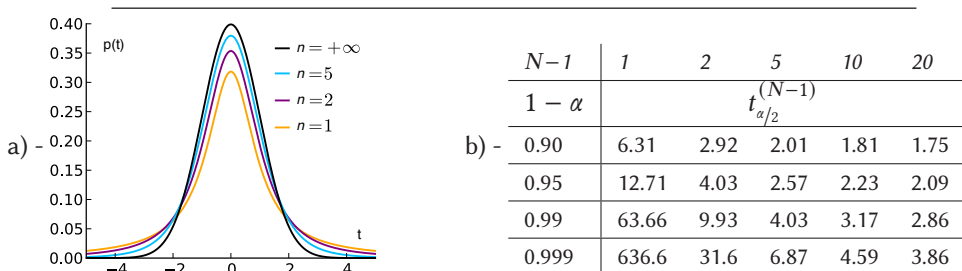


Figura 6.8: Densità di probabilità di Student al variare dei gradi di libertà m e relativi percentili

in cui i percentili $t_{\alpha/2}$ di Student sono ricavati dalla tabella precedente⁶⁸, e $\hat{\sigma}_x^2$ è la varianza campionaria non polarizzata.

Esempio Torniamo al problema della stima dell'età media dei frequentatori di una biblioteca, e svolgiamo le interviste su di un campione di soli 6 individui. I nuovi percentili per un livello di confidenza $1 - \alpha$ pari al 95% e 99% sono rispettivamente $t_{\alpha/2} = 2.57$ e 4.03. Pertanto, ponendo di aver anche stavolta stimato una media campionaria pari ad $\hat{m} = 29$ anni ed una deviazione standard di $\hat{\sigma} = 8$ anni, i nuovi intervalli di confidenza per l'età media risultano (20.6 - 37.39) e (15.83 - 42.16).

Dimensione del campione Il risultato (6.33) è valido, anche per per valori di N piccoli, purché la v.a. di osservazione x sia essa stessa gaussiana, dato che in tal caso la media campionaria $\hat{m}_x = \frac{1}{N} \sum_{i=1}^N x_i$ valuta la somma di gaussiane, anch'essa gaussiana (vedi § 6.5.2); viceversa ($p_X(x)$ qualsiasi) è valida solo per N elevato, perché in tal caso la gaussianità di \hat{m}_x è garantita dal teorema centrale del limite. Altrimenti (N piccolo e x non gaussiana), l'intervallo di confidenza può essere determinato a seguito di una analisi specifica per le diverse d.d.p., che in alcuni casi⁶⁹ permettono soluzioni agevoli, mentre un caso di particolare rilevanza è trattato al § seguente.

Abbiamo altresì osservato come all'aumento di N corrisponda una riduzione dell'intervallo di confidenza: nel caso di v.a. x gaussiana trattato finora, qualora si desideri mantenere l'ampiezza dell'intervallo $d = \theta_2 - \theta_1$ inferiore ad un valore d_M , dovrà risultare $2 \cdot z_{\alpha/2} \frac{\sigma_x}{\sqrt{N}} \leq d_M$, e dunque $N \geq 4 \cdot \sigma_x^2 \cdot \left(\frac{z_{\alpha/2}}{d_M}\right)^2$. Qualora la σ_x non sia nota, cautelativamente si usa al suo posto un multiplo fino a 4-6 volte la variabilità osservata nel campione.

Esempio Una precedente analisi ci informa che la durata x delle telefonate in un call center è approssimativamente una v.a. gaussiana con media m_x incognita e deviazione standard $\sigma_x = 4$ minuti; si calcoli la dimensione minima N del campione necessaria a stimare la durata media m_x entro una precisione d_M pari a due minuti⁷⁰, con confidenza del 95%. Si ottiene allora $N \geq 4 \cdot 16 \left(\frac{1.96}{2}\right)^2 = 61.46$, dunque $N \geq 62$.

6.6.4.1 Intervallo di confidenza per una frazione

Consideriamo ora una v.a. discreta booleana (o *dicotomica*) x che assume il valore 1 o 0 con probabilità p e $q = 1 - p$ rispettivamente, e che può rappresentare un evento di errore, una scelta politica, o una qualunque domanda di un sondaggio, ovvero p rappresenta *la frazione* di una certa caratteristica rispetto al totale dei casi. Dato che il valore medio risulta pari a $m_x = E\{x\} = \sum_{i=0}^1 p_i x_i = q \cdot 0 + p \cdot 1 = p$, la media campionaria $\hat{m}_x = \frac{1}{N} \sum_{i=1}^N x_i$ rappresenta una stimatore \hat{p} di p . Invocando nuovamente una dimensione del campione N sufficientemente elevata, $\hat{m}_x = \hat{p}$ è assimilabile ad una v.a. gaussiana con media p e varianza⁷¹ $\sigma_{\hat{p}}^2 = \frac{p(1-p)}{N}$, che quindi dipende dalla stessa

⁶⁸Oppure dalla tabella presente nella pagina Wikipedia citata prima.

⁶⁹Ad esempio, per la d.d.p. esponenziale (§ 22.2.1) e per quella poissoniana (§ 22.2), vedi Papoulis.

⁷⁰Come dire, minuto più, minuto meno... ;-)

⁷¹Infatti, calcoliamo prima $E\{x^2\} = p \cdot 1^2 + (1-p) \cdot 0^2 = p$ e dunque troviamo $\sigma_x^2 = E\{x^2\} - (E\{x\})^2 = p - p^2 = p(1-p)$; inoltre, tuttora risulta $\sigma_{\hat{p}}^2 = \frac{\sigma_x^2}{N}$.

quantità da stimare; utilizziamo dunque al suo posto $\sigma_{\hat{p}}^2 = \frac{\hat{p}(1-\hat{p})}{N}$. Ci troviamo pertanto nelle condizioni di poter applicare il risultato (6.32) con l'accortezza di sostituire a $\frac{\sigma_x}{\sqrt{N}}$ il termine $\sqrt{\frac{\hat{p}(1-\hat{p})}{N}}$, dando luogo ad un intervallo di confidenza espresso come $(\theta_1, \theta_2) = \hat{p} \pm z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{N}}$ in cui $z_{\alpha/2}$ è il quantile per la gaussiana normalizzata, con il livello di confidenza $1 - \alpha$ desiderato. Osserviamo ora che, a parità di N , l'estensione dell'intervallo risulta maggiore per valori di p attorno allo 0.5, e si riduce fino ad annullarsi per $p \rightarrow 0$ e $p \rightarrow 1$.

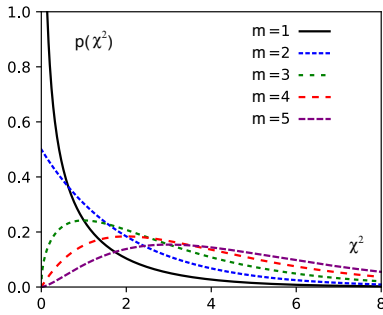
Esempio A seguito di una campagna pubblicitaria viene svolto un sondaggio mediante $N = 200$ interviste chiedendo se il prodotto oggetto della campagna viene ricordato. 85 di queste rispondono affermativamente, determinando una media campionaria pari a $\hat{p} = 85/200 = 0.425$, affetta dalla varianza $\sigma_{\hat{p}}^2 = \sqrt{\frac{\hat{p}(1-\hat{p})}{N}} = \sqrt{\frac{0.425 \cdot 0.575}{200}} = 0.035$. L'intervallo di confidenza al 95% corrisponde pertanto a $(\theta_1, \theta_2) = \hat{p} \pm z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{N}} = 0.425 \pm 1.96 \cdot 0.035 = 0.425 \pm 0.0686$ ovvero dal 36 al 49%, ossia uno *spread* di 13 punti. Se invece le risposte affermative fossero state solo 15, avremo ottenuto $\hat{p} = 15/200 = 0.075$ e dunque $\sigma_{\hat{p}}^2 = \sqrt{\frac{0.075 \cdot 0.925}{200}} = 0.019$, fornendo un intervallo $p = 0.075 \pm 1.96 \cdot 0.019 = 0.075 \pm 0.036$ ovvero dal 3.9 all'11.1%, cioè uno *spread* di 7.2 punti.

Dimensione campionaria Abbiamo osservato come la variabilità $\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{N}}$ per la stima di frazione dipenda, oltre che da N , anche dallo stesso parametro p oggetto di stima. Ci si pone allora nel caso peggiore, ovvero si suppone $p = 0.5$ a cui corrisponde $p(1-p) = 0.25$ e una massima deviazione standard $\sigma_{\hat{p}} = \frac{0.5}{\sqrt{N}}$, da cui $(\theta_1, \theta_2) = \hat{p} \pm z_{\alpha/2} \cdot \frac{0.5}{\sqrt{N}}$; pertanto il valore N minimo che permette di ottenere un intervallo di estensione $d_M = 2z_{\alpha/2} \cdot \frac{0.5}{\sqrt{N}}$ al livello $1 - \alpha$ risulta in prima approssimazione pari a $N \geq \left(\frac{z_{\alpha/2}}{d_M}\right)^2$.

Esempio Si intende svolgere un sondaggio per stimare la proporzione degli elettori che voteranno per un certo partito politico, desiderando un intervallo di confidenza che al livello $1 - \alpha = 0.95$ non superi l'ampiezza di 2 punti percentuali (la cosiddetta *forchetta*), ovvero $d_M = 0.02$. Avremo allora bisogno di un campione composto da almeno $N \geq \left(\frac{z_{\alpha/2}}{d_M}\right)^2 = \left(\frac{1.96}{0.02}\right)^2 = 9604$ soggetti.

6.6.5 Test chi quadro e verifica di distribuzione

Alcuni degli aspetti discussi fino ad ora pongono come condizione quella di avere a che fare con grandezze gaussiane. La tecnica che discutiamo appresso è spesso utilizzata per verificare che la frequenza dei valori osservati rispecchi quella prevista da una distribuzione di probabilità prefissata. A tale scopo, consideriamo una v.a. x di cui sono state collezionate N determinazioni, mediante le quali si costruisce un istogramma (pag. 150) suddividendo il campo di variabilità di x in m intervalli I_i , ed indichiamo con k_i il numero di realizzazioni di x che ricadono entro l'intervallo I_i , $i = 1, 2, \dots, m$, risultando $\sum_{i=1}^m k_i = N$.



| m | 1 | 2 | 5 | 10 | 20 |
|--------------|----------------------|-------|-------|------|-------|
| $1 - \alpha$ | $\chi^2_{\alpha}(m)$ | | | | |
| 0.9 | 0.016 | 0.211 | 1.61 | 4.86 | 12.44 |
| 0.95 | 0.004 | 0.103 | 1.15 | 3.94 | 10.85 |
| 0.99 | 0 | 0.02 | 0.554 | 2.56 | 8.26 |
| 0.995 | 0 | 0.01 | 0.412 | 2.16 | 7.43 |

Figura 6.9: Densità di probabilità χ^2 al variare dei gradi di libertà m e relativi percentili

Il test tenta di verificare l'ipotesi H_0 che la v.a. x abbia effettivamente origine dalla d.d.p. prevista, nel cui caso i valori k_i dovrebbero essere molto prossimi a Np_i , in cui $p_i = \int_{I_i} p_X(x) dx$ è la probabilità che la d.d.p. assegna all'evento $x \in I_i$. Pertanto se l'ipotesi H_0 è vera il valore della nuova v.a.

$$\chi^2 = \sum_{i=1}^m \frac{(k_i - Np_i)^2}{Np_i} \tag{6.34}$$

dovrebbe essere vicino allo zero, mentre valori più elevati (di un valore di soglia λ) suggeriscono di rigettare l'ipotesi. Se N è sufficientemente elevato⁷² i valori k_i sono grandezze gaussiane con valore atteso $m_{k_i} = Np_i$ e varianza $\sigma_{k_i}^2 \simeq (Np_i)^2$ ⁽⁷³⁾, dunque la (6.34) rappresenta la somma dei quadrati di v.a. gaussiane normalizzate, e χ^2 risulta essere una v.a. detta *chi quadro* ad $m - 1$ gradi di libertà⁷⁴, la cui d.d.p. assume l'andamento riportato in figura 6.9.

L'ipotesi sulla d.d.p. di x viene dunque accettata con un intervallo di confidenza $1 - \alpha$ nel caso in cui il valore calcolato per χ^2 risulti inferiore ad una soglia stabilita in base alle tabelle⁷⁵ del percentile $\chi^2_{\alpha}(m - 1)$, che corrisponde ad una probabilità di errore (decidere H_0 quando invece non è vera) pari ad α . Notiamo che, a differenza dei casi precedenti, il percentile si riferisce *all'intera* area α anziché alla sua metà, in quanto l'intervallo di confidenza è comunque delimitato a sinistra dallo zero.

Esempio Una moneta viene lanciata 200 volte, producendo 115 volte testa e 85 croce. Si può ritenere la moneta *non truccata*, con un livello di significatività del 95% ? Evidentemente in questo caso $m = 2$ e $p_1 = p_2 = 0.5$, cosicché il calcolo della (6.34)

⁷²Effettivamente è richiesto anche un numero di osservazioni k_i maggiori di 5-10 per qualunque i ; se ciò non fosse vero, è possibile ridurre il numero m degli intervalli I_i , raggruppando tra loro quelli meno popolati.

⁷³In realtà essendo k_i il numero di casi favorevoli $x \in I_i$ rispetto al totale N , esso ha una d.d.p. *binomiale* (§ 22.1) per la quale $m_{k_i} = Np_i$ e varianza $\sigma_{k_i}^2 = Np_i(1 - p_i)$. Al crescere di N , e dunque degli intervalli m , i termini $1 - p_i$ divengono circa unitari, e la binomiale viene approssimata da una *poissoniana* (§ 22.2), per la quale appunto $\sigma^2 = Np_i$. Tale approssimazione è descritta come test di *Pearson*, vedi https://it.wikipedia.org/wiki/Test_chi_quadrato_di_Pearson

⁷⁴Vedi ad es. https://it.wikipedia.org/wiki/Distribuzione_chi_quadrato. I gradi di libertà sono $m - 1$ anziché m in virtù del vincolo $\sum_i p_i = 1$.

⁷⁵Vedi ad es.

https://it.wikipedia.org/wiki/Distribuzione_chi_quadrato#Tabella_dei_valori_critici

porta a $\chi^2 = \frac{(115-100)^2}{100} + \frac{(85-100)^2}{100} = 4.5$. Per $m - 1 = 1$ grado di libertà la tabella dei percentili fornisce $\chi_{0,05}^2(1) = 0.004$, pertanto dato che $\chi^2 = 4.5 > 0.004 = \chi_{0,05}^2(1)$ il test rigetta l'ipotesi, che può essere ammissibile solo accettando una probabilità di errore α maggiore del 95%, vedi tabella di cui all'ultima nota.

Come ultimo argomento di questa sezione, vediamo come poter ridurre il rumore sovrapposto ad un segnale *ripetuto*.

6.6.6 Stima di forma d'onda

Il problema è quello di voler stimare, anziché un singolo valore (o parametro), una intera forma d'onda, ovvero l'andamento di un segnale limitato nel tempo, di cui si possono disporre di più repliche, ognuna affetta da una fonte di rumore additivo (§ 8.4), oppure qualora la realizzazione del segnale possa essere influenzata da fattori diversi da quelli che si intendono studiare. Un esempio tipico di tale esigenza si ha in campo biomedico, in cui deboli segnali elettrici (dell'ordine dei μV) di natura biologica sono captati mediante elettrodi disposti sul corpo, alla ricerca ad es. di particolarità in un segnale elettrocardiografico semiperiodico⁷⁶, che per sua natura presenta fisiologiche variazioni ed oscillazioni, o nel caso dei *potenziali evocati*⁷⁷ che si producono nelle vie nervose a seguito di stimolazione.

Allo scopo di ridurre per quanto possibile la variabilità dovuta alle cause di disturbo che potrebbero impedire il riscontro di caratteristiche del segnale $s(t)$ di significato diagnostico, le N repliche del segnale osservato $x_i(t) = s(t) + n_i(t)$ con $i = 1, 2, \dots, N$ in cui $n_i(t)$ rappresenta il disturbo ogni volta diverso, vengono *mediate* tra loro producendo

$$\hat{s}(t) = \frac{1}{N} \sum_{i=1}^N [s(t) + n_i(t)] = s(t) + \frac{1}{N} \sum_{i=1}^N n_i(t)$$

Osserviamo subito che se il processo $n(t)$ è a media nulla allora il risultato è *non polarizzato*, in quanto $E\{\hat{s}(t)\} = s(t) + \frac{1}{N} \sum_{i=1}^N E\{n_i(t)\} = s(t)$. Per quanto riguarda la varianza di $\hat{s}(t)$, aggiungendo l'ipotesi di *incorrelazione* (vedi § 7.1.2) tra le possibili coppie (i, j) di realizzazioni del processo $n(t)$ otteniamo un risultato del tutto simile a quello della media campionaria eq. (6.29), ovvero

$$\begin{aligned} \sigma_{\hat{s}(t)}^2 &= E\{(\hat{s}(t) - s(t))^2\} = E\left\{\left(\frac{1}{N} \sum_{i=1}^N n_i(t)\right)^2\right\} = \\ &= \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N E\{n_i(t) n_j(t)\} = \frac{1}{N^2} \sum_{i=1}^N E\{n_i^2(t)\} = \frac{1}{N^2} N \sigma_n^2 = \frac{\sigma_n^2}{N} \end{aligned}$$

e quindi anche in questo caso la varianza di rumore risulta ridotta grazie all'operazione di media.

Assumendo ora che le realizzazioni del disturbo $n(t)$ appartengano ad un processo gaussiano bianco con densità di potenza $P_n(f) = \frac{N_0}{2}$ (vedi § 7.2.4), osserviamo che la corrispondente varianza σ_n^2 risulta dipendere dalla banda W del filtro passa basso che precede l'amplificatore di ingresso allo strumento di acquisizione, ovvero $\sigma_n^2 = N_0 W$ (vedi § 15.4.1). Essendo i segnali di interesse limitati in banda a non più di ~ 60 Hz,

⁷⁶<https://it.wikipedia.org/wiki/Elettrocardiogramma>

⁷⁷https://it.wikipedia.org/wiki/Potenziali_evocati

viene tipicamente usato un filtro con frequenza di taglio pari a 100 Hz. Tale filtro diviene però fonte di correlazione tra realizzazioni di rumore⁷⁸ tra loro vicine, per cui le osservazioni $x_i(t) = s(t) + n_i(t)$ devono essere effettuate facendo intercorrere tra una acquisizione e la seguente un intervallo almeno pari al doppio della durata della $h(t)$.

Per rimanere nel campo biomedico, difficilmente il segnale $s(t)$ è sempre lo stesso per tutte le acquisizioni $x_i(t)$, bensì presenta una variabilità intrinseca legata sia alla durata dell'esame sia ad alee di origine biologica, come ad es. la latenza variabile per le risposte dei potenziali evocati: in tal caso sono necessarie ulteriori tecniche, la cui discussione trascende dagli scopi di questo testo.

6.7 Appendici

6.7.1 Approfondimento sulla gaussiana

Sviluppiamo qui i passaggi relativi ad alcuni risultati riguardanti la v.a. gaussiana monodimensionale presentata al § 6.2.4.

Integrale di Gauss Si dimostra⁷⁹ che $\int_{-\infty}^{\infty} e^{-x^2} dx = \sqrt{\pi}$, da cui $\int_{-\infty}^{\infty} e^{-\frac{z^2}{2}} dz = \sqrt{2\pi}$. Pertanto il fattore di scala $\frac{1}{\sqrt{2\pi}\sigma}$ che compare nell'espressione della gaussiana è necessario per rendere la (6.10) una d.d.p.: infatti con il cambio di variabile $z = \frac{x-m}{\sigma}$ da cui $dx = \sigma \cdot dz$ si ottiene

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-m)^2}{2\sigma^2}} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \frac{\sigma}{\sigma} e^{-\frac{z^2}{2}} dz = \frac{\sqrt{2\pi}}{\sqrt{2\pi}} = 1$$

Calcolo del valor medio Sebbene dal grafico di fig.6.1 sia evidente, data la simmetria di $p_X(x)$, che il suo valor medio sia pari al parametro m che compare nella (6.10), volendo una conferma analitica si può applicare la definizione $E\{x\} = \int_{-\infty}^{\infty} xp_X(x) dx = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} xe^{-\frac{(x-m)^2}{2\sigma^2}} dx$ da cui, operando il cambio di variabile $\frac{x-m}{\sqrt{2}\sigma} = t$ si ottiene $dx = \sqrt{2}\sigma dt$ e quindi

$$E\{x\} = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} (\sqrt{2}\sigma t + m) e^{-t^2} dt = \frac{\sqrt{2}\sigma}{\sqrt{\pi}} \int_{-\infty}^{\infty} te^{-t^2} dt + \frac{m}{\sqrt{\pi}} \int_{-\infty}^{\infty} e^{-t^2} dt = m$$

in quanto il primo termine risulta nullo dato che la funzione integranda è dispari, mentre il secondo termine vale m in quanto $\int_{-\infty}^{\infty} e^{-t^2} dt = \sqrt{\pi}$, vedi *integrale di Gauss* al paragrafo precedente.

⁷⁸ Al § 7.4 si mostra come l'autocorrelazione $R_{nn}(\tau)$ di un processo $n(t)$ che attraversa un filtro divenga pari a $R_{vv}(\tau) = R_{nn}(\tau) * R_{hh}(\tau)$, in cui $R_{hh}(\tau)$ è l'autocorrelazione della risposta impulsiva del filtro.

⁷⁹ Vedi https://it.wikipedia.org/wiki/Integrale_di_Gauss, in cui si fa uso della teoria esposta al § 6.4.

Calcolo della varianza Anche in questo caso applichiamo il cambio di variabile

$\frac{x-m}{\sqrt{2\sigma}} = t$ alla definizione $E\{(x-m)^2\} = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} (x-m)^2 e^{-\frac{(x-m)^2}{2\sigma^2}} dx$ ottenendo

$$\begin{aligned} E\{(x-m)^2\} &= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} 2\sigma^2 t^2 e^{-t^2} \sqrt{2\sigma} dt = \frac{\sigma^2}{\sqrt{\pi}} \int_{-\infty}^{\infty} t \cdot 2te^{-t^2} dt = \\ &= \frac{\sigma^2}{\sqrt{\pi}} \left[-te^{-t^2} \Big|_{-\infty}^{\infty} + \int_{-\infty}^{\infty} e^{-t^2} dt \right] = \frac{\sigma^2}{\sqrt{\pi}} \sqrt{\pi} = \sigma^2 \end{aligned}$$

dove alla seconda riga si è applicata la tecnica di integrazione per parti⁸⁰ in cui prima si è integrato $2te^{-t^2} \rightarrow -e^{-t^2}$ e quindi si è derivato $t \rightarrow 1$. A questo punto si osserva che $-te^{-t^2} \Big|_{-\infty}^{\infty}$ è pari a zero dato che per $t \rightarrow \pm\infty$ il fattore e^{-t^2} decresce più rapidamente di quanto cresca t , mentre anche qui $\int_{-\infty}^{\infty} e^{-t^2} dt = \sqrt{\pi}$.

Momenti centrati di ordine superiore Per una v.a. gaussiana x sussiste la relazione

$E\{(x-m_x)^p\} = \begin{cases} 0 & p \text{ dispari} \\ \sigma^p (p-1)!! & p \text{ pari} \end{cases}$ in cui $n!!$ indica il *semifattoriale*, pari a $n(n-2)(n-4) \dots$ in cui l'ultimo fattore sarà un 2 oppure un 1 a seconda se n è pari o dispari.

Posizione dei flessi I punti di flesso di (6.10) si trovano in corrispondenza dei valori $x = m \pm \sigma$: infatti l'azzeramento della derivata seconda⁸¹ determina la condizione $(x-m)^2 = \sigma^2$ ovvero $x-m = \pm\sigma$.

6.7.2 Teorema centrale del limite

Fornisce la dimostrazione che la somma di v.a. statisticamente indipendenti ed identicamente distribuite, ovvero con d.d.p. qualsiasi ma uguale per tutte, tende ad assumere una d.d.p. gaussiana qualora il loro numero n sia molto elevato, e dunque permette di ritenere valida l'ipotesi di gaussianità per molti fenomeni naturali, come ad es. il rumore termico (§ 8.4.2.1). Indichiamo con m e σ^2 i valori di media e varianza (finita) per ciascuna delle v.a. x_i che prendono parte alla somma, e consideriamo la nuova v.a. $x^S = \sum_{i=1}^n \frac{x_i}{\sqrt{n}}$: il teorema afferma che $x^S \rightarrow N(n \cdot m, \sigma^2)$ qualora $n \rightarrow \infty$.

Per semplificare la dimostrazione conviene definire $y_i = x_i - m$ come nuove v.a. a media nulla e varianza σ^2 , ed affrontare lo studio della v.a. somma $z = \sum_{i=1}^n \frac{y_i}{\sqrt{n}}$. Iniziamo

⁸⁰Si ricorda che la regola di integrazione per parti stabilisce che $\int_a^b [f'(x)g(x)] dx = f(x)g(x) \Big|_a^b - \int_a^b [f(x)g'(x)]$

⁸¹La derivata prima di $p(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-m)^2}{2\sigma^2}}$ risulta pari a $p'(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-m)^2}{2\sigma^2}} \left[-\frac{(x-m)}{\sigma^2} \right]$ e dunque

$$p''(x) = \frac{1}{\sqrt{2\pi}\sigma} \left\{ e^{-\frac{(x-m)^2}{2\sigma^2}} \left[\frac{(x-m)^2}{\sigma^2} \right] + e^{-\frac{(x-m)^2}{2\sigma^2}} - \frac{1}{\sigma^2} \right\} = \frac{1}{\sqrt{2\pi}\sigma^3} e^{-\frac{(x-m)^2}{2\sigma^2}} \left[\frac{(x-m)^2}{\sigma^2} - 1 \right]$$

con il verificare che

$$\sigma_z^2 = E \left\{ \left(\sum_{i=1}^n \frac{y_i}{\sqrt{n}} \right)^2 \right\} = \sum_{i=1}^n E \left\{ \frac{y_i^2}{n} \right\} = \frac{1}{n} \sum_{i=1}^n \sigma^2 = \frac{1}{n} n \sigma^2 = \sigma^2$$

in cui la seconda eguaglianza sussiste in virtù dell'indipendenza statistica⁸² tra le y_i , e la terza in quanto identicamente distribuite. Passiamo ora a calcolare la *funzione caratteristica* (§ 6.2.5) $\Phi_z(\omega)$ di z , ovvero

$$\Phi_z(\omega) = E_z \{ e^{j\omega z} \} = E_z \left\{ e^{j\omega \sum_{i=1}^n \frac{y_i}{\sqrt{n}}} \right\} = \prod_{i=1}^n E_{y_i} \left\{ e^{j\omega \frac{y_i}{\sqrt{n}}} \right\} = \prod_{i=1}^n \Phi_{y_i}(\omega) = [\Phi_y(\omega)]^n \quad (6.35)$$

dove anche per questo caso la terza eguaglianza sussiste in virtù dell'indipendenza statistica tra le y_i , e la quinta in quanto identicamente distribuite, essendo $\Phi_y(\omega)$ la f.c. di ciascuna delle v.a. y_i normalizzate, ovvero $\Phi_y(\omega) = E_y \left\{ e^{j\omega \frac{y}{\sqrt{n}}} \right\} = \int p_Y(y) e^{j\omega \frac{y}{\sqrt{n}}} dy$ in cui $p_Y(y)$ è la d.d.p. comune a tutte le y_i . La dimostrazione procede sostituendo all'esponenziale $e^{j\omega \frac{y}{\sqrt{n}}}$ il relativo sviluppo in serie di potenze $e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!}$ arrestato al secondo ordine ovvero

$$e^x = 1 + x + \frac{x^2}{2} + o\left(\frac{x^2}{2}\right)$$

in cui $o\left(\frac{x^2}{2}\right)$ significa un infinitesimo di ordine superiore (per x molto piccolo) a $\frac{x^2}{2}$, in modo da ottenere

$$\begin{aligned} \Phi_y(\omega) &= E_y \left\{ e^{j\omega \frac{y}{\sqrt{n}}} \right\} = E_y \left\{ \sum_{k=0}^{\infty} \frac{(j\omega y)^k}{(\sqrt{n})^k k!} \right\} = E_y \left\{ 1 + \frac{j\omega y}{\sqrt{n}} + \frac{(j\omega y)^2}{2n} + o\left[\frac{(\omega y)^2}{2n}\right] \right\} = \\ &= 1 + \frac{j\omega}{\sqrt{n}} m_y - \frac{1}{n} \frac{\omega^2 m_y^2}{2} + o\left(\frac{1}{n}\right) = 1 - \frac{1}{n} \frac{\omega^2 \sigma^2}{2} + o\left(\frac{1}{n}\right) \end{aligned} \quad (6.36)$$

in quanto $E_y \{y\} = m_y = 0$ per ipotesi, e $E_y \{y^2\} = m_y^2 = \sigma^2$ per lo stesso motivo. A questo punto non resta che sostituire la (6.36) nella (6.35) per ottenere $\Phi_z(\omega) = [\Phi_y(\omega)]^n = \left[1 - \frac{1}{n} \frac{\omega^2 \sigma^2}{2} + o\left(\frac{1}{n}\right) \right]^n$ e ricordare (!) che $\lim_{n \rightarrow \infty} \left(1 + \frac{\alpha}{n} \right)^n = e^\alpha$, per arrivare a

$$\lim_{n \rightarrow \infty} \Phi_z(\omega) = e^{-\frac{\omega^2 \sigma^2}{2}}$$

che è esattamente pari alla funzione caratteristica (6.13) di una v.a. gaussiana con media nulla e varianza σ^2 , e quindi il teorema è dimostrato.

Qualora le n v.a. i.i.d. x_i abbiano media m e la loro somma avvenga senza il fattore di normalizzazione $\frac{1}{\sqrt{n}}$, la nuova v.a. $x^S = \sum_{i=1}^n x_i$ per $n \rightarrow \infty$ risulta tendere ad una $N(n \cdot m, n \cdot \sigma^2)$.

⁸²Infatti, per il caso semplice di due v.a. y_1 ed y_2 a media nulla si ottiene

$$\sigma_y^2 = E \{ (y_1 + y_2)^2 \} = E \{ y_1^2 + y_2^2 + 2y_1 y_2 \} = E \{ y_1^2 \} + E \{ y_2^2 \} + 2E \{ y_1 y_2 \} = \sigma_{y_1}^2 + \sigma_{y_2}^2$$

in cui la terza eguaglianza è conseguenza della proprietà distributiva dell'integrale (6.7) che definisce il valore atteso, e la quarta discende dal fatto che l'indipendenza statistica implica incorrelazione (§ 7.1.2).

6.7.3 Matrice di covarianza e forma quadratica associata

La matrice di covarianza Σ_x introdotta ai §§ 6.5 e seguenti ha dimensione $n \times n$ con elementi σ_{ij} definiti come

$$\sigma_{ij} = E \left\{ (x_i - m_{x_i}) (x_j - m_{x_j}) \right\}$$

in cui x_i con $i = 1, 2, \dots, n$ sono v.a. e m_{x_i} è la rispettiva media. Se le x_i sono congiuntamente gaussiane, la stessa Σ_x compare nell'espressione (6.26) della relativa d.d.p. Peraltro, Σ_x mantiene il suo significato indipendentemente dalla natura delle v.a. a cui si riferisce, e presenta interessanti proprietà di cui ora discutiamo, e la cui conoscenza può tornare utile in alcune circostanze, come ad esempio nella analisi delle prestazioni del *filtro di Wiener*, vedi pag. 18.4.2.

Simmetria e autovettori Notiamo innanzitutto che Σ_x è una matrice *simmetrica*, in quanto $\sigma_{ij} = \sigma_{ji}$. In questo caso è possibile dimostrare che i suoi autovalori λ sono *reali*⁸³, ed in numero pari al suo rango. Ricordiamo che gli autovalori λ sono definiti come le possibili radici del *polinomio caratteristico* $q(\lambda) = |\Sigma_x - \lambda \mathbf{I}|$ di Σ_x , di grado n ; ad ogni autovalore λ corrisponde un autovettore $\boldsymbol{\gamma}$ ⁽⁸⁴⁾ tale che

$$\Sigma_x \boldsymbol{\gamma} = \lambda \boldsymbol{\gamma} \quad (6.37)$$

Per le matrici simmetriche si dimostra⁸⁵ che gli autovettori $\boldsymbol{\gamma}_i$ e $\boldsymbol{\gamma}_j$ associati ad autovalori *distinti* sono *ortogonali*, ovvero $\boldsymbol{\gamma}_i^\top \boldsymbol{\gamma}_j = 0$. Pertanto è possibile *assemblare* una matrice $\Gamma = [\boldsymbol{\gamma}_1 \boldsymbol{\gamma}_2 \dots \boldsymbol{\gamma}_p]$ di n righe e p colonne pari agli autovettori $\boldsymbol{\gamma}_i$, per la quale risulta⁸⁶

$$\Gamma^\top \Gamma = \Gamma \Gamma^\top = \mathbf{I}, \quad \Gamma^{-1} = \Gamma^\top, \quad \text{e} \quad \Gamma^\top \Sigma_x \Gamma = \Lambda \quad (6.38)$$

in cui Λ è una matrice quadrata $p \times p$ diagonale, con i valori degli autovettori λ_i sulla diagonale. Sia Σ_x che Λ hanno lo stesso determinante⁸⁷, e quindi $\det(\Sigma_x) = \prod_{i=1}^n \lambda_i$, che dunque si annulla se qualche $\lambda_i = 0$; pertanto, il numero $p \leq n$ di autovalori non nulli determina il rango di Σ_x .

⁸³Facciamo uso del *prodotto Hermitiano* definito come $\langle \mathbf{x}, \mathbf{y} \rangle = \bar{\mathbf{x}}^\top \mathbf{y} = \sum_{i=1}^n \bar{x}_i y_i$, in cui la soprallineatura $\bar{}$ rappresenta l'operazione di coniugazione. In generale per matrici e vettori reali risulta $\langle \mathbf{A}\mathbf{x}, \mathbf{y} \rangle = (\mathbf{A}\mathbf{x})^\top \mathbf{y} = \mathbf{x}^\top \mathbf{A}^\top \mathbf{y} = \langle \mathbf{x}, \mathbf{A}^\top \mathbf{y} \rangle$, ma se oltre a ciò \mathbf{A} è simmetrica si ha $\mathbf{A}^\top = \mathbf{A}$ e dunque $\langle \mathbf{A}\mathbf{x}, \mathbf{y} \rangle = \langle \mathbf{x}, \mathbf{A}\mathbf{y} \rangle$. Indicando ora con $\bar{\lambda}$ il coniugato di un autovalore di \mathbf{A} (per assurdo) complesso, possiamo scrivere $\bar{\lambda} \langle \mathbf{x}, \mathbf{x} \rangle = \langle \bar{\lambda} \mathbf{x}, \mathbf{x} \rangle = \langle \mathbf{A}\bar{\lambda} \mathbf{x}, \mathbf{x} \rangle = \langle \mathbf{A}\mathbf{x}, \mathbf{x} \rangle = \langle \mathbf{x}, \mathbf{A}\mathbf{x} \rangle = \langle \mathbf{x}, \lambda \mathbf{x} \rangle = \lambda \langle \mathbf{x}, \mathbf{x} \rangle$, ma dato che $\langle \mathbf{x}, \mathbf{x} \rangle$ è positivo, dovrebbe essere $\lambda = \bar{\lambda}$, il che è impossibile: dunque tutti gli autovalori sono reali.

⁸⁴Gli autovettori si considerano normalizzati, ovvero $\boldsymbol{\gamma}^\top \boldsymbol{\gamma} = 1$, altrimenti ad uno stesso autovalore ne corrisponderebbero infiniti. Gli autovettori sono inoltre definiti a meno di un termine di fase, dato che se $\boldsymbol{\gamma}$ è un autovettore, lo è anche $\boldsymbol{\gamma} e^{j\theta}$ con $0 < \theta < 2\pi$.

⁸⁵Vedi ad es. http://dssm.unipa.it/chiodi/teaching/files/Statistica3_First/MLAmatrix2012.pdf

⁸⁶La prima relazione è conseguenza dell'ortogonalità, la seconda discende dalla prima, e la terza deriva dalla premoltiplicazione di ambo i membri della (6.37) per $\boldsymbol{\gamma}_j^\top$, che produce $\begin{cases} \boldsymbol{\gamma}_i^\top \Sigma_x \boldsymbol{\gamma}_i = \lambda_i & \text{se } i = j \\ \boldsymbol{\gamma}_j^\top \Sigma_x \boldsymbol{\gamma}_i = 0 & \text{se } i \neq j \end{cases}$

⁸⁷In quanto $\det(\Sigma_x) = \det(\Gamma) \det(\Lambda) \det(\Gamma^\top)$, e $\det(\Gamma) = \det(\Gamma^\top) = \det(\Gamma^{-1}) = 1$.

Forma quadratica e segno Una *forma quadratica* individua l'espressione

$$Q(\mathbf{c}) = \sum_{i=1}^n \sum_{j=1}^n c_i c_j \sigma_{ij} = \mathbf{c}^T \Sigma_x \mathbf{c} \quad (6.39)$$

che rappresenta il valore di un polinomio *omogeneo* di secondo grado nelle variabili c_i . Qualora per qualsiasi scelta di $\mathbf{c} \neq \mathbf{0}$ risulti $Q(\mathbf{c}) \geq 0$, sia Σ_x che $Q(\mathbf{c})$ vengono dette di segno *semidefinito positivo*⁸⁸.

Condizione necessaria e sufficiente per avere $Q(\mathbf{c}) \geq 0$ è che Σ_x sia una *matrice di covarianza*: infatti, la varianza σ_y^2 (non negativa per definizione) di una combinazione lineare $y = \sum_{i=1}^n c_i x_i$ di v.a. x_i può essere espressa⁸⁹ in notazione matriciale appunto come $\sigma_y^2 = \mathbf{c}^T \Sigma_x \mathbf{c}$. Inoltre, se $\det(\Sigma_x) \neq 0$ ovvero Σ_x ha rango pieno⁹⁰, sia Σ_x che $Q(\mathbf{c})$ sono *definite positive*, ovvero $Q(\mathbf{c}) > 0$, in quanto in tal caso tutti gli n autovalori λ_i sono *positivi*⁹¹. Se viceversa $\det(\Sigma_x) = 0$ il rango di Σ_x risulta $p < n$, con $n - p$ autovalori nulli, e sia Σ_x che $Q(\mathbf{c})$ sono *semidefinite positive*.

Analisi della convessità La ricerca di un punto *di minimo* \mathbf{c}_m per $Q(\mathbf{c})$ tale che $Q(\mathbf{c}_m) < Q(\mathbf{c})$ è molto semplice. Occorre innanzitutto verificare che in tale punto si annulli il vettore gradiente $\mathbf{g}(\mathbf{c})|_{\mathbf{c}=\mathbf{c}_m}$ con elementi $g_i(\mathbf{c}) = \frac{\partial Q(\mathbf{c})}{\partial c_i} = 2\Sigma_x \mathbf{c}$, che corrisponde ad impostare un sistema omogeneo $\Sigma_x \mathbf{c} = \mathbf{0}$, che come noto ammette soluzioni $\neq \mathbf{0}$ solo se $\det(\Sigma_x) = 0$. Quindi, se Σ_x non ha autovalori nulli ed è definita positiva, $\mathbf{c}_m = \mathbf{0}$ è l'unico punto singolare.

Per stabilire se si tratti di un minimo od un massimo, occorre valutare il segno della matrice *Hessiana* $\mathbf{H}(\mathbf{c})$ con elementi $h_{ij}(\mathbf{c}) = \frac{\partial^2 Q(\mathbf{c})}{\partial c_i \partial c_j}$: si dimostra che se $\mathbf{H}(\mathbf{c}_m)$ è *definita positiva* allora $Q(\mathbf{c})|_{\mathbf{c}=\mathbf{c}_m}$ è *convessa*⁹², e \mathbf{c}_m corrisponde ad un minimo, oppure ad un massimo se $\mathbf{H}(\mathbf{c}_m)$ è definita negativa. Ma dato che per $Q(\mathbf{c})$ risulta $\mathbf{H}(\mathbf{c}) = 2\Sigma_x$ indipendentemente da \mathbf{c} , il segno di \mathbf{H} è lo stesso di Σ_x , e dunque se quest'ultima è definita positiva, $Q(\mathbf{c})$ è *convessa* per ogni \mathbf{c} , e $\mathbf{c}_m = \mathbf{0}$ corrisponde ad un *minimo globale*.

Nel caso di $n = 2$ è possibile classificare $Q(\mathbf{c})$ come un *paraboloide ellittico*⁹³,

⁸⁸Vedi https://it.wikipedia.org/wiki/Matrice_definita_positiva

⁸⁹
$$\begin{aligned} \sigma_y^2 &= E \left\{ (y - m_y)^2 \right\} = E \left\{ \left(\sum_{i=1}^n c_i x_i - \sum_{i=1}^n c_i m_i \right)^2 \right\} = \\ &= \sum_{i=1}^n \sum_{j=1}^n c_i c_j E \{ x_i x_j \} - \sum_{i=1}^n \sum_{j=1}^n c_i c_j m_i m_j = \sum_{i=1}^n \sum_{j=1}^n c_i c_j \sigma_{ij} \end{aligned}$$

⁹⁰Ossia nessuna tra le v.a. x_i presenta dipendenza *lineare* da una o più altre.

⁹¹Tenendo infatti conto che dalla (6.38) si ottiene $\Sigma_x = \Gamma \Lambda \Gamma^T$, possiamo scrivere $Q(\mathbf{c}) = \mathbf{c}^T \Sigma_x \mathbf{c} = \mathbf{c}^T \Gamma \Lambda \Gamma^T \mathbf{c}$, che ponendo $\mathbf{d} = \Gamma^T \mathbf{c}$ riscriviamo ancora come $Q(\mathbf{c}) = \mathbf{d}^T \Lambda \mathbf{d} = \sum_{i=1}^p \lambda_i d_i^2$. Se qualche λ_i fosse negativo o nullo, si potrebbe trovare un vettore \mathbf{d} nullo tranne per l'unica componente corrispondente al $\lambda_i \leq 0$, e produrre una $Q(\mathbf{c}) \leq 0$, in contrasto con l'ipotesi. Pertanto è vero anche il viceversa, cioè Σ_x è definita positiva se $\lambda_i > 0 \quad \forall i$.

⁹²Vedi https://it.wikipedia.org/wiki/Funzione_convessa. La condizione sulla matrice Hessiana definita positiva è *analoga* alla proprietà nota per la derivata seconda di una funzione monovariata, ma per una dimostrazione si può visitare ad es.

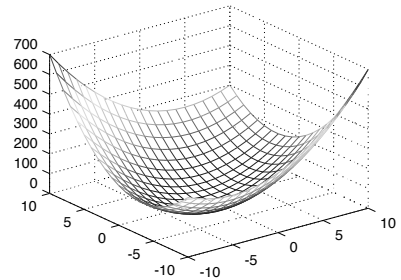
<http://www.statistica.unimib.it/utenti/matematica/AM2/appunti/conv.pdf>.

⁹³Vedi <https://it.wikipedia.org/wiki/Quadrica>. In particolare, la proprietà di una matrice definita positiva di avere n autovalori positivi è quella che in due dimensioni determina questo risultato,

raffigurato nella figura a lato per il caso

$$\Sigma_x = \begin{bmatrix} 2 & -1 \\ -1 & 3 \end{bmatrix}$$

ovvero $Q(\mathbf{c}) = z = 2x^2 - 2xy + 3y^2$. La superficie si dice ellittica dato che una sua intersezione con il piano $Q(\mathbf{c}) = cost$ individua un *iper-ellissoide*.



Interpretazione geometrica Notiamo ora che $Q(\mathbf{c}) = \mathbf{c}^T \Sigma_x \mathbf{c}$ può essere riscritta in forma di prodotto scalare $\langle \Sigma_x \mathbf{c}, \mathbf{c} \rangle$ tra \mathbf{c} e $\mathbf{d} = \Sigma_x \mathbf{c}$, il cui valore risulta $|\mathbf{c}| |\mathbf{d}| \cos \alpha$, ovvero positivo o negativo a seconda se \mathbf{c} e \mathbf{d} condividono o meno lo stesso semipiano, oppure nullo se sono ortogonali.

In definitiva, una matrice Σ_x definita positiva individua una trasformazione che mappa un *qualunque* vettore \mathbf{c} in uno $\mathbf{d} = \Sigma_x \mathbf{c}$ che giace nello *stesso semipiano*, o in quello opposto per una matrice definita *negativa*.

L'opera

Trasmissione dei Segnali e Sistemi di Telecomunicazione

è il risultato di un progetto ventennale di cultura libera, aggiornato di continuo ed evolutosi fino alla forma attuale. La sua disponibilità pubblica è regolata dalle norme di licenza CREATIVE COMMONS

*Attribuzione - Non commerciale -
Condividi allo stesso modo*



<https://creativecommons.org/licenses/by-nc-sa/4.0/deed.it>

e tutte le risorse relative al testo sono accessibili presso

<https://teoriadeisignali.it/libro/>

Puoi contribuire al suo successo promuovendone la diffusione e supportarne lo sviluppo attraverso una donazione, in buona parte devoluta ai progetti *open source*¹ che ne hanno resa possibile realizzazione e divulgazione. Ai donatori viene accordato un accesso *vitalizio* al formato PDF *navigabile* di tutte le edizioni presenti *e future*.

1

- . Lyx - <http://www.lyx.org/>
- . L^AT_EX - <https://www.latex-project.org/>
- . TeX Users Group - <https://tug.org/>
- . Inkscape - <http://www.inkscape.org/>
- . Gnuplot - <http://www.gnuplot.info/>
- . Octave - <http://www.gnu.org/software/octave/>
- . Geany - <https://www.geany.org/>
- . Linux - <https://www.linux.it/>
- . Free Software Foundation - <https://shop.fsf.org/>
- . GNOME Foundation - <https://www.gnome.org/>
- . Mozilla Foundation - <https://www.mozilla.org/it/>
- . Wikipedia - <https://it.wikipedia.org>
- . Internet Archive - <https://archive.org/about/>
- . Creative Commons - <https://creativecommons.it/chapterIT/>
- . WordPress - <https://it.wordpress.org/>
- . Phplist - <https://www.phplist.org/>