

Capitolo 7

Processi aleatori

7.1 Premessa

L'incertezza è inerente in tutti i fenomeni naturali, ma il livello di incertezza varia. Quando questo livello è relativamente piccolo, i problemi associati possono essere trattati in modo deterministico, cioè possono essere formulati in termini di medie, trascurando le variazioni intorno alla media. Data l'intrinsecità delle incertezze in tutti i problemi, l'effetto delle variazioni intorno alla media vengono tenute in considerazioni nell'approccio deterministico mediante concetti come il fattore di sicurezza, che riflettono una stima grossolana delle incertezze. In problemi in cui il livello di incertezza è elevato o si necessita un alto grado di precisione, l'approccio probabilistico fornisce un metodo più razionale e realistico. Esso consente una trattazione sistematica della incertezza in termini quantitativi.

Le prime teorie che sfruttarono razionalmente e quantitativamente questi concetti probabilistici furono:

- La teoria dei giochi;
- La teoria dei moti browniani (Einstein, 1905);

e successivamente:

- La teoria delle comunicazioni per trattare i problemi di rumore nei segnali (Rice, Middleton; 1945-1960).

Nell'analisi dei fenomeni fluidodinamici (come in effetti in ogni altro fenomeno fisico) ci si trova ad analizzare segnali che non hanno caratteristiche deterministiche, che cioè presentano "a priori" incertezza e aleatorietà circa l'andamento effettivo. Tra i principali problemi che si affrontano quanto si fa della sperimentazione c'è quello di analizzare risultati random, cioè non prevedibili a priori, e quello di dover ottenere informazioni significative da misurare affette da "rumore", cioè con alte componenti casuali: quando il risultato si presenta come una funzione (del tempo o della frequenza), la "teoria delle funzioni aleatorie", o dei "processi random", consente di ricavare le suddette insperate informazioni, ottenendo quindi stime corrette e affidabili dei valori delle grandezze in gioco.

In generale si può dire che la non ripetibilità delle misure fatte con la stessa configurazione di prova, ed attribuibile all'interazione di più cause che interagiscono in vario modo con il fenomeno in studio, può essere vista come "fluttuazione statistica" per cui i risultati ottenuti debbono essere trattati seguendo le indicazioni e i metodi proposti dalla teoria delle probabilità e dalla teoria delle funzioni aleatorie.

7.2 Cenni sul calcolo delle probabilità

7.2.1 Eventi

Spesso nel linguaggio comune, o anche tecnico, si usano definire attraverso frasi o relazioni matematiche o in qualsiasi altro modo, determinati *eventi* il cui verificarsi o meno non è a priori noto. Per esempio la frase:

“uscirà il n° 27 alla prossima estrazione del lotto sulla ruota di Bari”

identifica un evento sul cui verificarsi sussistono a priori incertezze.

Nel linguaggio comune si usa dire che tali eventi sono più o meno probabili. Lo scopo che si vuole raggiungere nel presente paragrafo è di individuare il significato ed il valore numerico da attribuire a frasi del tipo:

“probabilità che un certo evento si verifichi”

Per poter dare significato oggettivo alla frase suddetta è necessario preliminarmente:

- a) poter individuare in modo preciso un “esperimento (o fenomeno) aleatorio” di cui non si conosce a priori il risultato effettivo, ma di cui sia possibile individuare l’insieme dei risultati possibili e rispetto al quale l’evento considerato sia una qualificazione, mediante proprietà o attributi, del risultato.

In altri termini, l’evento in oggetto deve potersi pensare costituito da una elencazione di attributi o proprietà a cui il risultato di un certo ben determinato esperimento aleatorio può più o meno soddisfare; fatto l’esperimento, si potrà allora dire se l’evento in oggetto si è o meno verificato, a seconda che il risultato soddisfi o meno tutte le proprietà dichiarate dell’evento.

- b) immaginare di poter ripetere molte volte (al limite, infinite volte) l’esperimento aleatorio con tutte le condizioni che lo qualificano immutate.

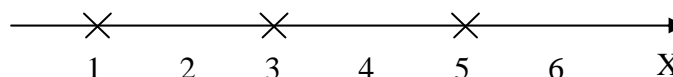
Nel presente paragrafo prenderemo in esame solo quelli esperimenti aleatori per i quali ciascun risultato sia rappresentabile mediante un numero reale e quindi mediante un punto di un asse coordinato avente tale numero come ascissa.

I seguenti esempi cercano di chiarire quanto finora esposto.

Esempio 1:

Si consideri l’evento: “Uscita di un numero dispari nel lancio di un dado a 6 facce”

Tale evento fa riferimento al fenomeno aleatorio costituito dal lancio di un dado a 6 facce; possibili risultati sono i numeri interi da 1 a 6, quindi rappresentabili su una retta mediante punti di ascisse 1,2,...,6 (vedi figura seguente).



Possibili risultati

× Risultati favorevoli all’evento dell’esempio

L’evento in oggetto si verifica se il risultato del lancio è 1,3, oppure 5; non si verifica negli altri casi.

Esempio 2:

Si consideri l’evento: “Sortita di un picche da un mazzo di 40 carte”

Tale evento fa riferimento al fenomeno aleatorio costituito dalla estrazione di una carta da un mazzo contenente 10 carte di cuori, 10 di quadri, 10 di picche e 10 di fiori. I possibili risultati sono 40, e non rappresentati da valori numerici. Tuttavia è possibile numerare tutte le carte da 1 a 40 con una certa regola e in questo modo rendere numerico il risultato.

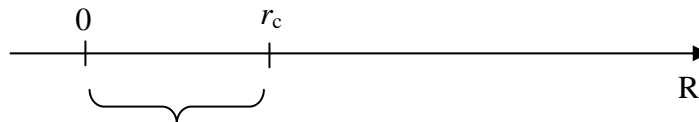
Ad esempio, numeriamo da 1 a 10 le carte di cuori cominciando dall’asso e finendo con il re, da 11 a 20 le carte di quadri, dal 21 al 30 quelle di fiori e dal 31 al 40 quelle di picche, con lo stesso ordine per quanto riguarda il valore. Con questa associazione, il risultato è

stato reso numerico; inoltre l'evento considerato si verifica se il risultato è compreso tra 31 e 40, non si verifica nel caso opposto.

Esempio 3:

Si consideri l'evento: "Colpire il cerchio centrale di un determinato bersaglio"

In questo caso il fenomeno aleatorio è costituito da una certa persona che, con una determinata arma, e a determinata distanza, cerca di colpire il "centro" di un bersaglio (di cui il cerchio centrale abbia raggio r_c). Come risultato del fenomeno aleatorio si può assumere la distanza del colpo dal centro del bersaglio.



Insieme dei risultati favorevoli all'evento

Indicando con r tale distanza è chiaro che l'insieme dei possibili risultati è costituito dall'insieme di tutti i possibili valori di r e cioè dall'insieme dei numeri reali positivi (0 incluso), ed è quindi associabile a tutti i punti di una semiretta (vedi figura precedente). Se il risultato soddisfa la relazione:

$$0 \leq r \leq r_c$$

l'evento considerato si è verificato; se non la soddisfa non si è verificato.

In quanto mostrato precedentemente si è cercato di puntualizzare e concretizzare alcuni concetti di base nel calcolo delle probabilità. Conviene ora procedere in maniera più precisa, e per cominciare si formalizzeranno, mediante definizione, alcuni termini. Si definisce:

Fenomeno aleatorio: *la descrizione opportunamente dettagliata di una procedura di natura qualsiasi, che dà luogo, se attuata, a uno e uno solo tra un insieme di possibili risultati. Il risultato effettivamente ottenuto si deve ritenere non noto a priori (e cioè prima dell'effettuazione dell'esperimento), mentre si deve ritenere noto l'insieme dei possibili risultati ottenibili.*

Si considereranno nel seguito del presente paragrafo solo quei fenomeni aleatori i cui risultati sono rappresentabili, o direttamente o con opportuni artifici, mediante numeri reali, o equivalentemente mediante punti su un asse coordinato aventi tali numeri come ascisse.

Si definisce:

Variabile aleatoria: *il generico valore numerico, a priori non noto, assunto dal risultato di un fenomeno aleatorio. Le variabili aleatorie verranno indicate con lettere maiuscole (X, per esempio).*

Si definisce:

Determinazione: *il valore numerico assunto da una variabile aleatoria in seguito ad una attualizzazione del fenomeno aleatorio. Essa verrà indicata con una lettera minuscola, eventualmente con indice (x_j , per esempio).*

La variabile aleatoria non è un numero; i valori che di volta in volta assume, a seguito dell'attuazione del fenomeno aleatorio, prendono un nome diverso, chiamandosi determinazioni. Il corrispondente geometrico della variabile aleatoria è l'insieme di punti su un asse coordinato che corrispondono a tutti i possibili risultati dell'esperimento; per questo motivo tale asse viene di solito indicato con lo stesso simbolo della variabile aleatoria, mentre i punti rappresentativi dei possibili risultati (che sono chiaramente il corrispondente geometrico delle determinazioni) verranno indicati con la corrispondente lettera minuscola, eventualmente con indice.

Per esempio, nell'esempio 1 la variabile aleatoria dell'esperimento, detta X, può assumere 6 determinazioni; risulterà, indicando con x_1, x_2, x_3, x_4, x_5 e x_6 tali determinazioni: $x_1=1, x_2=2, \dots, x_6=6$. Nell'esempio 2 la variabile aleatoria dell'esperimento, detta Y, può assumere 40 determinazioni, che possono indicarsi con y_1, y_2, \dots, y_{40} . Nell'esempio 3, infine, la variabile aleatoria dell'esperimento, detta Z, può assumere un'infinità non numerabile di determinazioni ciascuna indicata con z.

Nei primi due esempi la variabile aleatoria viene detta *discreta*, nel senso che può assumere un numero discreto (finito o infinito, in generale) di determinazioni, mentre nel terzo esempio viene detta *continua*.

Preferibilmente, le determinazioni di una variabile aleatoria discreta, che sono finite o infinite, si indicheranno mediante lettere minuscole con un indice intero, mentre le determinazioni di una variabile aleatoria continua verranno indicate mediante lettere minuscole senza indice.

Cerchiamo ora, nel contesto che andiamo delineando, di collocare il concetto di evento.

Come già detto in precedenza, si definisce:

Evento: *l'affermazione di una proprietà o attributo del risultato di un fenomeno aleatorio, e quindi della variabile aleatoria ad esso associata.*

E' dunque possibile stabilire se ogni determinazione soddisfi o meno la proprietà o attributo di un generico evento (ovvero, più propriamente, se tale determinazione è *favorevole* o *sfavorevole* all'evento); da questa osservazione discende che ogni evento individua un insieme di determinazioni costituite da tutte e sole quelle ad esso favorevoli e reciprocamente ogni insieme di determinazioni individua un particolare evento a cui esse (e solo esse) sono favorevoli.

In altre parole, vi è una corrispondenza biunivoca tra eventi ed insiemi di determinazioni, nel senso che ad ogni evento corrisponde un insieme di determinazioni, e viceversa.

Per esempio, l'evento di cui all'esempio 1 individua ed è individuato dall'insieme di determinazioni x_1, x_2, x_3 ; l'evento di cui all'esempio 2 individua ed è individuato dalle determinazioni $y_{31}, y_{32}, \dots, y_{40}$; infine l'evento di cui all'esempio 3 individua ed è individuato dal segmento $0-r_c$.

7.2.2 Probabilità di un evento descritto da una variabile aleatoria discreta

A questo punto siamo finalmente in grado di definire e valutare la probabilità di un certo evento ε_x , che verrà indicata con il simbolo $\Pr\{\varepsilon_x\}$.

Allo scopo, supponiamo di ripetere un certo numero di volte l'esperimento aleatorio su cui è definito l'evento ε_x e indichiamo con N_ε il numero di volte in cui si è verificato un risultato favorevole all'evento ε_x ; è un fatto osservabile sperimentalmente che in ogni fenomeno aleatorio ben definito, il rapporto N_ε / N tende ad apparire sempre meno soggetto a variabilità quando N aumenta, tanto che al limite, per $N \rightarrow \infty$, può assumersi che il suddetto rapporto tenda sempre ad un valore limite definito; sulla base di questa osservazione, è possibile definire come *probabilità dell'evento* la quantità:

$$\Pr\{\varepsilon_x\} = \lim_{N \rightarrow \infty} \frac{N_\varepsilon}{N} \geq 0$$

Osservazione:

Ciò è vero e possibile solo se le condizioni dell'esperimento sono indicate in modo esauriente. Tale condizione, sebbene vaga, è tuttavia importantissima ai fini della definizione della probabilità di un certo evento. Per esempio, è intuitivo che ripetendo continuamente il lancio di un dado a struttura fisica uniforme, il numero di volte che esce un valore, il 2 ad esempio, tende ad essere pari a 1/6 del numero di prove, ma questo è vero solo se il dado conserva nel tempo la sua omogeneità di struttura.

E' possibile esprimere la quantità precedente in termini di grandezze più facilmente misurabili o calcolabili.

Cominceremo con il considerare il caso che X sia una variabile aleatoria discreta, e supponiamo che l'evento ε_x sia semplicemente l'affermazione che si verifichi la determinazione x_j (prefissata ma generica) per la variabile aleatoria X e ciò, simbolicamente:

$$\varepsilon_x: \quad X = x_j$$

Allora, per definizione, la **probabilità** di tale evento ha la seguente espressione:

$$\Pr\{X = x_j\} \doteq P_X(x_j) = \lim_{N \rightarrow \infty} \frac{N_j}{N} \geq 0$$

in cui N_j rappresenta il numero di volte in cui si è verificato il risultato x_j sul totale delle ripetizioni dell'esperimento. Alcuni autori (Von Mises ed altri) considerano questa relazione come la definizione (frequentistica) della probabilità. A noi serve solo per connettere tra loro il punto di vista teorico con quello sperimentale. Un'altra definizione (classica) fa ricorso al concetto di "ugualmente probabile": "se un evento può capitare in N modi distinti e ugualmente probabili, e se n di questi modi hanno un attributo A , allora la probabilità di A è n/N ".

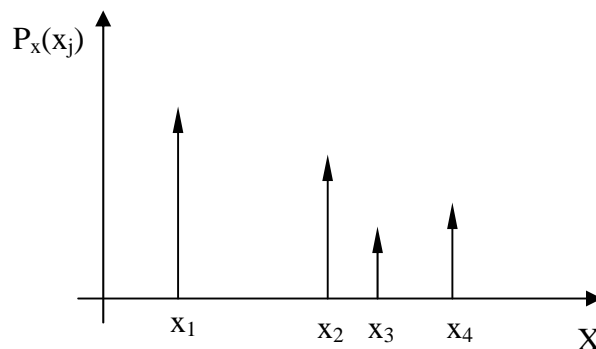
La probabilità risulta una funzione della determinazione x_j : per questo motivo essa prende il nome di **funzione di probabilità** della variabile aleatoria discreta X .

Per mettere in evidenza il fatto che:

1. l'evento considerato è il semplice verificarsi di una determinata x_j
2. che la variabile aleatoria, di cui x_j è una determinazione, è X ,

alla probabilità si usa dare il simbolo particolare $P_X(x_j)$.

La funzione ora definita (un esempio è rappresentato nella seguente figura) è chiaramente sempre positiva o nulla e minore o uguale a 1, cioè



Inoltre, poiché per ogni N , la somma di tutti gli N_i relativi alle determinazioni x_i è pari a N , risulta vera la seguente proprietà:

$$\sum_j P_X(x_j) = \frac{1}{N} \sum_j N_j = 1$$

Sono utili a questo punto le seguenti osservazioni:

- a) Se la probabilità di una determinazione x_k è nulla, vuol dire che il numero di volte in cui tale determinazione si è verificata è trascurabilmente piccolo. Al limite: x_k si è verificato un numero finito di volte su un numero infinito di esperimenti. In questo caso la determinazione x_k si dice **impossibile**.
- b) Se la probabilità di una determinazione x_k è pari a 1 (per cui tutte le altre determinazioni hanno probabilità zero), vista l'equazione precedente, vuol dire che il numero di volte in cui tale determinazione non si è verificata è trascurabilmente piccolo (nel senso prima detto). In questo caso la determinazione x_k si dice **certa**.

La definizione usata per la funzione di probabilità della variabile aleatoria X è operativa, nel senso che prevede la valutazione della funzione di probabilità sulla base dei risultati di un numero molto grande di esperimenti. Non sempre è però necessario procedere in questo modo: in molte situazioni si verifica la circostanza che tutte le possibili determinazioni hanno intuitivamente la stessa probabilità, per motivi di simmetria e perché non vi sono ragionevoli motivi per ritenere che sia diversamente. In queste situazioni (verificate per es. negli esempi 1 e 2), detto n il numero di determinazioni distinte, risulterà:

$$P_X(x_j) = \frac{1}{n} \quad \text{per } j = 1, 2, \dots, n$$

Consideriamo ora il caso in cui la variabile aleatoria X è ancora discreta ma ε_x è un generico evento, per cui la sua probabilità è data dalla

$$\Pr\{\varepsilon_x\} = \lim_{N \rightarrow \infty} \frac{N_\varepsilon}{N} \geq 0.$$

Indicando con X_ε l'insieme delle determinazioni ad esso favorevoli, risulterà che il numero N_ε di risultati favorevoli ad ε_x in N ripetizioni del fenomeno è pari alla somma dei numeri N_j dei risultati favorevoli alle determinazioni x_j purché $x_j \in X_\varepsilon$; risulterà pertanto:

$$\Pr\{\varepsilon_x\} \doteq \lim_{N \rightarrow \infty} \frac{N_\varepsilon}{N} = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{x_j \in X_\varepsilon} N_j = \sum_{x_j \in X_\varepsilon} \lim_{N \rightarrow \infty} \frac{N_j}{N}$$

da cui, ricordando la definizione di probabilità:

$$\Pr\{\varepsilon_x\} = \sum_{x_j \in X_\varepsilon} P_X(x_j)$$

In altre parole, l'equazione precedente mostra che la funzione di probabilità $P_X(x_j)$ consente di valutare la probabilità di un qualsiasi evento definito sul fenomeno aleatorio di cui X costituisce la relativa variabile aleatoria.

Esercizio 1: Si calcoli la probabilità che, nell'estrazione di un numero da un'urna contenente 90 numeri (da 1 a 90), venga ottenuto un numero tale che la somma delle due cifre sia 8.

Per la risoluzione dell'esercizio proposto, così come in tutti i problemi inerenti al calcolo della probabilità di un certo evento, conviene procedere come segue:

a) *Descrizione del fenomeno aleatorio su cui è definito l'evento.*

Nel caso attuale il fenomeno aleatorio consiste nell'estrazione a caso di un numero da un'urna contenente 90 numeri, da 1 a 90.

b) *Individuazione della variabile aleatoria associata e di tutte le sue determinazioni.*

Nel caso attuale, la variabile aleatoria associata al fenomeno è costituita dal numero estratto. Essa quindi è discreta e ha 90 determinazioni ciascuna associata ad un numero intero compreso tra 1 e 90. Possiamo indicare con x_j [$j = 1, \dots, 90$] tali determinazioni: con $x_j = j$.

c) *Valutazione della funzione di probabilità.*

Nel caso attuale, le singole determinazioni appaiono intuitivamente equiprobabili, per cui è possibile applicare la

$$P_X(x_j) = \frac{1}{n} \Rightarrow P_X(x_j) = \frac{1}{90} \quad [j = 1, 2, \dots, 90]$$

d) *Individuazione dell'insieme delle determinazioni favorevoli all'evento.*

Nel caso attuale, l'evento di cui sopra individua ed è individuato da quelle determinazioni in cui la somma delle cifre dia come risultato 8; è ovvio allora che tali determinazioni siano $x_8, x_{17}, x_{26}, x_{35}, x_{44}, x_{53}, x_{62}, x_{71}, x_{80}$.

e) *Applicazione della* $\Pr\{\mathcal{E}_x\} = \sum_{x_j \in X_e} P_X(x_j)$

Nel caso attuale si ha :

$$\Pr\{\mathcal{E}_x\} = P_X(x_8) + \dots + P_X(x_{80}) = 9/90 = 0.1$$

Si osservi che, se tutte le determinazioni hanno uguale probabilità, la probabilità di un evento risulta avere la seguente espressione:

$$\Pr\{\mathcal{E}_x\} = \frac{\text{numero di determinazioni favorevoli}}{\text{numero di determinazioni possibili}}$$

facilmente deducibile dalla $P_X(x_j) = \frac{1}{n}$, e che è molto utile nelle applicazioni, in quanto evita di individuare singolarmente le determinazioni favorevoli all'evento, essendo sufficiente conoscerne il numero.

Esercizio 2: Si calcoli la probabilità che nell'estrazione di due carte da un mazzo di 40 carte si ottenga una coppia (due carte dello stesso valore).

Nella risoluzione dell'esercizio si seguirà la procedura indicata nel precedente.

a) Il fenomeno aleatorio è costituito dall'estrazione contemporanea, a caso, di due carte da un mazzo di 40 carte.

b) Il numero delle possibili determinazioni è dato dal numero di combinazioni di due oggetti in 340 posti senza ripetizioni, e cioè vale:

$$\binom{40}{2} = \frac{40 \times 39}{2} = 780$$

Non viene esplicitata la corrispondenza tra ogni determinazione e ogni possibile risultato in quanto non necessaria alla risoluzione del problema.

c) Poiché tutte le possibili determinazioni sono equiprobabili, si ha:

$$P_X(x_j) = \frac{1}{780}$$

d) Il numero di determinazioni favorevoli all'evento in oggetto è pari a $10 \times (4 \times 3) / 2$ in quanto vi sono 10 possibili coppie di valori coincidenti (asso con asso, re con re, ecc.) e per ciascuna di esse vi possono essere 4×3 accoppiamenti di seme (ciascuno dei 4 semi con uno dei rimanenti 3), di cui a coppie identificanti lo stesso evento (per esempio, le combinazioni "asso di cuori e asso di fiori" e "asso di fiori e asso di cuori" identificano lo stesso risultato favorevole).

e) Applicando la

$$\Pr\{\varepsilon_x\} = \frac{\text{numero di determinazioni favorevoli}}{\text{numero di determinazioni possibili}} \Rightarrow P_X\{\varepsilon_x\} = \frac{60}{780} = \frac{1}{13}$$

7.2.3 Probabilità di un evento descritto da una variabile aleatoria continua

E' evidente che applicando la definizione, la probabilità che una variabile aleatoria continua assuma un particolare valore è identicamente nulla (Attenzione: probabilità nulla non implica l'impossibilità dell'evento!). Di conseguenza non ha senso utilizzare la precedente definizione per eventi descritti da una variabile continua. Al fine di caratterizzare la probabilità anche in questo caso, si introduce una nuova grandezza, la funzione di ripartizione, valida sia per variabili discrete che per variabili continue.

a) Funzione di ripartizione:

Sia X una variabile casuale discreta o continua: definiamo la funzione di ripartizione $F(x)$ come:

$$F(x) = P(X \leq x)$$

La $F(x)$ rappresenta dunque la probabilità che la variabile casuale X assuma un valore non maggiore di x . Conoscendo la funzione $F(x)$ si può calcolare la probabilità che la variabile casuale X assuma un valore compreso tra due limiti a e b :

$$P(a < X \leq b) = F(b) - F(a)$$

Poiché la probabilità è una quantità definita positiva, la funzione $F(x)$ è una funzione non decrescente della variabile x .

Siano c e d i limiti di variabilità della variabile casuale X ; valgono le seguenti relazioni limite:

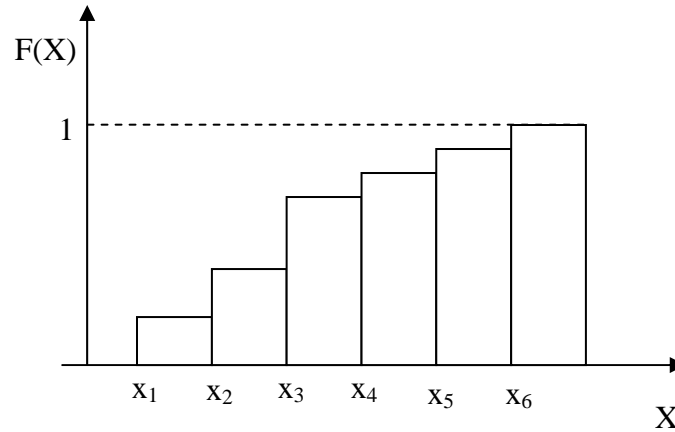
$$\lim_{x \rightarrow c} F(x) = 0 \qquad \lim_{x \rightarrow d} F(x) = 1$$

cioè la probabilità che la variabile casuale X sia uguale ad un certo valore a è uguale alla discontinuità della funzione di ripartizione nel punto a . Quindi se la funzione di ripartizione è continua, si ha:

$$P(X = a) = 0$$

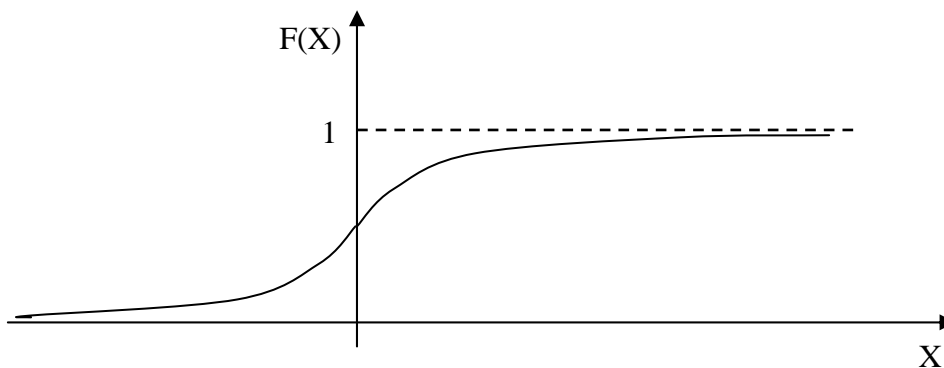
Si può dimostrare che l'insieme dei punti di discontinuità della funzione di ripartizione è numerabile.

La funzione di ripartizione di una variabile casuale discreta è una funzione a gradini con salti p_i nei punti x_i , essendo p_i la probabilità del valore x_i (vedi figura seguente).



b) Funzione densità di probabilità:

Una variabile casuale è detta continua quando la sua funzione di ripartizione è continua e differenziabile in ogni punto dell'asse reale (vedi figura seguente).



La derivata:
$$p(x) = F'(x) = \frac{dF}{dx}$$

Della funzione di ripartizione $F(x)$ per una variabile casuale continua è detta funzione densità di probabilità. Per come è stata definita, la $p(x)$ non può essere negativa. Valgono le seguenti relazioni:

$$P(a < x \leq b) = \int_a^b p(x)dx$$

$$F(x) = \int_{-\infty}^x p(x)dx$$

$$\int_{-\infty}^{+\infty} p(x)dx = 1$$

La quantità $p(x)dx$ rappresenta la probabilità infinitesima che la variabile casuale X assuma un valore compreso tra x e $x+dx$.

7.3 Caratteristiche di una distribuzione di probabilità

7.3.1 Momenti di una variabile random

Si è visto come la funzione densità di probabilità di una variabile aleatoria consenta di ricavare la probabilità di qualsiasi evento definito sul fenomeno aleatorio a cui essa è associata. La densità di probabilità costituisce pertanto una descrizione completa per tale fenomeno aleatorio.

Non sempre però è necessaria una conoscenza così completa, ma è sufficiente la conoscenza di alcune grandezze globali inerenti la variabile aleatoria: questi sono i **momenti della variabile aleatoria** e in particolare:

- **i momenti di ordine n** della variabile aleatoria X definiti come:

$$m_x^{(n)} = \int_{-\infty}^{+\infty} x^n p_x(x) dx \quad (\text{per variabile continua})$$

$$m_x^{(n)} = \sum_j x_j^n P(x_j) \quad (\text{per variabile discreta})$$

- **i momenti centrati di ordine n** della variabile aleatoria X :

$$\mu_x^{(n)} = \int_{-\infty}^{+\infty} (x - m_x^{(1)})^n p_x(x) dx$$

che, ricordando lo sviluppo binomiale:

$$(a + b)^n = \sum_{k=0}^n \binom{n}{k} a^k b^{n-k}$$

possono essere ricavati da quelli non centrati:

$$\mu_x^{(n)} = \int_{-\infty}^{+\infty} \sum_k \binom{n}{k} x^k [-m_x^{(1)}]^{n-k} p_x(x) dx = \sum_k \binom{n}{k} [-m_x^{(1)}]^{n-k} m_x^{(k)}$$

Per $n=1$ si ottiene:

$$m_x^{(1)} = E[x]$$

e prende il nome di **valore atteso**.

Se si considera una variabile discreta:
$$m_x^{(1)} = E[x] = \sum_{j=1}^N x_j \frac{N_j}{N}$$

cioè il momento di ordine 1 di una variabile discreta è proprio il suo valor medio. Ovviamente $\mu_x^{(1)} = 0$.

Per $n=2$ si ottiene:

$$m_x^{(2)} = E[x^2] = \int_{-\infty}^{+\infty} x^2 p(x) dx$$

che prende il nome di **valore quadratico medio**. Si ha naturalmente:

$$\mu_x^{(2)} = E[(x - m_x^{(1)})^2] = E[(x - E[x])^2]$$

che prende il nome di **varianza** o si indica con σ^2 . Essa è il valore quadratico medio degli scarti delle **determinazioni** (valori assunti dalla variabile random) dal valore medio. La radice quadrata della varianza è detta **deviazione standard** σ .

Si noti che:

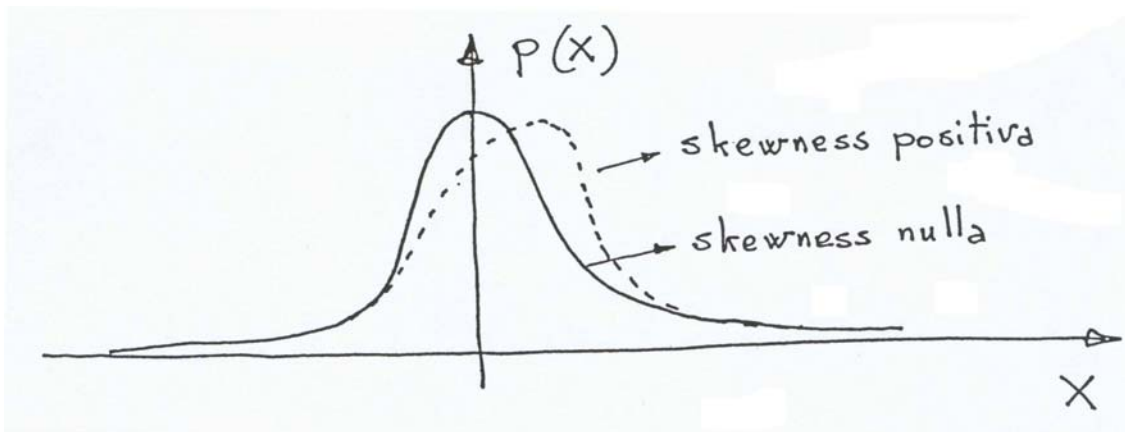
$$\sigma^2 = E[x^2 - 2xE[x] + (E[x])^2] = E[x^2] - 2E[x]E[x] + (E[x])^2 = E[x^2] - (E[x])^2$$

Pertanto la varianza coincide con il valore quadratico medio se e solo se $E[x]=0$.

Per i momenti centrati di ordine superiore si preferisce prendere in considerazione i momenti normalizzati rispetto alla potenza n -esima della deviazione standard, cioè:

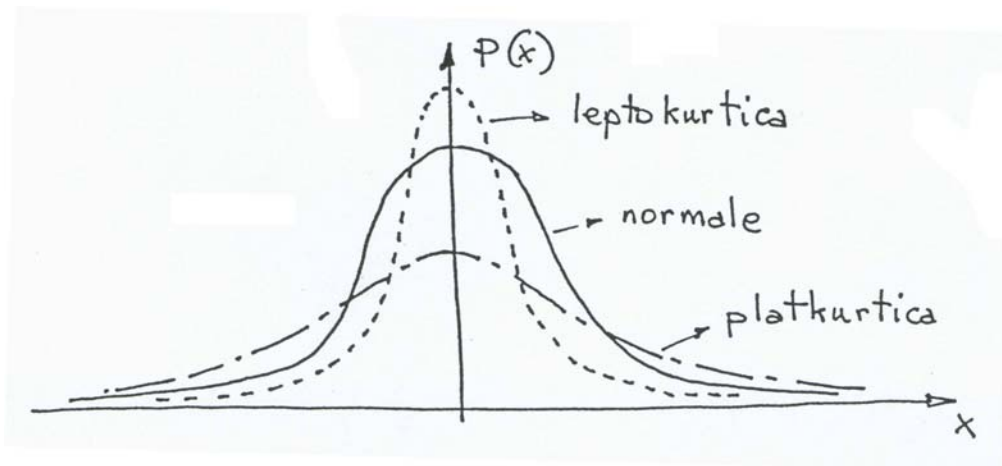
$$\mu_x^{(n)} = \frac{E[(x - E[x])^n]}{\sigma^n}$$

Il terzo momento ($n=3$) normalizzato prende il nome di **skewness** ed è una *misura della asimmetria della distribuzione* (vedi figura seguente):



Per una distribuzione gaussiana (che vedremo in seguito e rappresenta una distribuzione di riferimento) la skewness è nulla. Una skewness uguale a zero implica che i valori, specialmente quelli lontani dalla media, sono distribuiti in modo simmetrico da entrambi le parti della media.

Il quarto momento ($n=4$) normalizzato è noto come **kurtosis** o **flatness**, ed è una *misura della "piattezza" della distribuzione*. In altre parole, se la maggior parte dei valori si distribuisce intorno alla media, allora la curva di distribuzione ha un picco piuttosto alto intorno alla media e cade a zero molto rapidamente da entrambe le parti. In questo caso il kurtosis ha un basso valore e la curva di distribuzione si dice **leptokurtica**. La situazione opposta si ha se il kurtosis è alto e la distribuzione si dice **platokurtica** (vedi figura seguente). Per la distribuzione gaussiana il kurtosis vale 3. Il kurtosis è utile per molte applicazioni, ad esempio per caratterizzare segnali ottenuti in flussi turbolenti o segnalare l'iniziale apparizione di guasti (rotture) nei macchinari.



Poiché si è detto che il kurtosis della distribuzione gaussiana è pari a 3, molto spesso esso viene definito nel seguente modo:

$$k = \frac{E[x - E[x]]^4}{\sigma^2} - 3$$

in modo che una distribuzione leptokurtica ha kurtosis negativo, mentre una distribuzione platokurtica ha kurtosis positivo.

Esempio 3:

Si può facilmente dimostrare che un processo descritto da una funzione sinusoidale con fase aleatoria ha i seguenti quattro momenti:

$$m = E[x] = 0$$

$$\sigma^2 = E[x^2] = 0.5$$

$$sk = \frac{E[x^3]}{\sigma^3} = 0$$

$$k = \frac{E[x^4]}{\sigma^4} - 3 = -1.5$$

7.4 Variabili aleatorie pluridimensionali. Densità di probabilità di ordine superiore

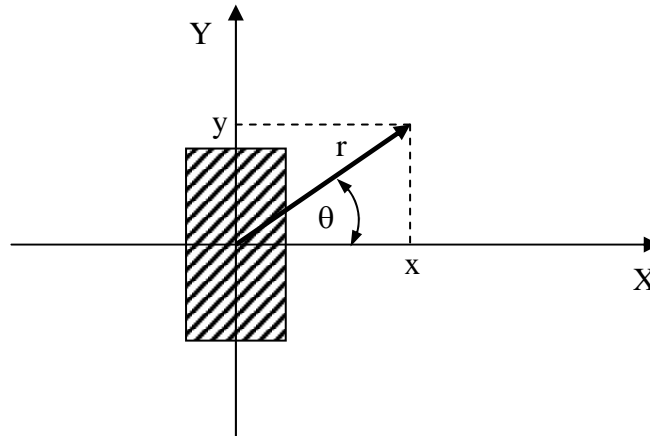
In taluni casi, l'impostazione illustrata nei precedenti paragrafi per il calcolo della probabilità di un evento non risulta soddisfacente: questo si verifica nei casi, esclusi a suo tempo, in cui i risultati dell'esperimento aleatorio non possono essere rappresentati (o non è conveniente che lo siano) mediante punti in un'unica dimensione, ma richiedono l'uso di due o più dimensioni, e quindi di due o più variabili aleatorie unidimensionali.

Limitiamo l'esame, nel presente paragrafo, ai casi in cui ogni risultato sia associabile ad un punto in uno spazio bidimensionale; pertanto ogni determinazione sarà in effetti costituita da una coppia di numeri reali $\omega = (x_1, x_2)$ e la corrispondente variabile aleatoria $W = (X_1, X_2)$ sarà anche essa

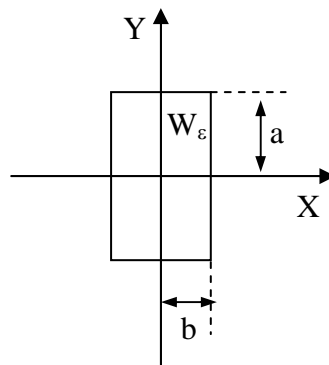
bidimensionale, costituita dall'insieme di due variabili aleatorie unidimensionali, che prendono il nome di *marginali* e che a loro volta descrivono due componenti marginali del fenomeno aleatorio. Questo caso si verifica in generale quando i risultati di un esperimento aleatorio sono per loro natura pluridimensionali e l'evento di cui si vuol considerare la probabilità è tale da implicare l'uso di due di tali dimensioni.

Per esempio, l'evento:

ε_0 : colpire il bersaglio entro il rettangolo tratteggiato della seguente figura.



fa riferimento ad un fenomeno aleatorio identico a quello a cui si riferisce l'evento dell'esempio 3 precedentemente visto. Tuttavia per la rappresentazione di ciascun risultato è necessario l'impiego di una coppia di variabili aleatorie, e cioè di una variabile aleatoria bidimensionale W , se si vuole rappresentare correttamente l'insieme di determinazioni (ovvero di punti nello spazio W) favorevoli all'evento. In particolare, se la coppia di variabili aleatorie è costituita dall'ascissa X e dall'ordinata Y del colpo [per cui $W = (X, Y)$] riferite al sistema coordinato della precedente figura, all'evento in oggetto sono favorevoli i risultati rappresentati dai punti contenuti nell'insieme W_ε della seguente figura.



Un caso molto importante in cui tale situazione si verifica è quando l'evento è definito dall'insieme di due frasi opportunamente connesse tra loro, di cui la prima costituisce un evento (che possiamo dire marginale) che vincola le determinazioni della sola variabile aleatoria marginale X , e la seconda costituisce un altro evento (anch'esso marginale) che vincola le determinazioni della sola variabile aleatoria marginale Y .

Sia dato dunque un fenomeno aleatorio i cui risultati siano rappresentabili da un punto in uno spazio bidimensionale e sia $W = (X, Y)$ la corrispondente variabile bidimensionale (in cui X e Y costituiscono le componenti unidimensionali marginali).

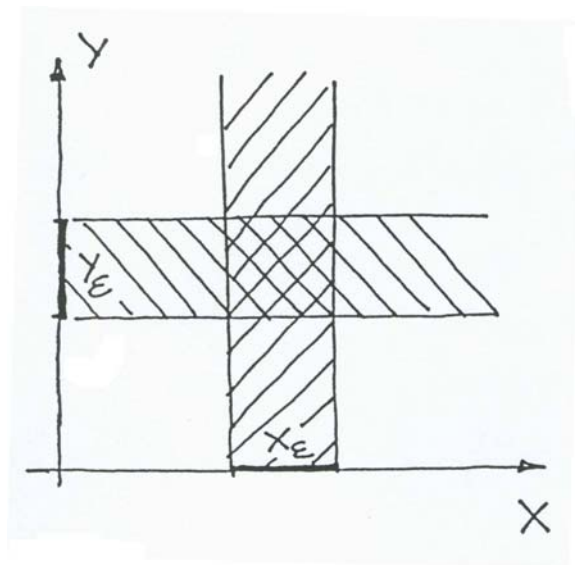
Consideriamo poi due generici eventi di tipo marginale ε_X e ε_Y a cui corrispondono gli insiemi di punti X_ε e Y_ε degli assi X e Y che comprendono tutte e sole le determinazioni favorevoli rispettivamente a ε_X e ε_Y .

Nello spazio bidimensionale, gli eventi ε_X e ε_Y possono esprimersi, per definizione, nel seguente modo:

ε_X : ε_X ed una qualunque determinazione per Y

ε_Y : ε_Y ed una qualunque determinazione per X .

Pertanto l'insieme W_X di punti favorevoli a ε_X nello spazio bidimensionale è costituito da tutti e soli quei punti che hanno ascissa contenuta in X_ε (figura seguente), mentre l'insieme W_Y di punti favorevoli a ε_Y è costituito da tutti e soli quei punti che hanno ordinata contenuta in Y_ε .



insieme dei punti del piano $W(X,Y)$ favorevoli all'evento:

$$\varepsilon_W = (\varepsilon_X) \text{ e } (\varepsilon_Y)$$

Esaminiamo adesso i tipi più comuni di connessione tra le frasi qualificanti i due eventi.

Un primo tipo di connessione è fornito dalla particella “e”; l'evento ε in questo caso è dato dalla frase:

$$\varepsilon : (\varepsilon_X) \text{ e } (\varepsilon_Y)$$

A tale evento sono favorevoli tutti e soli i risultati favorevoli ad entrambi gli eventi (ε_X) e (ε_Y) , e corrisponde all'insieme di punti comuni agli insiemi W_X e W_Y . Un evento di questo tipo prende il nome di **evento congiunto**.

Un secondo tipo di connessione è fornito dalla particella “o”; l'evento in questo caso è dato dalla frase:

$$\varepsilon : (\varepsilon_X) \text{ o } (\varepsilon_Y)$$

A tale evento sono favorevoli tutti e soli i risultati favorevoli ad almeno uno degli eventi (ε_X) e (ε_Y) , e corrisponde all'insieme di punti comuni agli insiemi W_X e W_Y . Un evento di questo tipo prende il nome di **evento totale**.

Il terzo e ultimo tipo di connessione è fornito dalla particella “quando”; l'evento in questo caso è dato dalla frase:

$$\varepsilon : (\varepsilon_X) \text{ quando } (\varepsilon_Y)$$

ovvero

$$\varepsilon : (\varepsilon_X) / (\varepsilon_Y)$$

A tale evento sono favorevoli tutti e soli i risultati favorevoli a ε_X purché il risultato sia favorevole anche a ε_Y ; se quest'ultima condizione non è soddisfatta, il risultato non è né favorevole né

sfavorevole a ε , ma l'esperimento aleatorio si deve considerare *non effettuato* (cioè non viene conteggiato nel novero delle prove effettuate). Un evento di questo tipo prende il nome di *evento condizionato* (corrispondentemente, ε_Y prende il nome di *condizione* o *evento condizionante*).

Detto ε_W un generico evento che si riferisce ad un fenomeno aleatorio la cui corrispondente variabile aleatoria risulti di tipo bidimensionale (non necessariamente composta dalla connessione tra eventi marginali), la sua probabilità è ovviamente fornita ancora dalla

$$\Pr\{\varepsilon_x\} = \lim_{N \rightarrow \infty} \frac{N_\varepsilon}{N} \geq 0$$

per cui valgono le osservazioni fatte a suo tempo a proposito di tale espressione, e in particolare alla tendenza ad un valore infinito del rapporto N_ε/N , quando N tende all'infinito. Quello di cui ci si vuole ora occupare è il modo con cui calcolare tale probabilità in termini di una funzione di densità di probabilità opportunamente definita, in analogia a quanto fatto nel paragrafo 1 di questo capitolo. Allo scopo, e senza ripeterci ulteriormente, si può dimostrare che, se la variabile aleatoria bidimensionale gode della proprietà di "regolarità" e detto W_ε l'insieme di punti nello spazio $[X, Y]$ corrispondenti all'evento ε_W , supposto misurabile, la probabilità di ε_W è data dalla seguente espressione:

$$\Pr\{\varepsilon\} = \int_{W_\varepsilon} p_{XY}(x, y) dx dy$$

ove $p_{XY}(x, y)$ è una funzione delle variabili X e Y detta *densità di probabilità* della variabile aleatoria bidimensionale $W=(X, Y)$ ovvero *densità di probabilità congiunta* delle variabili aleatorie unidimensionali X e Y .

Vediamo ora di applicare la formula generale appena vista ad alcuni eventi di tipo particolare introdotti in precedenza.

7.4.1 Probabilità di evento marginale

Sia ε_X un evento marginale e X_ε il corrispondente insieme di determinazioni della variabile aleatoria marginale unidimensionali X ad esso favorevoli. Nello spazio bidimensionale, all'evento ε_X corrisponde l'insieme di punti che hanno ordinata qualunque e ascissa appartenente a X_ε (vedi figura precedente); pertanto la probabilità ε_X risulta pari a:

$$\Pr\{\varepsilon_X\} = \int_{y=-\infty}^{\infty} \int_{X_\varepsilon} p_{XY}(x, y) dx dy = \int_{X_\varepsilon} \left\{ \int_{y=-\infty}^{\infty} p_{XY}(x, y) dy \right\} dx$$

Ora X è una variabile aleatoria unidimensionali e pertanto ammette una densità di probabilità $p_X(x)$; poiché ε_X è definito solo su tale variabile aleatoria, la probabilità di ε_X è data, in termini di $p_X(x)$, dall'espressione $\Pr(\varepsilon_X) \int_{X_\varepsilon} p_X(x) dx$; confrontando quest'ultima con la precedente, e tenendo conto del fatto che X_ε è generico, si ricava l'uguaglianza:

$$p_X(x) = \int_{y=-\infty}^{\infty} p_{XY}(x, y) dy$$

che fornisce la densità di probabilità della variabile aleatoria marginale, nota la densità di probabilità congiunta. Con analogo ragionamento, si ottiene la seguente espressione per la densità di probabilità della variabile aleatoria marginale Y :

$$p_Y(y) = \int_{x=-\infty}^{\infty} p_{XY}(x, y) dx$$

7.4.2 Probabilità di un evento congiunto

Siano ε_X e ε_Y due eventi marginali relativi alle variabili aleatorie marginali X e Y , siano X_ε ed Y_ε gli insiemi delle determinazioni delle variabili aleatorie X e Y favorevoli a tali eventi e sia infine ε_W l'evento congiunto.

Allora, la probabilità di ε_X e ε_Y è data dalla relazione

$$\Pr\{\varepsilon_W = (\varepsilon_X)e(\varepsilon_Y)\} = \int_{X_\varepsilon} \int_{Y_\varepsilon} p_{XY}(x, y) dx dy$$

Ora, poiché $p_{XY}(x, y)$ è sempre ≥ 0 , e poiché l'insieme dei punti favorevoli a $(\varepsilon_X)e(\varepsilon_Y)$ è compreso in quello favorevole a ε_Y , risultano vere le seguenti disuguaglianze:

$$\Pr\{(\varepsilon_X)e(\varepsilon_Y)\} \leq \Pr\{\varepsilon_X\}$$

$$\Pr\{(\varepsilon_X)e(\varepsilon_Y)\} \leq \Pr\{\varepsilon_Y\}$$

da cui:

$$[\Pr\{(\varepsilon_X)e(\varepsilon_Y)\}]^2 \leq \Pr\{\varepsilon_X\} \cdot \Pr\{\varepsilon_Y\}$$

che costituisce l'unico legame esistente a priori tra le probabilità degli eventi ε_X , ε_Y , $(\varepsilon_X)e(\varepsilon_Y)$.
Se risulta

$$\Pr\{(\varepsilon_X)e(\varepsilon_Y)\} = \Pr\{\varepsilon_X\} \cdot \Pr\{\varepsilon_Y\}$$

gli eventi ε_X e ε_Y si dicono **indipendenti**.

Affinché l'equazione precedente sia verificata, è necessario e sufficiente che la densità di probabilità soddisfi alla seguente relazione:

$$\int_{X_\varepsilon} \int_{Y_\varepsilon} [p_{XY}(x, y) - p_X(x) \cdot p_Y(y)] dx dy = 0$$

Se infine la densità di probabilità congiunta è tale che la $\Pr\{(\varepsilon_X)e(\varepsilon_Y)\} = \Pr\{\varepsilon_X\} \cdot \Pr\{\varepsilon_Y\}$ sia verificata qualunque siano gli eventi marginali ε_X e ε_Y ovvero, qualunque siano gli insiemi X_ε e Y_ε , le variabili aleatorie marginali X e Y si dicono **st statisticamente indipendenti**; condizione necessaria e sufficiente affinché ciò si verifichi è che la precedente equazione valga qualunque sia X_ε e Y_ε , quindi che risulti verificata la relazione:

$$p_{XY}(x, y) = p_X(x) \cdot p_Y(y)$$

(fattorizzabilità della densità di probabilità congiunta).

Tale condizione esprime analiticamente il fatto che le variabili aleatorie marginali del fenomeno aleatorio non si influenzano reciprocamente e cioè non esiste un legame di natura qualsiasi che renda più o meno probabile ciascuna determinazione per una variabile aleatoria in funzione della determinazione ottenuta per l'altra.

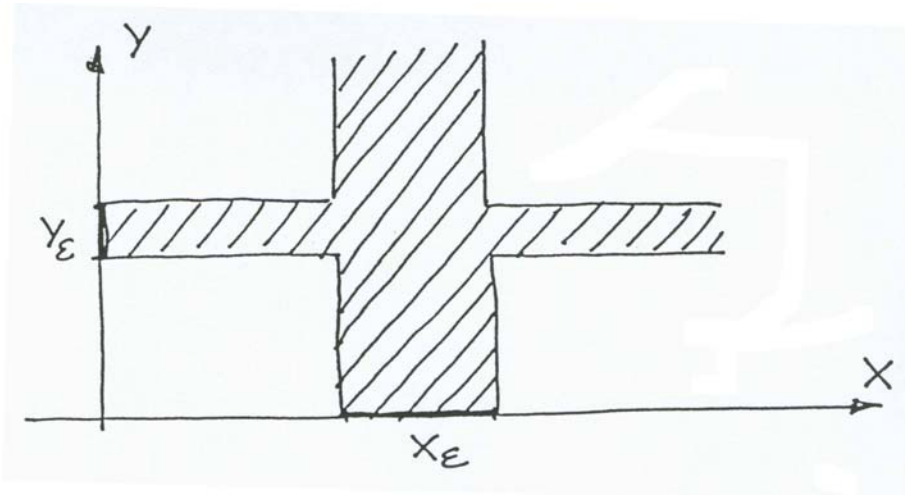
7.4.3 Probabilità di un evento totale

Siano ε_X e ε_Y due eventi marginali relativi alle variabili aleatorie marginali X e Y , siano X_ε e Y_ε gli insiemi delle determinazioni delle variabili aleatorie X e Y favorevoli a tali eventi e sia infine ε_W : $(\varepsilon_X) \cup (\varepsilon_Y)$ l'evento totale.

Allora la probabilità di ε_W è data dalla seguente espressione:

$$\Pr\{(\varepsilon_X) \cup (\varepsilon_Y)\} = \int_{\lambda_\varepsilon} p_{XY}(x, y) dx dy$$

ove λ_ε è l'insieme tratteggiato della seguente figura:



Osservando i domini di integrazione per gli eventi ε_X , ε_Y e ε_W , si ricavano facilmente le seguenti disuguaglianze:

$$\Pr\{(\varepsilon_X) \cup (\varepsilon_Y)\} \geq \Pr\{\varepsilon_X\}$$

$$\Pr\{(\varepsilon_X) \cup (\varepsilon_Y)\} \geq \Pr\{\varepsilon_Y\}$$

e infine:

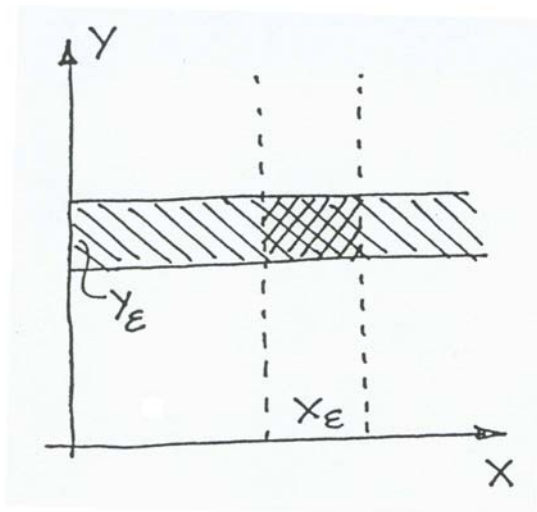
$$\Pr\{(\varepsilon_X) \cup (\varepsilon_Y)\} = \Pr\{\varepsilon_X\} + \Pr\{\varepsilon_Y\} - \Pr\{(\varepsilon_X) \cap (\varepsilon_Y)\}$$

Questa relazione è nota come il **teorema delle probabilità totali**, e mostra come la probabilità di un evento totale sia minore o uguale alla somma delle probabilità dei rispettivi eventi marginali; in particolare vale l'eguaglianza se e solo se $\Pr\{(\varepsilon_X) \cap (\varepsilon_Y)\} = 0$ cioè se e solo se la probabilità che i due eventi si verifichino contemporaneamente è nulla. In tal caso i due eventi si dicono **mutuamente escludentisi**, per ovvi motivi.

7.4.4 Probabilità di un evento condizionato

Siano ε_X e ε_Y due eventi marginali relativi alle variabili aleatorie marginali X e Y , siano X_ε e Y_ε gli insiemi delle determinazioni delle variabili aleatorie X e Y favorevoli a tali eventi e sia infine ε_W : $\varepsilon_X/\varepsilon_Y$ l'evento condizionato.

A tale evento condizionato sono favorevoli gli stessi punti del piano (X,Y) favorevoli all'evento congiunto ε_X e ε_Y (vedi figura seguente),



ma la sua probabilità non è pari a $\Pr(\varepsilon_X \text{ e } \varepsilon_Y)$ in quanto in N ripetizioni del fenomeno aleatorio non tutti i risultati vanno considerati utili, ma solo quelli favorevoli a ε_Y , e cioè compresi nella fascia tratteggiata della figura precedente; pertanto, per definizione, detti N_{XY} , N_X ed N_Y il numero di risultati in N prove, favorevoli rispettivamente a (ε_X ed ε_Y), ε_X ed ε_Y , si ha:

$$\Pr\{\varepsilon_X / \varepsilon_Y\} = \lim_{N \rightarrow \infty} \frac{N_{XY}}{N_Y} = \lim_{N \rightarrow \infty} \left[\frac{N_{XY}}{N} / \frac{N_Y}{N} \right]$$

e quindi:

$$\Pr\{\varepsilon_X / \varepsilon_Y\} = \frac{\Pr\{\varepsilon_X \text{ e } \varepsilon_Y\}}{\Pr\{\varepsilon_Y\}}$$

Questa formula mostra che la probabilità di un evento ε_X condizionato ad un evento ε_Y è data dal rapporto della probabilità dell'evento congiunto divisa la probabilità dell'evento condizionante ε_Y (**teorema delle probabilità composte**).

Poiché il secondo membro dell'equazione precedente può anche scriversi nel seguente modo:

$$\frac{\int_{X_\varepsilon} \int_{Y_\varepsilon} p_{XY}(x, y) dx dy}{\Pr\{\varepsilon_Y\}} = \int_{X_\varepsilon} \left[\frac{\int_{Y_\varepsilon} p_{XY}(x, y) dy}{\Pr\{\varepsilon_Y\}} \right] dx$$

risulta vera la seguente espressione alternativa per il teorema delle probabilità composte:

$$\Pr\{\varepsilon_X / \varepsilon_Y\} = \int_{X_\varepsilon} \left[\frac{\int_{Y_\varepsilon} p_{XY}(x, y) dy}{\Pr\{\varepsilon_Y\}} \right] dx$$

Confrontando questa relazione con la $\Pr\{\varepsilon_x\} = \lim_{N \rightarrow \infty} \frac{N_\varepsilon}{N} \geq 0$, si vede che la probabilità di un

generico evento marginale ε_X condizionato ε_Y , può calcolarsi mediante la medesima espressione che si adopererebbe nel caso di evento non condizionato, purché si impieghi come integrando non la

densità di probabilità $p_X(x)$ ma una diversa funzione, che prende il nome di **densità di probabilità di X condizionata a ε_Y** , indicata con $p_X/\varepsilon_Y(x/\varepsilon_Y)$ e definita nel seguente modo:

$$p_X/\varepsilon_Y(x/\varepsilon_Y) = \frac{1}{\Pr\{\varepsilon_Y\}} \int_{y_\varepsilon} p_{XY}(x, y) dy = \frac{\int_{y_\varepsilon} p_{XY}(x, y) dy}{\int_{y_\varepsilon} p_{XY}(y) dy}$$

Questa funzione gode di tutte le proprietà di una densità di probabilità; in particolare risulta maggiore o uguale a zero, e inoltre il suo integrale da $-\infty$ a $+\infty$ vale 1. Infatti:

$$\int_{-\infty}^{+\infty} p_X/\varepsilon_Y(x/\varepsilon_Y) dx = \frac{1}{\Pr\{\varepsilon_Y\}} \int_{-\infty}^{+\infty} p_{XY}(x, y) dy = \frac{\Pr\{\varepsilon_Y\}}{\Pr\{\varepsilon_Y\}} = 1$$

Essa, come è ben evidente, costituisce la densità di probabilità della variabile aleatoria condizionata $\{X/\varepsilon_Y\}$.

Infine si osservi che, mentre nel caso di eventi congiunti o totali vi è asimmetria tra gli eventi marginali, non altrettanto è vero per eventi condizionati.

In particolare, gli eventi:

$$\varepsilon : \varepsilon_X/\varepsilon_Y$$

$$\varepsilon : \varepsilon_Y/\varepsilon_X$$

sono diversi e quindi hanno diversa probabilità. E' di interesse la determinazione del legame sussistente tra tali probabilità. Applicando il teorema delle probabilità condizionate, si ottiene:

$$\Pr\{\varepsilon_Y/\varepsilon_X\} = \frac{\Pr\{\varepsilon_X \text{ e } \varepsilon_Y\}}{\Pr\{\varepsilon_X\}} = \frac{\Pr\{\varepsilon_Y\}}{\Pr\{\varepsilon_X\}} \Pr\{\varepsilon_X/\varepsilon_Y\}$$

Questa relazione è nota con il nome di **Teorema di Bayes**.

In effetti la formulazione di tale teorema è leggermente diversa e più generale, ma per semplicità non è stata riportata qui in tale forma.

Un ultimo risultato di notevole interesse si ha nella determinazione delle probabilità condizionata nel caso di eventi ε_X e ε_Y indipendenti o mutuamente escludentisi; nel primo caso risulta:

$$\Pr\{\varepsilon_Y/\varepsilon_X\} = \frac{\Pr\{\varepsilon_X \text{ e } \varepsilon_Y\}}{\Pr\{\varepsilon_X\}} = \Pr\{\varepsilon_Y\}$$

$$\Pr\{\varepsilon_Y/\varepsilon_X\} = \frac{\Pr\{\varepsilon_X \text{ e } \varepsilon_Y\}}{\Pr\{\varepsilon_Y\}} = \Pr\{\varepsilon_X\}$$

ovvero, la probabilità che si verifichi l'evento ε_Y quando si è verificato ε_X è uguale alla probabilità di ε_Y (e cioè non dipende dall'essersi verificato o meno l'evento ε_X), e viceversa, il che giustifica pienamente il nome dato a tali eventi.

Nel caso di eventi mutuamente escludentisi si ha $\Pr\{\varepsilon_X \text{ e } \varepsilon_Y\} = 0$, e quindi risulta:

$$\Pr\{\varepsilon_Y/\varepsilon_X\} = \Pr\{\varepsilon_X/\varepsilon_Y\} = 0$$

Questo risultato ha una giustificazione intuitiva immediata, in quanto se ε_X e ε_Y sono mutuamente escludentesi, ε_X (ovvero ε_Y) non si verificherà mai, una volta verificatosi ε_Y (ovvero ε_X).

7.5 Momenti di variabili aleatorie pluridimensionali

Nel precedente paragrafo si è introdotto il concetto di variabile aleatoria pluridimensionale a cui si è associata la densità di probabilità, mediante la quale è possibile valutare la probabilità di un qualsiasi evento a cui sia favorevole un insieme di punti dello spazio n -dimensionale.

La densità di probabilità descrive pertanto completamente il fenomeno aleatorio.

Analogamente al caso delle variabili monodimensionali è possibile, anche in questo caso, definire delle grandezze globali che talvolta sono sufficienti a descrivere adeguatamente il fenomeno: i momenti delle variabili aleatorie pluridimensionali.

Tra questi definiamo:

- I **momenti di ordine** (k_1, k_2, \dots, k_n) della variabile pluridimensionale $\omega(x_1, x_2, \dots, x_n)$, definiti come:

$$m_{\omega}^{(k_1, k_2, \dots, k_n)} = \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} x_1^{k_1} x_2^{k_2} \dots x_n^{k_n} p_{\omega}(x_1, \dots, x_n) dx_1 \dots dx_n$$

e, in particolare, per una variabile bidimensionale ($n=2$):

$$m_{x_1, x_2}^{(k_1, k_2)} = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} x_1^{k_1} x_2^{k_2} p_{x_1, x_2} dx_1 dx_2$$

- I **momenti misti centrati di ordine** (k_1, k_2, \dots, k_n) che per $n=2$ diventa:

$$\mu_{x_1, x_2}^{(k_1, k_2)} = \int \int_{-\infty}^{+\infty} [x_1 - E[x_1]]^{k_1} [x_2 - E[x_2]]^{k_2} p_{x_1, x_2} dx_1 dx_2$$

Qui particolare interesse acquista il momento misto centrato di ordine (1,1) per la variabile bidimensionale che si scrive:

$$\mu_{x_1, x_2}^{(1,1)} = \int \int_{-\infty}^{+\infty} [x_1 - E[x_1]] [x_2 - E[x_2]] p_{x_1, x_2} dx_1 dx_2 = m_{x_1, x_2}^{(1,1)} - E[x_1]E[x_2]$$

e che prende il nome di **covarianza**.

Si noti che

$$\mu_{x_1, x_2}^{(1,1)} = m_{x_1, x_2}^{(1,1)}$$

se e solo se almeno uno dei valori attesi di x_1 o x_2 è nullo.

Naturalmente la covarianza è nulla se x_1 e x_2 sono statisticamente indipendenti, mentre non è in genere vero il contrario.

7.6 Alcune importanti distribuzioni di probabilità

7.6.1 La distribuzione binomiale o di Bernoulli

La distribuzione binomiale o di Bernoulli governa eventi che avvengono a caso in modo discreto. Essa si applica quando:

- In ogni prova di una certa serie si possono presentare *solo* due possibili risultati (ad es. nel lancio di una moneta vi è la possibilità che si presenti testa o croce; nel lancio di un dado vi è la possibilità che si presenti un determinato numero oppure no, etc...).
- Le prove successive sono indipendenti.
- La probabilità che si presenti un determinato risultato è costante per tutte le prove.
- Il risultato per tutte le prove è casuale.

Quando queste condizioni sono soddisfatte può presentarsi il problema: qual è la probabilità $P(n, k)$ che un dato risultato si presenti k volte in n prove?

La probabilità che in n prove si presenti una certa sequenza (con un determinato ordine di risultati) in cui un dato risultato che ha probabilità p di avvenire compare k volte e quello opposto (probabilità $q=1-p$) $n-k$ volte è, per la legge delle probabilità composte (eventi indipendenti):

$$p^k q^{n-k}$$

A noi tuttavia non interessa il particolare ordine con cui i due diversi eventi si susseguono.

Così considerando il lancio di una moneta sono possibili i due eventi: croce (A) o testa (B).

Se ora cerchiamo le probabilità che ad esempio in tre lanci si verifichi due volte l'evento A, ed una volta l'evento B, le possibili sequenze che danno luogo a questa situazione sono:

$$A,A,B \quad A,B,A \quad B,A,A$$

e ciascuna di esse ha una eguale probabilità di manifestarsi. Pertanto la probabilità dell'evento considerato si ottiene moltiplicando la $p^k q^{n-k}$ per il numero delle possibili sequenze (nell'esempio

considerato tre). Tale numero è dato dalle combinazioni di n oggetti presi a k a k , cioè: $\binom{n}{k}$.

Si ottiene quindi

$$P(n, k) = \binom{n}{k} p^k q^{n-k} = \frac{n!}{k!(n-k)!} p^k q^{n-k}$$

che è la formula trovata da Bernoulli e prende il nome di **legge di distribuzione binomiale** o di **Bernoulli**. Essa rappresenta la probabilità che un evento che ha probabilità p di accedere e probabilità $q=1-p$ di non accedere si presenti k volte in un gruppo di n prove.

Essa si applica evidentemente solo se k e n sono dei numeri interi.

Si vede che questa probabilità rappresenta il termine generico dello sviluppo del binomio

$$(p+q)^n = p^n + \binom{n}{1} p^{n-1} q + \dots + \binom{n}{k} p^k q^{n-k} + \dots + q^n$$

D'altra parte poiché $p + q = 1$ la somma di tutte le possibili probabilità con cui si possono presentare gli eventi è 1 come vuole la definizione stessa di probabilità..

Esempio 4: Come esempio consideriamo tre lanci di un dado. Per mezzo della distribuzione binomiale si può calcolare la probabilità che un dato numero (per es. il 2) si presenti nei tre lanci tre volte, due volte, una volta o nessuna volta. Infatti la probabilità che esca il numero è $p = 1/6$ mentre la probabilità che non esca (cioè che esca un qualsiasi altro numero) è $q = 1 - p = 5/6$.

Quindi la probabilità che in tre lanci si presenti una volta il 2 è:

$$P(3,1) = \frac{3!}{1!(3-1)!} p q^{3-1} = \frac{3 \cdot 2}{2} \frac{1}{6} \left(\frac{5}{6}\right)^2 = \frac{25}{72} = \frac{75}{216}$$

La probabilità che in tre lanci si presenti due volte il numero 2 è:

$$P(3,2) = \frac{3!}{2!(3-2)!} p^2 q^{3-2} = \frac{3 \cdot 2}{2} \left(\frac{1}{6}\right)^2 \frac{5}{6} = \frac{5}{72} = \frac{15}{216}$$

La probabilità che in tre lanci si presenti tre volte il numero 2 è:

$$P(3,3) = \frac{3!}{3!(3-3)!} p^3 q^0 = \left(\frac{1}{6}\right)^3 = \frac{1}{216}$$

La probabilità che non si presenti mai è:

$$P(3,0) = q^3 = \left(\frac{5}{6}\right)^3 = \frac{125}{216}$$

Si noti come la somma delle probabilità sia

$$\sum_{k=0}^3 P(3,k) = \frac{75 + 15 + 1 + 125}{216} = \frac{216}{216} = 1$$

mentre la probabilità che il numero non si presenti affatto è

$$\frac{125}{216} = 58\%$$

Il valore medio del risultato di n prove è:

$$m^{(1)} = \sum_{k=0}^n k P(n,k) = np$$

mentre i momenti del 2° ordine valgono:

$$m^{(2)} = \sum_{k=0}^n k^2 P(n,k) = np(1-p) + n^2 p^2$$

$$\mu^{(2)} = \sum_{k=0}^n [k - np]^2 P(n, k) = np(1 - p)$$

La distribuzione è asimmetrica eccetto per il caso in cui $p=1/2$. Si vede che, quando $p \neq 1/2$, l'asimmetria diminuisce al crescere di k .

La distribuzione binomiale è molto importante. In un certo senso essa può essere considerata la distribuzione fondamentale. Infatti le distribuzioni di Gauss e Poisson possono essere considerate casi speciali di questa. La distribuzione di Poisson si ottiene per n tendente ad infinito e p tendente a zero in modo tale che np resti finito (nella fattispecie $np \ll \sqrt{n}$).

La distribuzione di Gauss può ottenersi passando al limite per n tendente ad infinito e p non troppo piccolo (in pratica questa distribuzione è sufficientemente buona quando $np > 5$).

7.6.2 Distribuzione di Poisson

La distribuzione di Poisson serve a descrivere tutti i processi che avvengono a caso, la cui probabilità p di verificarsi è piccola ma costante. Si pensi ad esempio ad eventi casuali come la vincita ad una lotteria, disintegrazioni nucleari, etc. La distribuzione di Poisson si può dedurre come caso limite dalla distribuzione binomiale, per n tendente all'infinito e p tendente a zero.

Si considerino gli eventi numerabili che si verificano in modo random nello spazio e nel tempo (terremoti in una certa regione, numero di veicoli che attraversano un ponte, ecc.).

Sia $N(t)$ il numero di eventi random nell'intervallo $(0, t)$. Risulta $N(0)=0$ con probabilità unitaria.

Il processo random numerabile è detto di Poisson se:

$$1) \quad P[N(t_1) - N(s_1) = n / N(t_2) - N(s_2) = m] = P[N(t_1) - N(s_1) = n]$$

se gli intervalli (s_1, t_1) (s_2, t_2) sono indipendenti, cioè i futuri arrivi non sono affetti dagli arrivi passati.

2) Il processo ha incrementi stazionari, cioè:

$$P[N(t) - N(s) = n] = P[N(t + a) - N(s + a) = n]$$

per ogni a : cioè la probabilità dell'evento è funzione di $t-s$, ma non da t e s individualmente.

3) La probabilità di più di un arrivo in un intervallo infinitesimo $(t, t + \Delta t)$ è trascurabile, cioè:

$$P[N(t + \Delta t) - N(t) = 1] = v \cdot \Delta t + o(\Delta t)$$

$$P[N(t + \Delta t) - N(t) \geq 2] = o(\Delta t)$$

con v velocità media degli arrivi.

Sotto queste condizioni si ha:

$$P[N(t) - N(s) = n] = \frac{v(t-s)^n}{n!} e^{-v(t-s)}$$

Si ponga $s = 0$: $N(0) = 0$, e sia

$$P(n, t) = P[N(t) - N(0) = n]$$

Si ha:

$$P[n,t] = \frac{(vt)^n e^{-vt}}{n!} \quad \text{distribuzione di Poisson}$$

Si noti che nella distribuzione di Poisson il valore medio non è quello cui compete la massima probabilità e che gli scarti positivi e negativi intorno alla media hanno la stessa probabilità.

Il valor medio è vt .

7.6.3 Distribuzione normale (o di Gauss)

La terza distribuzione di grandissimo interesse è quella di Gauss detta anche distribuzione normale. La distribuzione di Gauss si usa quando il numero delle prove che si fanno è molto grande e la grandezza osservabile non assume valori discreti come nella distribuzione binomiale ma può assumere ogni valore compreso tra $-\infty$ e $+\infty$.

Consideriamo in dettaglio la distribuzione di Gauss in quanto essa ci permette in particolare di ricavare come si distribuiscono i valori degli scarti di una data misura intorno al valor medio.

Si può far vedere che uno dei modi per ricavare la legge di Gauss è quello di far partire dalla distribuzione binomiale imponendo che:

1. il numero n di prove sia molto grande;
2. la probabilità p che avvenga l'evento considerato sia sufficientemente grande perché $np \gg 1$ (al limite infinito).

In pratica la distribuzione normale è valida con buona approssimazione fintanto che è $n > 10$ e $np > 5$. La funzione densità di probabilità di questa distribuzione è:

$$f(x) = p(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}} \quad -\infty < x < +\infty$$

nella quale m e σ sono rispettivamente la media e la deviazione standard.

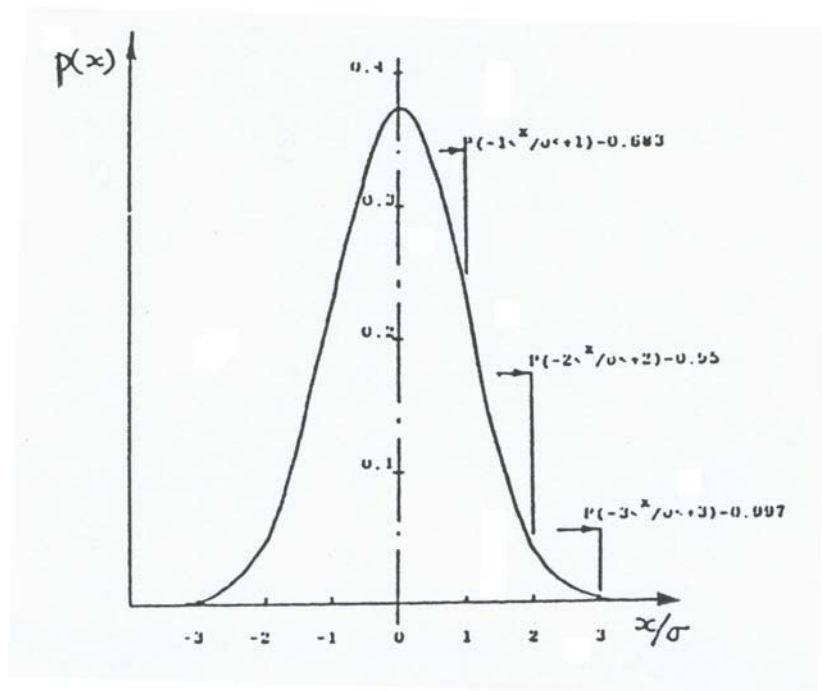
La funzione di ripartizione corrispondente è:

$$F(x) = P(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{(v-m)^2}{2\sigma^2}} dv$$

Il grafico della funzione normalizzata:

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

che si ottiene per $\sigma=1$ ed $m=0$ e che prende il nome di **distribuzione normale standardizzata** è mostrato nella seguente figura:



Nel grafico sono state indicate le aree con $|X| = |x/\sigma|$ inferiore a 1, 2 e 3 (cioè tra $x/\sigma = -1$ e $+1$, tra $x/\sigma = -2$ e $+2$, tra $x/\sigma = -3$ e $+3$) pari rispettivamente al 68.27%, 95,45% e 99.73% dell'area totale che è 1. Questo significa che:

$$P(-1 \leq X \leq 1) = 0.6827$$

$$P(-2 \leq X \leq 2) = 0.9545$$

$$P(-3 \leq X \leq 3) = 0.9973$$

L'importanza particolare che di solito viene attribuita alla distribuzione Gaussiana risiede nelle seguenti proprietà:

1. Le variabili random generate da operazioni lineari su variabili random normali sono anch'esse normali.
2. Teorema del limite centrale: se X_1, X_2, \dots, X_n sono n variabili random mutuamente indipendenti con media e varianza finita, la somma:

$$S_n = \sum_{i=1}^n X_i$$

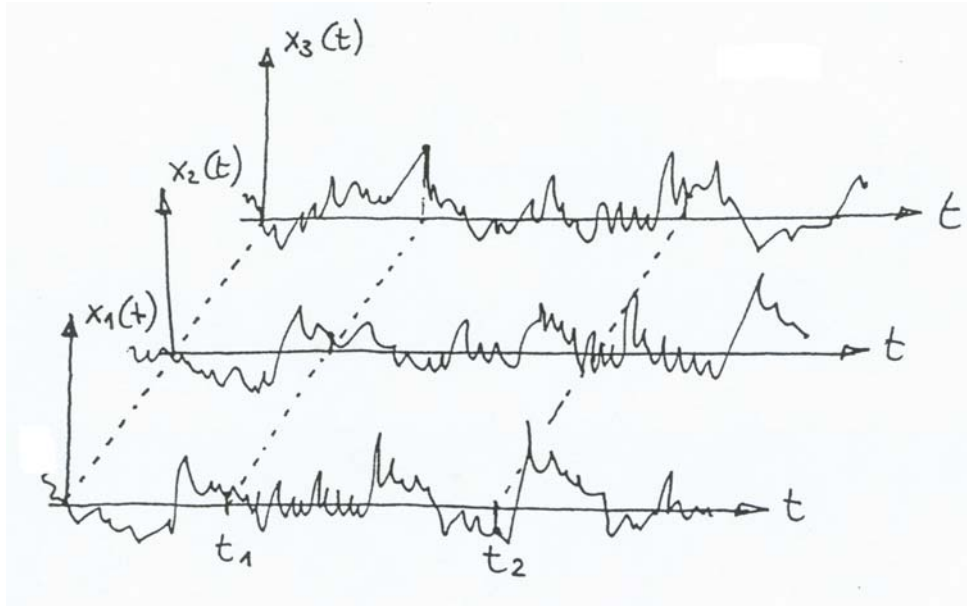
tende ad essere una variabile random normale per $n \rightarrow \infty$, purché nessuna delle variabili X_i contribuisca significativamente alla somma.

Il teorema fornisce una giustificazione matematica all'asserzione che una larga classe di fenomeni random è normale.

7.7 Caratteristiche di un processo stocastico

7.7.1 Processo stazionario. Medie di insieme

Si consideri l'accelerazione del centro di gravità di un passeggero seduto su una autovettura. Se si fanno continue registrazioni di tale accelerazione per molte vetture che percorrono la stessa strada, si ottengono una serie di registrazioni come nella seguente figura:



L'accelerazione rappresentata da questa serie (teoricamente infinita) di registrazioni è un processo stocastico (*random*) $\{x(t)\}$.

Ogni funzione di accelerazione $x_k(t)$ – **funzione campione del processo** – è funzione del tempo e della particolare autovettura considerata, e può essere vista come risultato di una prova separata.

Per trattare adeguatamente il problema, il processo stocastico $\{x(t)\}$ può essere ricondotto a un insieme di variabili random a istanti fissi di tempo t_1, t_2, \dots, t_n .

La struttura probabilistica del processo è costituita dalla gerarchia delle densità di probabilità congiunte:

$$p(x_1, t_1)$$

$$p(x_1, t_1, x_2, t_2)$$

.....

$$p(x_1, t_1, x_2, t_2, \dots, x_n, t_n)$$

con $x_i = x(t_i)$, tale che:

$p(x_1, t_1) \cdot dx_1$ è la probabilità che l'ampiezza di $x(t)$ sia compresa, nell'istante t_1 , tra x_1 e $x_1 + dx_1$

$p(x_1, t_1, x_2, t_2) \cdot dx_1 dx_2$ è la probabilità congiunta che l'ampiezza di $x(t)$ sia compresa tra x_1 e $x_1 + dx_1$ nell'istante t_1 , e tra x_2 e $x_2 + dx_2$, nell'istante t_2 .

eccetera...

Ciò stabilito, il processo si dice stazionario se la sua struttura di probabilità è invariante sotto arbitrarie traslazioni temporali. In altre parole $\{x(t)\}$ è stazionario fino all'ordine n se, per tutti i t_1, t_2, \dots, t_n e per una costante a arbitraria si ha:

$$p(x_1, t_1, x_2, t_2, \dots, x_n, t_n) = p(x_1, t_1 + a, x_2, t_2 + a, \dots, x_n, t_n + a)$$

Si ponga $a = -t_1$. Si ottiene:

$$p(x_1, t_1) = p(x_1, 0)$$

$$p(x_1, t_1, x_2, t_2) = p(x_1, 0, x_2, t_2 - t_1)$$

ecc.

Se dunque il processo è stazionario la probabilità del primo ordine non dipende dal tempo, la probabilità del secondo ordine dipende dall'intervallo temporale $t_2 - t_1$, ma non, separatamente, dai singoli valori di t_1 e t_2 . Ragionamenti conseguenti possono essere sviluppati per la probabilità di ordine superiore.

Ricordando che:

$$\mu(t) = E[x(t)] = \int_{-\infty}^{+\infty} x p(x, t) dx$$

$$\sigma^2(t) = E[(x(t) - \mu(t))^2] = \int_{-\infty}^{+\infty} [x - \mu]^2 p(x, t) dx$$

$$\mu_{x_1, x_2}(t_1, t_2) = E[x(t_1), x(t_2)] = \int \int_{-\infty}^{+\infty} [x_1 - \mu][x_2 - \mu] p_{x_1, x_2}(x_1, t_1, x_2, t_2) dx_1 dx_2$$

e, ricordando quanto ora detto per un processo stazionario, si ha:

$$\mu(t) = \mu = \text{cost}$$

$$\sigma(t) = \sigma = \text{cost}$$

$$\mu_{x_1, x_2}(t_1, t_2) = \mu_{x_1, x_2}(t_2 - t_1) = \mu_{x_1, x_2}(\tau)$$

In altre parole la media e la deviazione standard di un processo stazionario sono indipendenti dal tempo, mentre la covarianza e le funzioni che da essa derivano (vedremo più avanti la autocorrelazione e la cross-correlazione) dipendono dall'intervallo temporale τ ma non dai valori assoluti che generano tale intervallo.

Ritornando ora al nostro modello di processo stocastico, si possono statisticamente considerare due diversi tipi di medie. In primo luogo si può considerare l'accelerazione al tempo t_1 per tutte le automobili e scrivere la media di insieme che è la media al tempo t_1 di tutte le automobili. Una simile operazione può essere fatta per gli istanti t_2, t_3, t_n ; per la definizione di processo stazionario tutte queste medie di insieme sono uguali tra di loro.

Molti processi naturali sono sufficientemente stazionari (si noti che il termine stazionario è riferito alla distribuzione di probabilità e non alle funzioni campione). Per esempio, se uno considera la componente verticale della velocità turbolenta nell'atmosfera, il valore medio e la distribuzione di probabilità variano chiaramente da giorno a giorno. Ma se si considera la risposta di un aereo a

questa turbolenza, è ragionevole assumere che durante il volo (o parte di un volo) la turbolenza sia stazionaria.

C'è poi da considerare che, poiché tutti i processi random “ingegneristici” hanno un inizio ed una fine, essi non possono mai essere considerati realmente stazionari, ma, ai fini pratici, è molto spesso sufficiente assumere che un processo è stazionario per la maggior parte della sua storia temporale, oppure che esso può essere diviso in periodi, ognuno dei quali può essere approssimativamente considerato stazionario.

Il termine “debolmente stazionario” è usato per descrivere quei processi stocastici in cui solo le densità di probabilità del primo e secondo ordine sono invarianti con il tempo; un processo è invece “strettamente stazionario” quando sono invarianti rispetto al tempo tutte le densità di probabilità di ordine superiore.

7.7.2 Processo ergodico. Medie temporali

Consideriamo un processo stocastico stazionario (statisticamente), si consideri una sola funzione campione $x_k(t)$, e si calcolino la media e le densità di probabilità lungo tutta la lunghezza del segnale. Le medie così calcolate sono medie temporali. La media temporale del segnale si scrive, ovviamente:

$$\bar{x} = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T x(t) dt$$

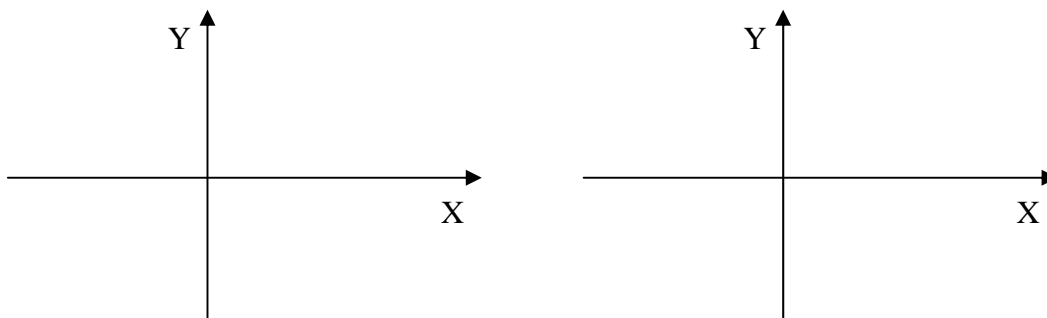
Se si verifica che le medie temporali sono uguali (o approssimativamente uguali) a quelle calcolate sull'insieme, il processo stazionario si dice **ergodico**. In questo caso ogni funzione campione è completamente rappresentativa del processo random, e pertanto si possono calcolare le medie temporali al posto delle medie di insieme. In altre parole, dalla misura di un solo campione possiamo predire le proprietà statistiche dell'intero processo.

E' molto difficile verificare sperimentalmente che un particolare processo sia ergodico (soprattutto è molto lungo). Se però si può assumere che esso è stazionario, può essere ragionevole, da un punto di vista ingegneristico, assumere che esso sia anche ergodico.

Si noti infine che se un processo è ergodico, esso deve anche essere stazionario, ma non è vero il contrario: infatti le medie calcolate lungo il campione sono indipendenti dal tempo che si considera variabile da $t = -\infty$ a $+\infty$; quindi anche le medie di insieme, che per definizione di ergodicità sono uguali a quelle sul campione generico, sono invarianti rispetto al tempo: da cui la stazionarietà.

7.8 Correlazione

Si consideri un insieme di coppie di 2 variabili random x e y : a queste coppie si possono far corrispondere dei punti sul piano xy (figura seguente).



Nelle precedenti figure sono rappresentati due casi estremi che si possono verificare: il primo è quello in cui non è possibile definire una qualsiasi relazione tra il valore della variabile x e quello di y ; il secondo rappresenta invece il caso in cui, ad un piccolo valore di x ne corrisponde uno piccolo di y e per un grande valore di x se ne trova uno proporzionalmente grande di y . In questo secondo caso pertanto si può parlare di “**correlazione**” tra le due variabili random in quanto è possibile cercare un legame di tipo funzionale, mentre nel primo dei casi esaminati non esiste alcun tipo di correlazione.

Se, nel secondo caso, si desidera esprimere un legame funzionale approssimato tra x ed y sotto forma lineare (una retta), un modo di procedere è quello di minimizzare il quadrato della deviazione del valore effettivo di y da quello prodotto dall'approssimazione usata. Se si assumono gli assi in modo che:

$$E[x] = E[y] = 0$$

e si fissa l'origine nel baricentro del sistema di punti, il legame funzionale è dato da: $y = mx$

La deviazione da minimizzare (metodo dei minimi quadrata) è $\Delta = y - mx$ e il valore medio del quadrato della deviazione è:

$$E[\Delta^2] = E[(y - mx)^2] = E[y^2] - 2mE[xy] + m^2E[x^2]$$

$$\frac{\partial E[\Delta^2]}{\partial m} = 0 = 2mE[x^2] - 2E[xy] \Rightarrow m = \frac{E[xy]}{E[x^2]}$$

l'equazione cercata è:

$$y = \frac{E[xy]}{E[x^2]} \cdot x$$

Alla precedente relazione si può dare una forma diversa considerando che:

$$\sigma_x^2 = E[x^2] \quad \sigma_y^2 = E[y^2]$$

Pertanto:

$$\frac{y}{\sigma_y} = \left\{ \frac{E[xy]}{\sigma_x \sigma_y} \right\} \frac{x}{\sigma_x}$$

L'equazione precedente è nota come **equazione della linea di regressione** di y su x . Analogamente si può scrivere l'equazione della linea di regressione di x su y come:

$$\frac{x}{\sigma_x} = \left\{ \frac{E[xy]}{\sigma_x \sigma_y} \right\} \frac{y}{\sigma_y}$$

Nel caso in cui il valor medio delle variabili random sia diverso da zero, le precedenti equazioni diventano:

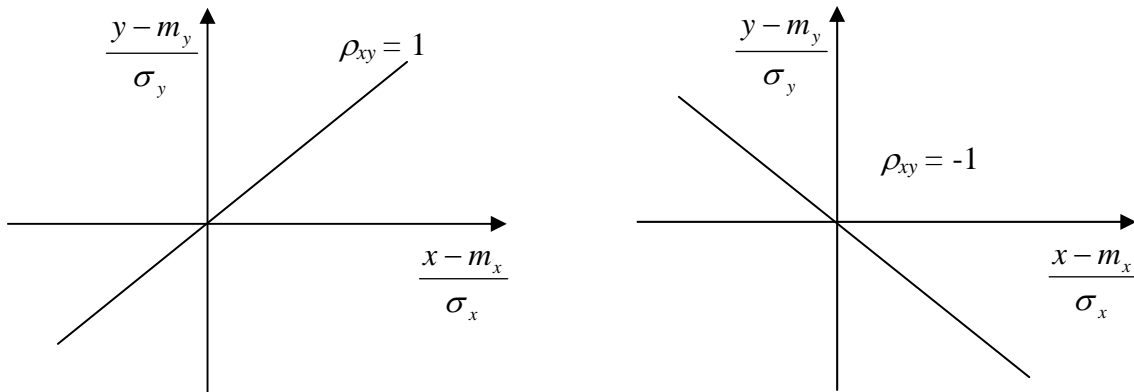
$$\frac{y - m_y}{\sigma_y} = \left\{ \frac{E[(x - m_x)(y - m_y)]}{\sigma_x \sigma_y} \right\} \frac{x - m_x}{\sigma_x}$$

$$\frac{x - m_x}{\sigma_x} = \left\{ \frac{E[(x - m_x)(y - m_y)]}{\sigma_x \sigma_y} \right\} \frac{y - m_y}{\sigma_y}$$

dove m_x e m_y sono, rispettivamente, i valori medi di x e y . Il parametro:

$$\rho_{xy} = \frac{E[(x - m_x)(y - m_y)]}{\sigma_x \sigma_y}$$

viene chiamato “*coefficiente di correlazione*” o “*covarianza normalizzata*”. Per $\rho \pm 1$ c’è una correlazione perfetta, per $\rho=0$ non c’è correlazione (vedi figure seguenti).



7.9 Descrizione dei segnali random nel dominio del tempo

Le medie e le distribuzioni di probabilità fanno perdere ogni informazione sulla periodicità e sulle frequenze presenti nella funzione campione. Medie e distribuzioni possono essere adeguate per considerazioni di carattere globale, ma altre informazioni possono essere ottenute attraverso l'analisi di altri indicatori statistici direttamente legati all'andamento temporale delle variabili random.

7.9.1 Funzione di autocorrelazione (o covarianza)

Una grandezza che prende in considerazione la periodicità dei segnali è la **funzione di autocorrelazione** o **covarianza** della variabile random $x(t)$ in due istanti t_1 e t_2 . Questa grandezza fornisce informazioni circa la dipendenza del valore di una variabile random in un istante rispetto al valore della variabile stessa in un altro istante. Come vedremo, essa riveste una notevole importanza nella caratterizzazione di flussi turbolenti.

La funzione di autocorrelazione è definita come il valor medio del prodotto $x(t)x(t + \tau)$, essendo il valor medio calcolato sull'insieme:

$$R_x(t, \tau) = E[x(t) \cdot x(t + \tau)]$$

Se il processo è stazionario ed ergodico, possiamo sostituire la media di insieme con la media temporale e scrivere $R_x(\tau)$ al posto di $R_x(t, \tau)$:

$$R_x(\tau) = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T x(t)x(t + \tau) dt$$

Perciò la funzione di autocorrelazione è solo funzione della differenza temporale τ o del ritardo τ tra i due punti campionati del segnale.

Si possono ora dedurre alcune proprietà di $R_x(\tau)$.

Se $x(t)$ è stazionario (statisticamente), il valore medio e la deviazione standard non dipendono dal tempo, cioè:

$$E[x(t)] = E[x(t + \tau)] = m$$

$$\sigma_x(t) = \sigma_x(t + \tau) = \sigma$$

Il coefficiente di correlazione, prima definito, per $x(t)$ e $x(t + \tau)$ è dato da:

$$\begin{aligned} \rho &= \frac{E[(x(t) - m)(x(t + \tau) - m)]}{\sigma^2} = \frac{E[(x(t)x(t + \tau)] - mE[x] - mE[x(t + \tau)] + m^2}{\sigma^2} = \\ &= \frac{E[(x(t)x(t + \tau)] - m^2}{\sigma^2} = \frac{R_x(\tau) - m^2}{\sigma^2} \end{aligned}$$

Deriva quindi che $R_x(\tau) = \sigma^2 \rho + m^2$, e, poiché i valori limiti di ρ sono ± 1 , si ha:

$$-\sigma^2 + m^2 \leq R_x(\tau) \leq \sigma^2 + m^2$$

Il valore della funzione di autocorrelazione non può quindi essere più grande del valore quadratico medio $E[x^2] = \sigma^2 + m^2$ e non può essere inferiore a $-\sigma^2 + m^2$.

Quando l'intervallo di separazione tra i due punti di misura è nullo, segue immediatamente che:

$$R_x(\tau = 0) = E[x^2]$$

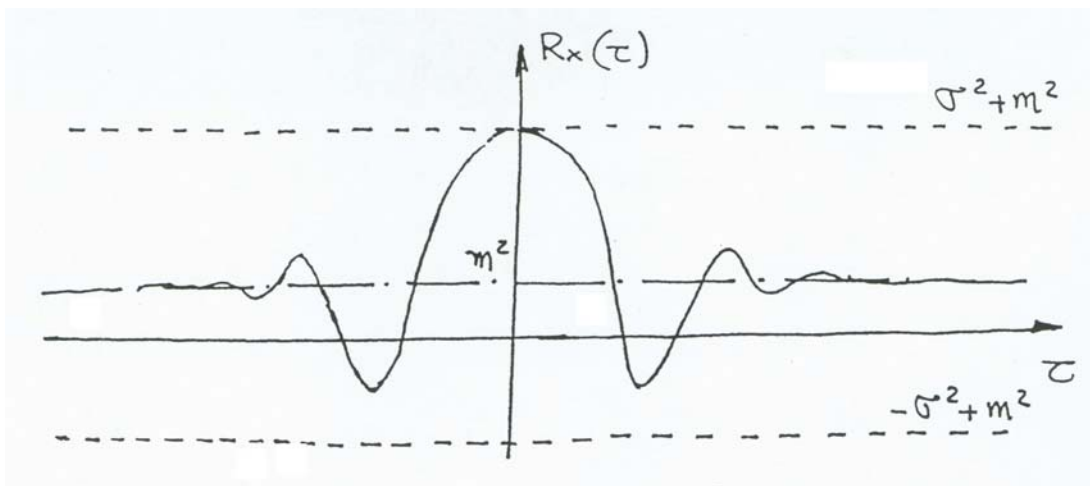
cioè il valore della funzione di autocorrelazione per $\tau = 0$ è uguale al valore quadratico medio del processo.

Per $\tau \rightarrow \infty$ invece, il valore del coefficiente di correlazione $\rho \rightarrow 0$, in quanto il processo non è correlato, cioè non c'è relazione coerente tra $x(t)$ e $x(t + \tau)$. In tal caso $R_x(\tau \rightarrow \infty) \rightarrow m^2$ (o a 0 se il valor medio è nullo).

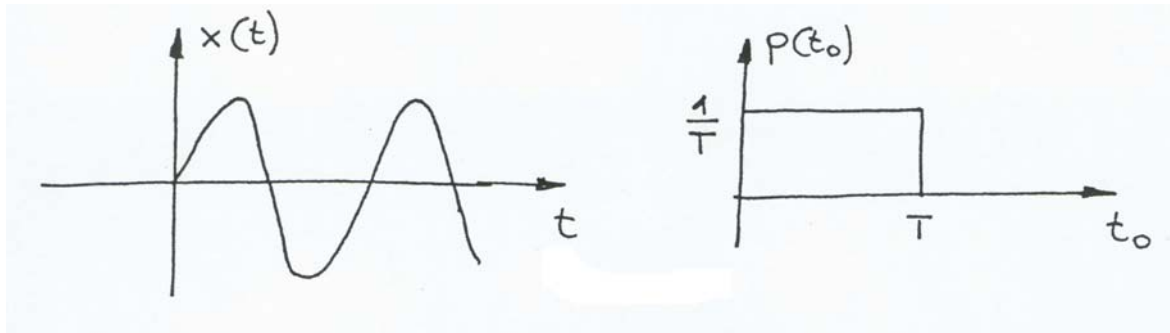
La funzione di autocorrelazione è una *funzione pari*. Infatti considerando che $R_x(\tau)$ è solo funzione di τ si può scrivere:

$$R_x(\tau) = E[x(t) \cdot x(t + \tau)] = E[x(t)x(t - \tau)] = R_x(-\tau)$$

Le proprietà ora viste possono essere illustrate nel seguente modo (figura seguente).



Si consideri ora il caso di un processo stocastico ergodico $x(t)$, le cui funzioni campione siano onde sinusoidali di uguale ampiezza e periodo $x(t) = \sin \omega t$: la fase sia una variabile random caratterizzata da una densità di distribuzione uniforme tra 0 e T (vedi figura seguente).



Si vuole calcolare la funzione di autocorrelazione.

Ci sono due modi di fare questo calcolo: il primo consiste nel calcolare una media di insieme guardando attraverso le funzioni campione in due istanti t_0 e $t_0 + \tau$. Il secondo fa uso del fatto che il processo è ergodico e pertanto ogni singola funzione campione è completamente rappresentativa del processo: tale modo di operare è quello di mediare un singolo campione pensando all'istante di campionamento t_0 come una variabile random uniformemente distribuita sull'asse reale. In tal caso:

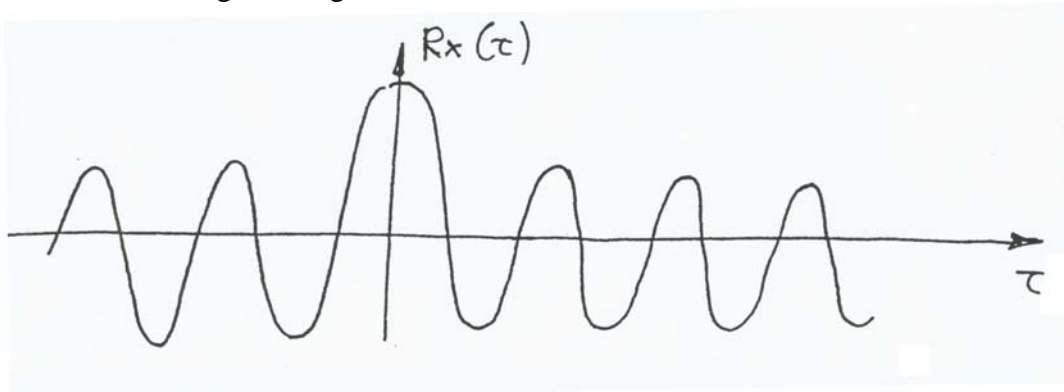
$$R_x(\tau) = E[x(t_0) \cdot x(t_0 + \tau)] = \int_{-\infty}^{+\infty} x(t_0)x(t_0 + \tau)p(t_0)dt_0$$

Poiché $x(t)$ è periodica, è necessario considerare soltanto un ciclo completo del tempo e come intervallo di valori di t_0 quello esteso tra 0 e T. Tutti i valori compresi in tale intervallo sono ugualmente probabili e quindi la distribuzione di probabilità è uniforme con valore $1/T$. Si ha pertanto:

$$R_x(\tau) = \int_0^T \frac{1}{T} \sin \omega t \sin \omega(t + \tau) dt = \frac{1}{T} \int_0^T \left(\sin^2 \omega t \cos \omega \tau + \frac{1}{2} \sin 2\omega t \sin \omega \tau \right) dt = \frac{1}{2} \cos \omega \tau$$

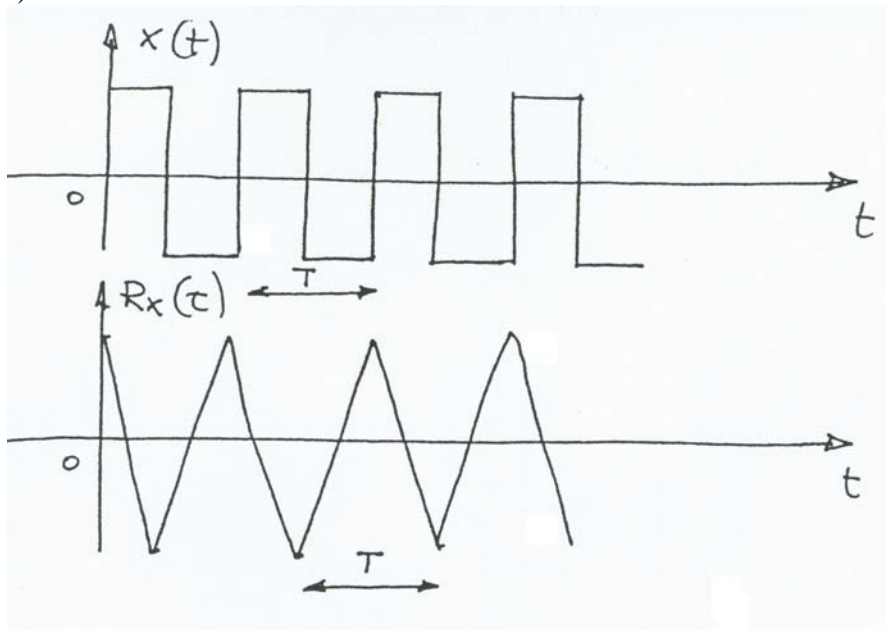
Da qui si può vedere che se $x(t)$ è periodica, allora anche $R_x(\tau)$ è periodica, e il valore massimo di $R_x(\tau)$ non si ha solo per $\tau = 0$ ma anche per valori di τ che sono multipli interi del periodo.

Riassumendo i risultati fin qui visti possiamo dire che se una variabile random ha una componente puramente random (senza periodicità) e una componente periodica, allora l'autocorrelazione sarà del tipo mostrato nella seguente figura.



Perciò la funzione di autocorrelazione viene spesso usata per identificare segnali periodici oscurati da rumore random.

Si noti infine che la funzione di autocorrelazione di un'onda quadra è dato da un'onda triangolare (figura seguente):



7.9.2 Funzione di cross-correlazione

Questa grandezza fornisce informazioni circa la dipendenza del valore di una variabile random in un istante rispetto al valore di un'altra variabile in un altro istante. Essa è definita nel seguente modo:

$$R_{xy}(\tau) = E[x(t)y(t + \tau)]$$

$$R_{yx}(\tau) = E[y(t)x(t + \tau)]$$

Per processi stazionari, caratterizzati dall'avere le medie di insieme indipendenti dal tempo, le precedenti relazioni si possono scrivere:

$$R_{xy}(\tau) = E[x(t - \tau)y(t)] = R_{yx}(-\tau)$$

$$R_{yx}(\tau) = E[y(t - \tau)x(t)] = R_{xy}(-\tau)$$

In generale $R_{xy}(\tau)$ e $R_{yx}(\tau)$ non sono le stesse e, contrariamente alla funzione di autocorrelazione, non sono funzioni pari. Ricordando l'espressione della covarianza normalizzata, la funzione di cross-correlazione può essere espressa nel seguente modo:

$$R_{xy}(\tau) = \sigma_x \sigma_y \rho_{xy}(\tau) + m_x m_y$$

$$R_{yx}(\tau) = \sigma_x \sigma_y \rho_{yx}(\tau) + m_x m_y$$

e, poiché i valori limite di ρ_{xy} sono ± 1 , si possono trovare gli intervalli di variazione delle funzioni di cross-correlazione:

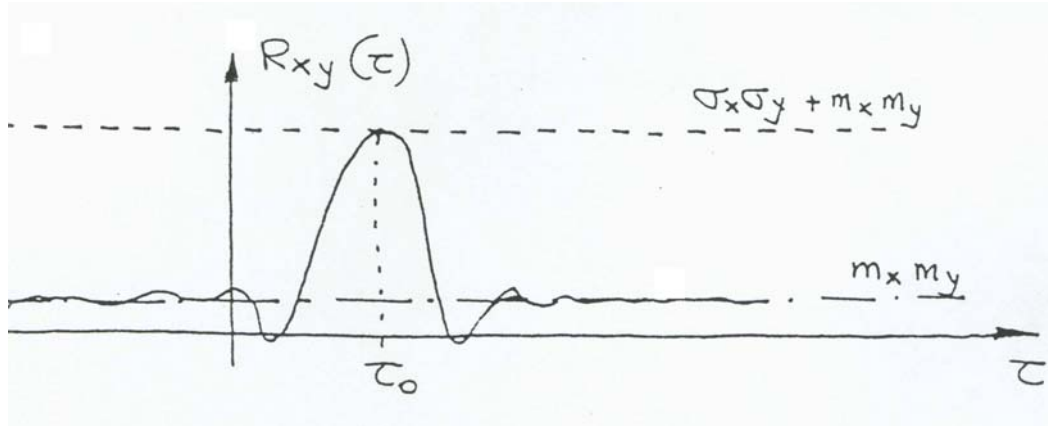
$$-\sigma_x \sigma_y + m_x m_y \leq R_{xy}(\tau) \leq \sigma_x \sigma_y + m_x m_y$$

Quando il tempo di separazione τ è molto grande, il coefficiente di correlazione tende a zero, per cui:

$$R_{xy}(\tau \rightarrow \infty) \rightarrow m_x m_y$$

$$R_{yx}(\tau \rightarrow \infty) \rightarrow m_y m_x$$

Queste proprietà sono illustrate nella seguente figura:



In questo caso si può vedere che i due processi random $x(t)$ e $y(t)$ mostrano la massima correlazione per $\tau = \tau_0$.

Esempio 5:

Si considerino due processi random $x(t)$ e $y(t)$, ognuno dei quali è costituito da funzioni campione che sono onde sinusoidali aventi la stessa ampiezza e frequenza. Un singolo campione del processo $x(t)$ è dato da:

$$x(t) = x_0 \sin(\omega t + \theta)$$

dove θ è l'angolo di fase. Se θ fosse lo stesso per tutte le funzioni campioni di $x(t)$, quest'ultimo non sarebbe un processo random. Si assume quindi che la fase vari, per ogni campione, in modo del tutto casuale. Se si assume che tutti gli angoli di fase tra 0 e 2π siano ugualmente probabili, si può scrivere che:

$$p(\theta) = \begin{cases} \frac{1}{2\pi} & \text{per } 0 \leq \theta \leq 2\pi \\ 0 & \text{altrove} \end{cases}$$

Ogni singolo campione del processo $y(t)$ è posto in relazione con uno del processo $x(t)$ nel seguente modo:

$$x(t) = x_0 \sin(\omega t + \theta)$$

$$y(t) = y_0 \sin(\omega t + \theta - \psi)$$

dove ψ è un angolo di fase costante ed uguale per tutti i componenti di $y(t)$. (Questo potrebbe essere il caso in cui $y(t)$ è derivato dal processo $x(t)$ da un sistema meccanico o elettrico che introduce nella propria risposta un ritardo di fase).

La funzione di cross-correlazione è:

$$R_{xy}(\tau) = E[x(t)y(t+\tau)] = E[x_0 y_0 \sin(\omega t + \theta) \sin(\omega t + \omega \tau + \theta - \psi)] = \\ = \int_0^{2\pi} x_0 y_0 \sin(\omega t + \theta) \sin(\omega t + \omega \tau + \theta - \psi) \frac{1}{2\pi} d\theta$$

Tenendo presente le formule di prostaferesi e sviluppando l'integrale si ottiene:

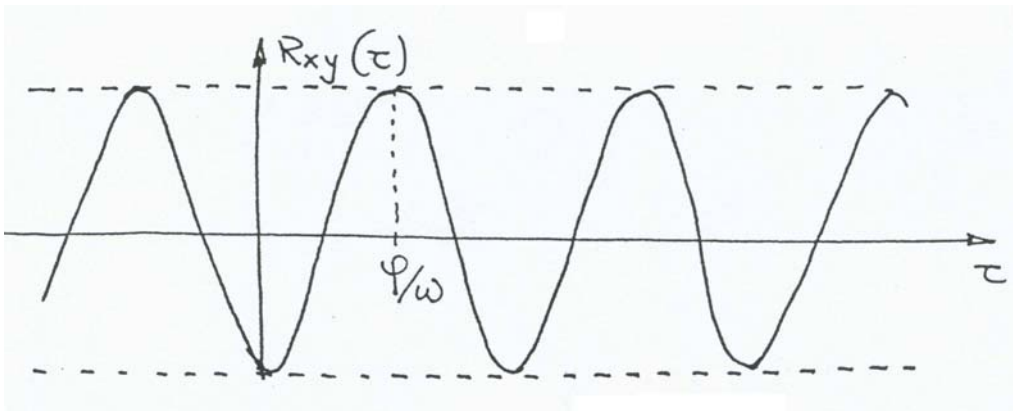
$$R_{xy}(\tau) = \frac{1}{2} x_0 y_0 \cos(\omega \tau - \psi)$$

L'altra funzione di cross-correlazione può essere calcolata in modo analogo, ottenendo:

$$R_{yx}(\tau) = \frac{1}{2} x_0 y_0 \cos(\omega \tau - \psi)$$

in accordo con la $R_{yx}(\tau) = R_{xy}(-\tau)$: infatti dovrebbe essere (vedi figura seguente):

$$R_{yx}(\tau) = R_{xy}(-\tau) = \frac{1}{2} x_0 y_0 \cos(-\omega \tau - \psi) = \frac{1}{2} x_0 y_0 \cos(\omega \tau + \psi)$$



Poiché è stato assunto che θ è uniformemente distribuito tra 0 e 2π , entrambi i processi $x(t)$ e $y(t)$ sono ergodici (se θ non fosse uniformemente distribuito le medie di insieme verrebbero a dipendere dal tempo). In base a questo si può affermare che esiste un'altra via per il calcolo delle funzioni di autocorrelazione: calcolare le medie, considerando due singoli campioni dei processi, in quanto tali campioni, se esiste la ergodicità, sono completamente rappresentativi dei processi stessi. Se si vuole seguire questo approccio, anziché calcolare le medie di insieme, si può procedere assumendo che le funzioni campione di $x(t)$ e $y(t)$ siano campionate ad un arbitrario istante di tempo t_0 . La variabile random è in questo caso l'istante t_0 la cui densità di probabilità si considera uniforme:

$$p(t_0) = \begin{cases} \frac{\omega}{2\pi} & \text{per } 0 \leq t_0 \leq \frac{2\pi}{\omega} \\ 0 & \text{altrove} \end{cases}$$

La funzione di cross-correlazione è quindi calcolata come media sul campione, cioè:

$$R_{xy}(\tau) = \int_0^{2\pi/\omega} x_0 y_0 \sin(\omega t + \theta) \sin(\omega t + \omega\tau + \theta - \psi) \frac{\omega}{2\pi} dt_0$$

che dà lo stesso risultato prima visto.

7.9.3 Applicazioni delle funzioni di correlazione

La principale applicazione della funzione di autocorrelazione è quella di stabilire l'influenza dei valori di una misura in un certo istante futuro. Poiché un segnale deterministico ha una autocorrelazione che persiste indefinitivamente nel tempo in contrapposizione a uno random che decade a zero per grandi ritardi temporali (assumendo che il valor medio sia nullo), una misura di autocorrelazione fornisce chiaramente un metodo per identificare dati deterministico offuscato dal rumore ambiente.

Alcune applicazioni della cross-correlazione sono le seguenti:

a) **Misure di tempi di ritardo:**

Si supponga di essere interessati a determinare il tempo che un segnale prende per attraversare un dato sistema. Se il sistema è lineare, una misura di cross-correlazione tra l'ingresso e l'uscita fornisce direttamente l'informazione su questo ritardo. La funzione di cross-correlazione infatti presenta un picco in corrispondenza al tempo richiesto perché il segnale passi attraverso il sistema.

Si deve notare però che questa tecnica spesso cade in difetto perché la velocità di trasmissione attraverso il sistema è funzione della frequenza. In questo caso al posto della cross-correlazione si può utilizzare la densità spettrale incrociata.

b) **Determinazione delle vie di trasmissione:**

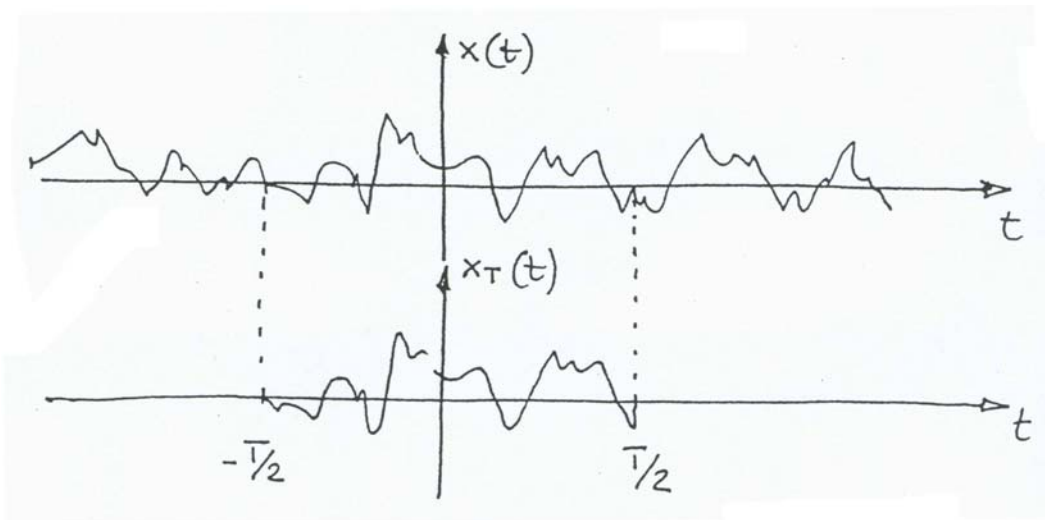
Si consideri il caso di un sistema lineare dove l'ingresso può seguire due o più vie di trasmissione per produrre un determinato output. (Per esempio, il funzionamento di macchinari in un'industria può indurre un livello di rumore e di vibrazioni eccessivo nelle aree destinate agli uffici: l'energia può essere trasferita in molti modi attraverso le strutture dell'edificio o, acusticamente, attraverso l'aria). Prima di effettuare operazioni di controllo, è necessario stabilire la corretta via di trasmissione. Questi problemi possono spesso essere risolti facendo misure di cross-correlazione tra l'ingresso e l'uscita. Poiché ogni via di trasmissione attraverso il sistema sarà generalmente associata con un differente ritardo, si verificheranno picchi separati nel cross-correlogramma per ogni via di trasmissione che contribuisce significativamente all'uscita. Se i tempi di ritardo associati con le possibili vie di trasmissione possono essere stimati, i ritardi attesi possono essere paragonati con i ritardi misurati dal cross-correlogramma per identificare le vie di trasmissione che contribuiscono significativamente all'uscita.

c) **Riconoscimento di segnali coperti da rumore:**

La funzione di autocorrelazione è utile per riconoscere un segnale periodico coperto da rumore, ma essa non può estrarre un segnale random da un rumore estraneo. In questo caso bisogna usare la cross-correlazione.

7.10 Descrizione dei segnali random nel dominio della frequenza

Fino ad ora le nostre considerazioni sui processi stocastici sono state ristrette al dominio temporale. Analizzeremo ora il dominio delle frequenze. Si è visto che le trasformate di Fourier possono essere applicate ai fenomeni periodici e transitori. Vedremo che, in qualche modo, queste trasformate possono anche essere applicate ai processi aleatori. Si consideri una funzione campione del processo random (figura seguente) e si prenda in considerazione una sua versione troncata $x_T(t)$, tale che $x_T = x(t)$ per $|t| < T/2$ e zero altrove:



Studiamo la decomposizione in frequenza della potenza del segnale. La trasformata di Fourier del segnale è:

$$x_T(t) = \int_{-\infty}^{+\infty} X_T(f) e^{j2\pi ft} df$$

L'energia totale del segnale è: $\int_{-\infty}^{+\infty} x_T^2(t) dt$

E questa quantità tende ad infinito per $T \rightarrow \infty$. E' pertanto utile considerare la media $\frac{1}{T} \int_{-\infty}^{+\infty} x_T^2(t) dt$ che ha le dimensioni di una **potenza**.

Da teorema di Parseval si ha:

$$\frac{1}{T} \int_{-\infty}^{+\infty} x_T^2(t) dt = \frac{1}{T} \int_{-T/2}^{+T/2} x_T^2(t) dt = \frac{1}{T} \int_{-\infty}^{+\infty} |X_T(f)|^2 df$$

Per $T \rightarrow \infty$ si ha: $\lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{+T/2} x_T^2(t) dt = \int_{-\infty}^{+\infty} \lim_{T \rightarrow \infty} \frac{|X_T(f)|^2}{T} df$

La quantità $|X_T(f)|^2 / T$ è chiamata lo spettro del campione e a questa quantità si dà la notazione $\hat{S}_{xx}(f)$. Poiché il termine a primo membro della precedente equazione è la potenza media della funzione campionata, è abbastanza logico pensare a:

$$\lim_{T \rightarrow \infty} \frac{|X_T(f)|^2}{T}$$

come una **densità spettrale di potenza**.

Sfortunatamente ciò non ha molto senso in quanto la precedente quantità non converge in senso statistico: si può dimostrare che se la precedente quantità è valutata utilizzando un $T' > T$, il risultato non risulta migliorato (ciò è dovuto al fatto che questo risultato è ottenuto attraverso una stima per la quale la proprietà di ergodicità non vale). Si può allora introdurre una nuova funzione, mediando questa quantità per rimuovere questo comportamento non corretto. Si considera perciò la media:

$$E \left[\lim_{T \rightarrow \infty} \frac{|X_T(f)|^2}{T} \right]$$

Per quanto sia statisticamente più corretto definire la funzione densità spettrale di potenza in questo modo, e sia questo poi il modo in cui la densità spettrale di potenza venga di fatto calcolata, qui per semplicità e analogia con i testi più autorevoli sull'argomento, chiameremo **densità spettrale di potenza** la funzione:

$$S_{xx}(f) = \lim_{T \rightarrow \infty} \frac{|X_T(f)|^2}{T}$$

oppure:

$$S_{xx}(\omega) = \lim_{T \rightarrow \infty} \frac{|X_T(\omega)|^2}{T}$$

Si noti che il valore quadratico medio della funzione $x(t)$ è:

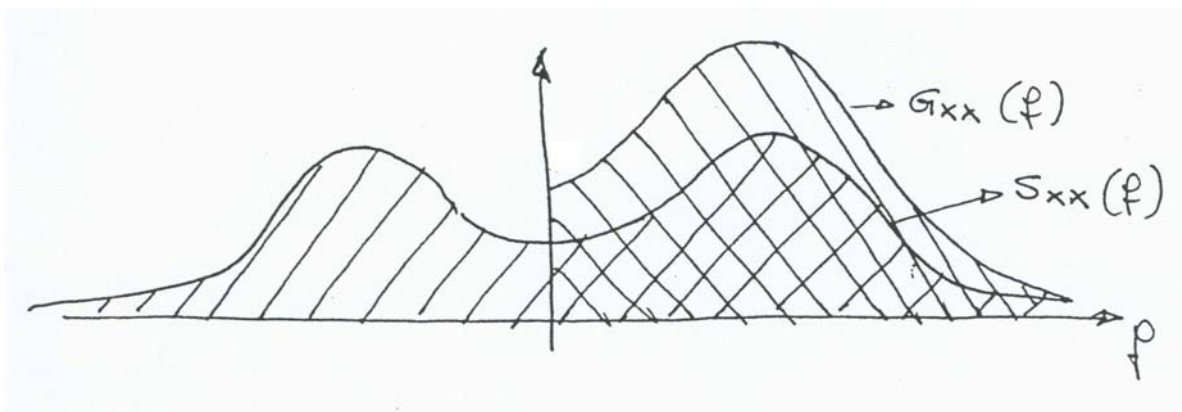
$$E[x^2(t)] = \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{+T/2} x_T^2(t) dt = \int_{-\infty}^{+\infty} \lim_{T \rightarrow \infty} \frac{|X_T(f)|^2}{T} df$$

e pertanto:

$$E[x^2(t)] = \int_{-\infty}^{+\infty} S_{xx}(f) df = \frac{1}{2\pi} \int_{-\infty}^{+\infty} S_{xx}(\omega) d\omega$$

cioè in questo caso (per segnali random) $S_{xx}(f)df$ è l'equivalente di $\frac{1}{2}c_n^2$ per i segnali periodici.

$S_{xx}(f)$ è una funzione reale pari, ed è definita nel campo $-\infty < f < +\infty$. L'integrale di questa funzione su tutto il campo è il valore quadratico medio del segnale. Talvolta è più conveniente usare, anziché la S_{xx} definita tra $-\infty$ e $+\infty$, la $G_{xx}(f)$ definita solamente per valori positivi di f . Sussiste allora ovviamente la relazione $G_{xx}(f) = 2S_{xx}(f)$ (vedi figura seguente).



Si noti che in genere la conoscenza della densità spettrale di un processo random non consente di determinare la sua densità di probabilità. Se però un processo è gaussiano, l'intera distribuzione è definita dal suo valore quadratico medio o varianza σ^2 , dalla relazione:

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

avendo ipotizzato che il valore medio sia nullo. Poiché in questo caso:

$$E[x^2(t)] = \sigma^2 = \int_{-\infty}^{+\infty} S_{xx}(f) df$$

una conoscenza della densità spettrale è sufficiente a dare la densità di probabilità e quindi la funzione di distribuzione, se necessaria. E' quindi una fortuna che tanti processi fisici aleatori abbiano una distribuzione di probabilità circa gaussiana.

Se una variabile random ha densità spettrale di potenza costante su tutto il campo di frequenza, viene chiamata **rumore bianco**. In pratica il termine rumore bianco viene usato per descrivere una funzione eccitante che ha la densità di potenza costante nel campo di frequenze che viene analizzato.

7.11 Relazioni tra la densità spettrale e la funzione di correlazione

Poiché sia la funzione densità spettrale che la funzione di autocorrelazione descrivono entrambe la stessa forma d'onda, è logico pensare che ci sia un legame tra di esse. Infatti le due funzioni sono legate tramite le trasformate di Fourier nel seguente modo:

$$R_{xx}(\tau) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} S_{xx}(\omega) e^{j\omega\tau} d\omega$$

$$S_{xx}(\omega) = \int_{-\infty}^{+\infty} R_{xx}(\tau) e^{j\omega\tau} d\tau$$

oppure:

$$R_{xx}(\tau) = \frac{1}{2\pi} \int_0^{+\infty} G_{xx}(\omega) \cos \omega\tau d\omega$$

$$G_{xx}(\omega) = 4 \int_0^{+\infty} R_{xx}(\tau) \cos \omega\tau d\tau$$

Queste relazioni sono note con il nome di **Wiener -Khintchine**. Esse indicano chiaramente che la densità spettrale di potenza è la trasformata di Fourier della funzione di correlazione e viceversa.

Alle relazioni precedenti si può pervenire partendo dalla definizione $S_{xx}(f) = \lim_{T \rightarrow \infty} \frac{|X_T(f)|^2}{T}$.

Si supponga che $x_T(t)$ coincida esattamente con la funzione $x(t)$ nel campo $-T/2 \leq t \leq T/2$ e sia nullo altrove. Si può scrivere:

$$R_{x_T x_T}(\tau) = E[x_T(t)x_T(t+\tau)]$$

Allora:

$$\begin{aligned} \int_{-\infty}^{+\infty} R_{x_T x_T}(\tau) e^{-j\omega\tau} d\tau &= \int_{-\infty}^{+\infty} E[x_T(t)x_T(t+\tau)] e^{-j\omega\tau} d\tau = \\ &= \int_{-\infty}^{+\infty} \frac{1}{T} \int_{-\infty}^{+\infty} x_T(t)x_T(t+\tau) dt e^{-j\omega\tau} d\tau = \end{aligned}$$

$$\begin{aligned}
&= \frac{1}{T} \int_{-\infty}^{+\infty} \left[\int_{-\infty}^{+\infty} x_T(t) x_T(t+\tau) e^{-j\omega\tau} d\tau \right] dt = \\
&= \frac{1}{T} \int_{-\infty}^{+\infty} \left[\int_{-\infty}^{+\infty} x_T(t) e^{j\omega t} x_T(t+\tau) e^{-j\omega(t+\tau)} dt \right] d\tau = \\
&= \frac{1}{T} \int_{-\infty}^{+\infty} x_T(t) e^{j\omega t} dt \int_{-\infty}^{+\infty} x_T(\xi) dt e^{-j\omega\xi} d\xi = \frac{1}{T} X_T^*(f) X_T(f)
\end{aligned}$$

Pertanto:

$$S_{xx}(f) = \int_{-\infty}^{+\infty} R_{xx}(\tau) e^{-j\omega\tau} d\tau = \lim_{T \rightarrow \infty} \frac{1}{T} [X_T^*(f) X_T(f)] = \lim_{T \rightarrow \infty} \frac{1}{T} |X_T(f)|^2 \quad \text{c.v.d.}$$

7.12 Cross-correlazione e densità spettrale incrociata

Già si è avuto occasione di parlare della cross-correlazione tra due funzioni $x_1(t)$ e $x_2(t)$:

$$R_{x_1 x_2}(\tau) = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T x_1(t) x_2(t+\tau) dt$$

(Se il processo è stazionario $R_{x_1 x_2}(t, \tau)$ è indipendente da t). La densità di potenza incrociata è ora definita come trasformata di Fourier della cross-correlazione. Poiché la cross-correlazione non è una funzione pari, la corrispondente trasformata di Fourier è una funzione complessa:

$$\begin{aligned}
S_{x_1 x_2}(\omega) &= \int_{-\infty}^{+\infty} R_{x_1 x_2}(\tau) e^{-j\omega\tau} d\tau \\
R_{x_1 x_2}(\omega) &= \frac{1}{2\pi} \int_{-\infty}^{+\infty} S_{x_1 x_2}(\omega) e^{j\omega\tau} d\omega
\end{aligned}$$

La densità di potenza incrociata (spesso chiamata **cross-spettro**) è particolarmente utile per studiare la risposta dei sistemi ad eccitazioni random e soprattutto per valutare la risposta complessa in frequenza.

7.13 Applicazioni delle densità spettrali

La principale applicazione della densità spettrale è quella di stabilire la composizione in frequenza dei dati random misurati. Inoltre essa fornisce un mezzo particolarmente valido per caratterizzare un sistema fisico. Anche in questo caso vedremo delle utili interpretazioni fisiche della densità spettrale nel caso di flussi turbolenti.

Per esempio si consideri un sistema caratterizzato da una risposta in frequenza $H(f)$, e si consideri un segnale random stazionario con densità spettrale $G_x(f)$ applicato all'ingresso del sistema. L'uscita del sistema sarà un segnale random stazionario caratterizzato da una densità spettrale:

$$G_y(f) = |H(f)|^2 G_x(f)$$

Pertanto la conoscenza di due di queste tre quantità consente di stimare la terza. Si noti che in questa relazione compare solo l'ampiezza della risposta in frequenza, ma non si hanno informazioni

sulla fase. La determinazione della fase necessita di una analisi di densità di potenza incrociata. Infatti in questo caso si ha:

$$G_{xy}(f) = H(f)G_x(f)$$

Questa relazione consente di determinare completamente la risposta complessa in frequenza di sistemi lineari.

Un'altra applicazione della densità di potenza incrociata è quella di fornire informazioni sui tempi di ritardo nel caso in cui lo sfasamento tra ingresso e uscita sia dipendente dalla frequenza.

Se $\theta_{xy}(f)$ è l'angolo di fase tra ingresso e uscita, il tempo di ritardo attraverso il sistema è:

$$\tau = \frac{\theta_{xy}(f)}{2\pi f}$$

E' così possibile determinare i tempi di ritardo in funzione della frequenza, ciò che non era direttamente ottenibile dalle misure di cross-correlazione.

7.14 La funzione di coerenza

La funzione di coerenza tra l'ingresso a un sistema $x(t)$ e l'uscita $y(t)$ è una quantità reale definita come:

$$\gamma_{xy}^2(f) = \frac{|G_{xy}(f)|^2}{G_x(f)G_y(f)} = \frac{|S_{xy}(f)|}{S_x(f)S_y(f)}$$

con ovvia notazione dei simboli.

Si può facilmente dimostrare che, per ogni f , la funzione di coerenza soddisfa la relazione:

$$0 \leq \gamma_{xy}^2(f) \leq 1$$

Per poter fornire la precedente relazione per la coerenza, bisogna assumere che sia $G_x(f)$ che $G_y(f)$ siano diversi da 0 e non contengano funzioni delta. Per eliminare le funzioni delta all'origine occorre rimuovere dai dati il valor medio (se diverso da zero) di ogni funzione prima di applicare la precedente relazione.

Considerando la relazione che intercorre tra l'ingresso e uscita di ogni sistema lineare a parametri costanti, la coerenza può anche essere scritta nel seguente modo:

$$\gamma_{xy}^2(f) = \frac{|H(f)|^2 G_x^2(f)}{G_x(f)|H(f)|^2 G_y(f)} = 1$$

dalla quale risulta che, per un tale sistema, la coerenza è sempre unitaria. Se invece $x(t)$ e $y(t)$ sono completamente incorrelati, la coerenza è nulla, mentre se la coerenza è compresa tra 0 e 1, sono possibili le seguenti situazioni:

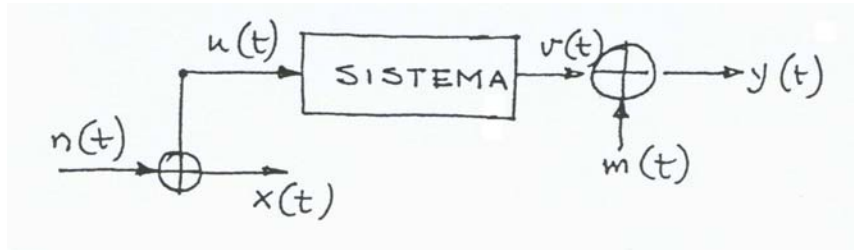
- a) c'è rumore estraneo presente nelle misure;
- b) il sistema che lega $x(t)$ e $y(t)$ è non lineare;
- c) $y(t)$ è in parte causato da $x(t)$, in parte da altri ingressi.

7.14.1 Effetto del rumore

Si consideri (figura seguente) che l'ingresso misurato $x(t)$ e l'uscita misurata $y(t)$ siano costituiti dai segnali veri $u(t)$ e $v(t)$ rispettivamente e dai rumori incorrelati $n(t)$ e $m(t)$. L'ingresso e l'uscita misurate sono quindi:

$$x(t) = u(t) + n(t)$$

$$y(t) = v(t) + m(t)$$



Le densità spettrali sono allora:

$$G_x(f) = G_u(f) + G_n(f)$$

$$G_y(f) = G_v(f) + G_m(f)$$

$$G_{xy}(f) = G_{uv}(f)$$

Per questo problema la *coerenza teorica* è:

$$\gamma_{uv}^2(f) = \frac{|G_{uv}(f)|^2}{G_u(f) G_v(f)}$$

mentre quella *misurata* è:

$$\begin{aligned} \gamma_{xy}^2(f) &= \frac{|G_{xy}(f)|^2}{G_x(f) G_y(f)} = \frac{|G_{uv}(f)|^2}{[G_u(f) + G_n(f)][G_v(f) + G_m(f)]} = \\ &= \frac{|G_{uv}(f)|^2}{1 + \frac{G_n(f)}{G_u(f)} + \frac{G_m(f)}{G_v(f)} + \frac{G_n(f)G_m(f)}{G_u(f)G_v(f)}} \leq \gamma_{uv}^2(f) \end{aligned}$$

La precedente relazione mostra ciò che succede quando rumore estraneo e incorrelato è presente sia all'ingresso che all'uscita. Si osservi che la coerenza misurata è teoricamente sempre minore della coerenza teorica quando esiste rumore o all'ingresso o all'uscita. Si noti inoltre che il contributo di ingressi incorrelati diversi dall'ingresso misurato $x(t)$ porta a un rumore incorrelato all'uscita.

7.15 Processi random a banda stretta e a banda larga

Le densità spettrali forniscono una misura della rappresentazione in frequenza di un processo random.

Per convenienza si presenterà la discussione in termini di processi random ergodici. I processi random sono spesso identificati dalla forma delle loro densità spettrali. In particolare distingueremo tra *processi a banda stretta* e *processi a banda larga*.

Un **processo a banda stretta** è caratterizzato da una densità spettrale di potenza concentrata prevalentemente intorno a una banda stretta di frequenze e presenta un picco piuttosto preminente. Un campione temporale di un tale processo contiene perciò un campo limitato di frequenze.

D'altra parte nel caso di un **processo a banda larga** la densità spettrale ha valori significativi su una larga banda di frequenza con ampiezza dello stesso ordine di grandezza della banda di frequenza intorno alla frequenza centrale. Ai due estremi del campo troviamo quindi:

- a) la densità spettrale di potenza costituita da una funzione delta, corrispondente a una funzione campione sinusoidale;
- b) una densità di potenza uniforme, corrispondente ad una funzione campione in cui tutte le frequenze sono ugualmente rappresentate.

La prima è naturalmente una *funzione deterministica* che può essere riguardata come random se l'angolo di fase è distribuito in modo random.

Il secondo caso è noto come *rumore bianco*: questo concetto è fisicamente impossibile perché implica un valore quadratico medio infinito, e quindi potenza infinita.

7.16 Stima dei segnali random

In questo paragrafo verranno analizzati e discussi i metodi di stima dei segnali random, basati su una singola registrazione del processo, considerando sia gli aspetti statistici che quelli computazionali.

7.16.1 Errori di stima ed accuratezza

Si supponga che il valor medio di un processo random sia μ_x (valore per noi incognito) e che si disponga solamente di una registrazione di lunghezza T del segnale. Una ragionevole stima \bar{x} di μ_x è:

$$\bar{x} = \frac{1}{T} \int_0^T x(t) dt$$

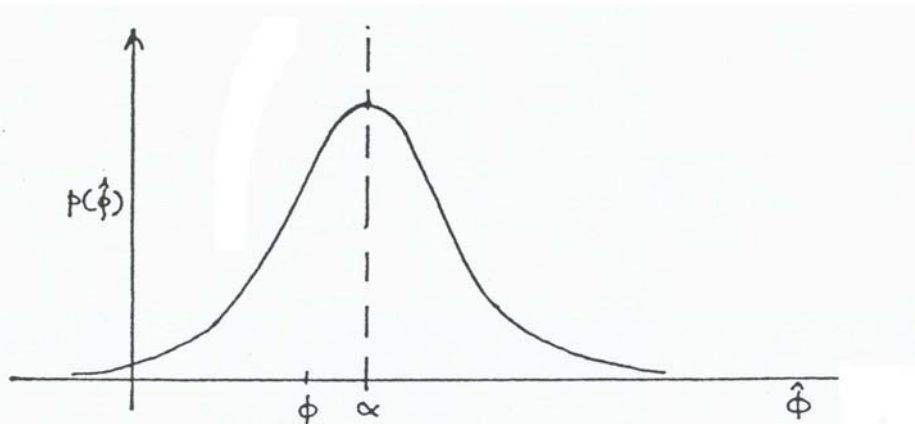
Il valore ottenuto è un valore campione di una variabile random \bar{X} , avente la sua propria distribuzione di probabilità e \bar{x} è una singola determinazione. Ogni \bar{x} calcolata da una differente registrazione di lunghezza T sarà in generale differente. Affinché il procedimento di stima sia utile, è opportuno che:

- lo scarto dei valori di \bar{x} non sia troppo grande e sia prossimo a μ_x ;
- più dati sono disponibili (più grande è T), migliore sia la stima.

Si possono formalizzare queste idee come segue.

Sia Φ il parametro che vogliamo stimare (nell'esempio precedente era μ_x) e sia $\hat{\Phi}$ uno stimatore di Φ . $\hat{\Phi}$ è ovviamente una variabile random con la sua propria distribuzione di probabilità, come

quella mostrata nella seguente figura, dove si può vedere che le stime di una variabile random si distribuiscono prevalentemente intorno ad α .



La distribuzione di probabilità $p(\hat{\Phi})$ è spesso difficile da ottenere e pertanto si definiscono alcune proprietà.

Errore di Bias (Distorsione):

Il bias di uno stimatore è definito nel seguente modo: $b(\hat{\Phi}) = E[\hat{\Phi}] - \Phi$

Esso è la differenza tra la media dello stimatore e il valore vero (incognito). Se $b(\hat{\Phi}) = 0$, allora $\hat{\Phi}$ è non distorto (unbiased). Mentre sembrerebbe opportuno usare uno stimatore non distorto, è talvolta prudente tollerare qualche distorsione purché si riesca a ridurre la variabilità della stima (relativamente a uno stimatore non distorto).

Varianza:

La varianza di uno stimatore è una misura della dispersione di $\hat{\Phi}$ rispetto al suo proprio valor medio.

$$\sigma^2[\hat{\Phi}] = E\left[(\hat{\Phi} - E(\hat{\Phi}))^2\right]$$

La radice quadratica della varianza è la deviazione standard $\sigma[\hat{\Phi}]$ dello stimatore. Sarebbe auspicabile che questo valore fosse piccolo, cioè che la densità di probabilità avesse un picco molto rilevante: purtroppo questo requisito spesso fa aumentare il bias.

Una misura della importanza relativa del bias e della varianza è il valore quadratico medio.

Valore quadratico medio (m.s.e.):

Il valore quadratico medio di uno stimatore è una misura della dispersione di $\hat{\Phi}$ rispetto a Φ , cioè:

$$m.s.e.[\hat{\Phi}] = \sigma^2[\hat{\Phi}] + b^2[\hat{\Phi}]$$

ciò mostra che il valore quadratico medio riflette sia la varianza che il bias.

Se si minimizza m.s.e. quando il bias è nullo, allora $\hat{\Phi}$ è uno stimatore non distorto a varianza minima. Se lo m.s.e. si riduce quando aumenta la dimensione del campione (quantità di dati usata), allora lo stimatore si dice *consistente*.

Talvolta gli errori sono resi adimensionali dividendo per la quantità che si desidera stimare (purché diversa da zero). Con le precedenti definizioni, si ottiene:

$$\varepsilon_r = \frac{\sigma[\hat{\Phi}]}{\Phi} \quad \varepsilon_b = \frac{b[\hat{\Phi}]}{\Phi} \quad \varepsilon_q = \frac{[m.s.e.(\hat{\Phi})]^{1/2}}{\Phi}$$

che sono, rispettivamente, l'errore standard, l'errore di bias e l'errore quadratico medio.

7.16.2 Intervalli di confidenza

Le stime $\hat{\Phi}$ a cui ci siamo riferiti prima sono stime puntuali, cioè singoli valori. E' spesso desiderabile sapere che un certo parametro cade probabilmente in un certo intervallo di valori. Per esempio, se la stima di un valore medio è $\hat{x} = 5$, probabilmente sarebbe utile sapere che μ_x è compreso nell'intervallo $4.5 \div 5.5$. Queste sono stime di intervallo: esse vengono chiamati *intervalli di confidenza* quando si attacca ad esse un valore che descrive la probabilità di un parametro di cadere dentro l'intervallo. Se, per esempio, stabiliamo che l'intervallo di confidenza per μ_x di stare tra 4.5 e 5.5 è 95%, questo significa che confidiamo che nel 95% dei casi μ_x cadrà in quell'intervallo. Si noti che questo non significa che la probabilità di μ_x di cadere nel suddetto intervallo sia 95%, perché μ_x non è una variabile random e non possiamo assegnare una probabilità ad essa. Piuttosto ciò significa che se potessimo analizzare un gran numero di campioni e trovare un intervallo di confidenza per μ_x per ogni campione, allora circa il 95% di questi intervalli conterrebbero μ_x .

Per calcolare gli intervalli di confidenza, si deve conoscere la distribuzione di probabilità dello stimatore.

7.17 Stimatore di un processo stocastico

Qui verrà dato un sommario su:

- definizione di stimatori comunemente usati;
- alcune proprietà statistiche degli stimatori;
- schemi di calcolo digitale per il calcolo delle stime.

Ci riferiamo a realizzazioni di un processo random, continuo e statisticamente stazionario. I dati si suppongono campionati con frequenze di campionamento sufficientemente alte da evitare aliasing.

7.17.1 Valore medio

Definizione: Uno stimatore per il valor medio di un campione lungo T è: $\hat{x} = \frac{1}{T} \int_0^T x(t) dt$

Proprietà statistiche: La stima è non distorta e il m.s.e. è:

$$m.s.e.(\hat{x}) = \frac{1}{T} \int_0^T \left(1 - \frac{|\tau|}{T}\right) R_{xx}(\tau) d\tau$$

Questo decresce all'aumentare di T, assicurando che lo stimatore è consistente. Questo risultato è utile per capire quanto deve lungo il tempo di acquisizione di una misura di quantità random (ad esempio la velocità in un flusso turbolento) fissando a priori un valore del m.s.e. accettabile.

Calcolo: Per dati campionati $\hat{x} = \frac{1}{N} \sum_{n=0}^{N-1} x(n)$

7.17.2 Valore quadratico medio

Definizione: Uno stimatore per lo m.s.e. da un campione lungo T è: $\hat{\Phi}^2 = \frac{1}{T} \int_0^T x^2(t) dt$

Proprietà statistiche: Lo stimatore è non distorto.

Calcolo: $\hat{\Phi}^2 = \frac{1}{N} \sum_{n=0}^{N-1} x^2(n)$

7.17.3 Varianza

Definizione: Uno stimatore per la varianza è: $\sigma^2 = \frac{1}{T} \int_0^T (x(t) - \bar{x})^2 dt$

Proprietà statistiche: Con l'uso della formula riportata qui di seguito, la varianza risulta non distorta. La varianza può anche essere calcolata dalla stima della funzione di covarianza per ritardo nullo.

Calcolo:
$$\sigma^2 = \frac{1}{N-1} \sum_{n=0}^{N-1} (x(n) - \bar{X})^2$$

7.17.4 Funzioni di correlazione – Autocorrelazione (autocovarianza)

Definizione: Se si ha un segnale definito in $0 \leq t \leq T$, si usano di solito due stimatori per $R_{xx}(\tau)$:

$$\hat{R}_{xx} = \frac{1}{T} \int_0^{T-|\tau|} x(t) x(t+|\tau|) dt \quad \text{per} \quad 0 \leq |\tau| \leq T$$

$$\hat{R}_{xx} = 0 \quad \text{per} \quad |\tau| > T$$

oppure:

$$\hat{R}'_{xx} = \frac{1}{T-|\tau|} \int_0^{T-|\tau|} x(t) x(t+|\tau|) dt \quad \text{per} \quad 0 \leq |\tau| \leq T$$

$$\hat{R}'_{xx} = 0 \quad \text{per} \quad |\tau| > T$$

La letteratura tecnica sembra preferire il secondo stimatore perché esso è non distorto, sebbene dovrebbe essere il m.s.e. il criterio per la scelta tra i due.

Considerazioni statistiche: Per τ nel campo $0 \leq |\tau| \leq T$, \hat{R}'_{xx} è non distorto, mentre

$$E[\hat{R}_{xx}] = R_{xx}(\tau) \left[1 - \frac{|\tau|}{T} \right]$$

cioè la prima definizione produce uno stimatore asintoticamente non distorto. Nella letteratura viene anche dimostrato che il primo stimatore è consistente. Quando $\tau \ll T$, non c'è molta differenza tra uno stimatore e l'altro, ma quando τ si avvicina a T , la varianza del secondo stimatore diverge.

Calcolo: Il calcolo del primo stimatore si ottiene con la formula:

$$\hat{R}_{xx}(m) = \frac{1}{N} \sum_{n=0}^{N-|m|-1} x(n) x(n+|m|)$$

mentre nel secondo bisogna dividere per $n - |m|$ anziché per N ; m è il ritardo e può assumere valori $0 \leq m \leq M - 1$, essendo N il massimo valore permesso per M . I calcoli da sviluppare nella relazione precedente sono molto lunghi se M non è piccolo, ed è preferibile utilizzare la FFT della densità spettrale di potenza per calcolare le correlazioni, con notevole vantaggio computazionali.

7.17.5 Densità spettrale di potenza

Definizione: Una stima per la densità spettrale da dati lunghi $2T$, (tra $-T$ e T) è:

$$\hat{S}_{xx}(f) = \frac{|X_T(f)|^2}{T} = \int_{-T}^T \hat{R}_{xx}(\tau) e^{-j2\pi f\tau} d\tau$$

Questa è una stima bruta in quanto risulta che la variabilità dello stimatore è indipendente dalla lunghezza dei dati: non può pertanto essere usato come stimatore. E' un esempio di stimatore per il quale non vale la proprietà di ergodicità.

Per ridurre le fluttuazioni dovute al campionamento, la stima deve essere "lisciata" ("smoothed") nel dominio delle frequenze, e questo può essere ottenuto moltiplicando la stima della funzione di autocorrelazione per una finestra di ritardo $\omega(\tau)$ che ha una trasformata di Fourier $W(f)$.

Questa operazione definisce lo stimatore:

$$\tilde{S}_{xx}(f) = \int_{-T}^T \hat{R}_{xx}(\tau) \omega(\tau) e^{-j2\pi f\tau} d\tau$$

che è la convoluzione della stima bruta di densità spettrale con $W(f)$, cioè $\hat{S}_{xx}(f) * W(f)$. Questa procedura di stima può essere vista come una operazione di regolarizzazione nel dominio della frequenza, o, nel dominio del tempo, si può riguardare la finestra di ritardo come una funzione capace di ridurre l'importanza dei valori \hat{R}_{xx} quando τ aumenta. Ciò è utile in quanto le oscillazioni di \hat{R}_{xx} aumentano quando τ aumenta.

Proprietà statistiche: Prima di discutere questo aspetto è necessario introdurre alcuni concetti statistici.

Distribuzione del χ^2

Se $\chi_1, \chi_2, \dots, \chi_n$ sono n variabili random indipendenti, ciascuna delle quali ha una distribuzione normale con media nulla e deviazione standard unitaria, e se:

$$\chi_n^2 = \chi_1^2 + \chi_2^2 + \dots + \chi_n^2$$

allora la distribuzione di χ^2 è chiamata distribuzione del chi quadro con "n gradi di libertà" (essendo i gradi di libertà il numero di variabili indipendenti che entrano nell'espressione). I primi due momenti del χ^2 sono:

$$E[\chi_n^2] = n \qquad \sigma[\chi_n^2] = 2n$$

Se si sommano insieme indipendenti di *chi*², cioè:

$$\chi_v^2 = \chi_{n_1}^2 + \chi_{n_2}^2 + \dots + \chi_{n_k}^2$$

i gradi di libertà complessivi sono $v = n_1 + n_2 + \dots + n_k$. Queste espressioni vengono utilizzate per descrivere le proprietà statistiche della densità spettrale di potenza.

Bias: La stima bruta della densità spettrale di potenza è asintoticamente non distorta.

Varianza:

Poiché $\hat{S}_{xx}(f)$ è la convoluzione di due funzioni, esso risulta distorto. Infatti, per T grande, risulta:

$$E[\tilde{S}_{xx}(f)] = \int_{-\infty}^{\infty} S_{xx}(g)W(f-g)dg$$

e pertanto: $bias[\tilde{S}_{xx}(f)] = \int_{-\infty}^{+\infty} [\omega(\tau) - 1]R_{xx} e^{-j2\pi f\tau} d\tau$

La prima espressione mostra come la media della stima risulti distorta dal fenomeno di leakage (convoluzione con la finestra che presenta lobi laterali). Chiaramente i picchi risultano sottostimati (Parseval) e le valli sovrastimate. Poiché \tilde{S}_{xx} è una somma pesata di valori \hat{S}_{xx} , si può dimostrare

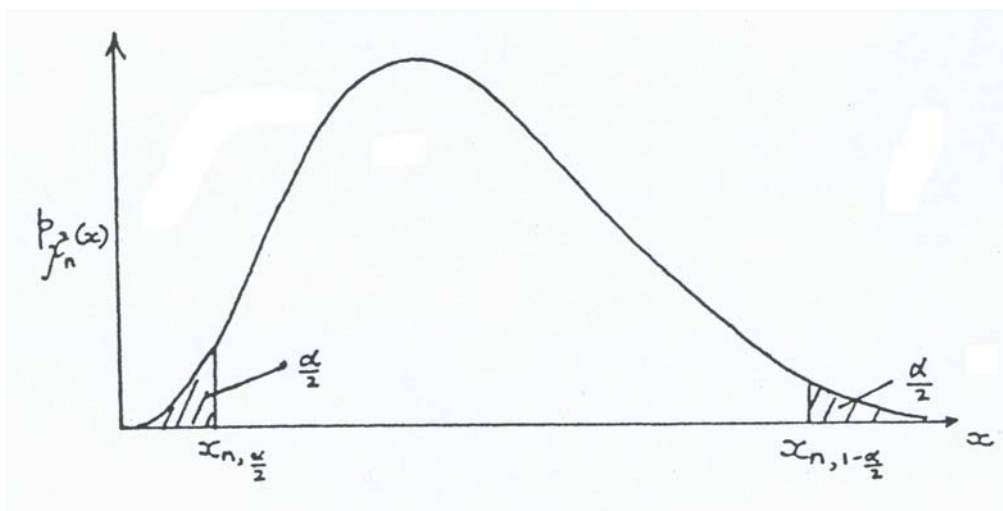
che l'espressione $\frac{n\tilde{S}_{xx}(f)}{S_{xx}(f)}$ è distribuita come χ_n^2 , essendo n (numero dei gradi di libertà) pari a 2BT. B è la banda di risoluzione e T il tempo di registrazione del segnale. Segue che:

$$\sigma^2[\tilde{S}_{xx}(f)] = \frac{S_{xx}(f)}{n/2} = \frac{S_{xx}(f)}{BT}$$

La scelta della lunghezza della finestra temporale è funzione dell'interesse che si dà alla stabilità (piccola varianza) o a un piccolo valore di bias. Se il bias è piccolo per tutti i valori di f si ha un'altra fedeltà. In generale ci si deve preoccupare di entrambi i fattori. Per esempio, se la funzione spettrale ha picchi stretti, è conveniente accettare qualche perdita di stabilità per ottenere una risoluzione accurata dei picchi, mentre se la funzione spettrale è piuttosto liscia, allora gli errori di bias non sono molto importanti.

Intervalli di confidenza:

Verranno ora introdotti intervalli di stima basati su stime puntuali, per una funzione spettrale piuttosto liscia. Poiché $\frac{n\tilde{S}_{xx}(f)}{S_{xx}(f)}$ è distribuita come una variabile random χ_n^2 , allora la densità di probabilità ha l'andamento mostrato nella seguente figura:



Se si sceglie un numero α con $0 < \alpha < 1$, in modo tale che le sezioni di area $\alpha/2$ si abbiano esternamente ai punti $x_{n,\alpha/2}$ e $x_{n,1-\alpha/2}$ possono essere ottenuti dalle tavole del χ_n^2 per valori differenti di α , e la disuguaglianza può essere risolta in funzione del valore S_{xx} che si vuole ottenere, nel seguente modo:

$$\frac{n\tilde{S}_{xx}(f)}{x_{n,1-\alpha/2}} \leq S_{xx}(f) \leq \frac{n\tilde{S}_{xx}(f)}{x_{n,\alpha/2}}$$

Pertanto, per una stima puntuale particolare \tilde{S}_{xx} , i limiti di confidenza 100(1- α)% per S_{xx} sono:

$$\frac{n}{x_{n,1-\alpha/2}} \tilde{S}_{xx} \quad \frac{n}{x_{n,\alpha/2}} \tilde{S}_{xx}$$

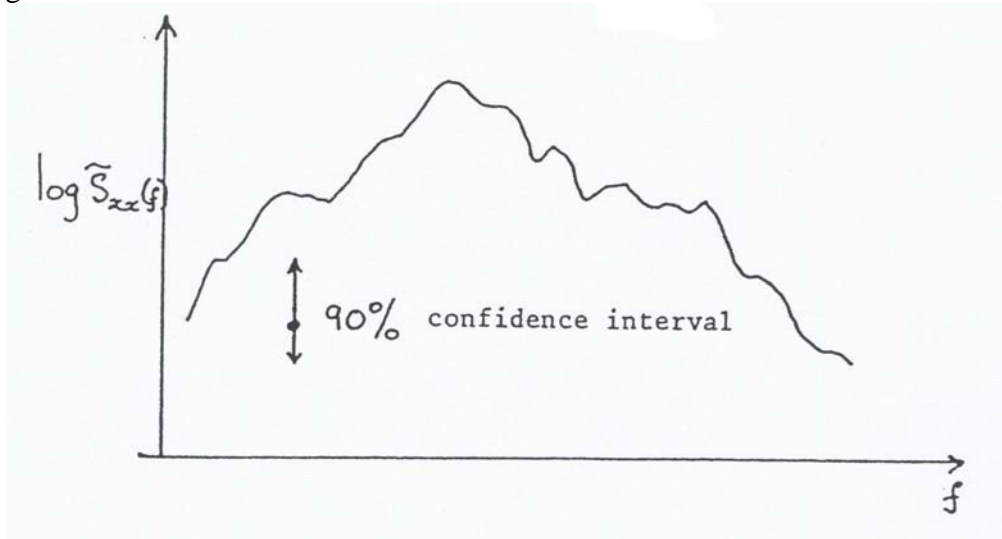
L'intervallo di confidenza è la differenza tra di essi.

Si noti che su una scala lineare l'intervallo di confidenza dipende dalla stima, ma su una scala logaritmica l'intervallo è:

$$\log\left(\frac{n}{x_{n,\alpha/2}}\right) - \log\left(\frac{n}{x_{n,1-\alpha/2}}\right)$$

che è un valore costante indipendente da \tilde{S}_{xx} .

In un diagramma logaritmico si possono attaccare bande di confidenza come mostrato nella seguente figura:



Calcolo:

Il calcolo di $\tilde{S}_{xx}(f)$ per dati campionati si ottiene partendo dalla seguente forma discreta, cioè:

$$\tilde{S}_{xx}(f) = \Delta \sum_{r=-(N-1)}^{N-1} \hat{R}_{xx}(r\Delta) \omega(r\Delta) e^{-j2\pi f r \Delta}$$

dove $T = N\Delta$. \hat{R}_{xx} è calcolata come mostrato nel paragrafo precedente.

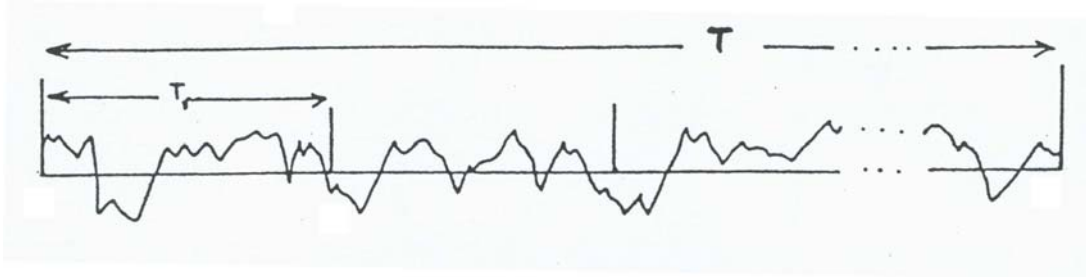
E' ora necessario decidere per quali valori di frequenza deve essere stimata \tilde{S}_{xx} . A causa della correlazione che esiste tra stime di \tilde{S}_{xx} per frequenze vicine, correlazione che dipende dalla finestra spettrale, è stato suggerito in letteratura che debbano essere calcolate solo stime incorrelate. Purtroppo le caratteristiche della densità spettrale andrebbero perse se si procedesse in questo modo e fosse necessaria una risoluzione in frequenza molto fine. Se, nella relazione precedente, la finestra fosse di dimensione $T = M\Delta$, allora:

$$\tilde{S}_{xx}(f) = \Delta \sum_{r=-(M-1)}^{M-1} \hat{R}_{xx}(r\Delta) \omega(r\Delta) e^{-j2\pi f r \Delta}$$

e, calcolando la f in $k/M\Delta$ (k intero), risulta una DFT. Per ottenere una risoluzione in frequenza più fine, Jenkins and Watt hanno proposto alcuni metodi ad hoc, come quello di "offuscare" la

correlazione, aggiungendo zeri prima della trasformazione. Un metodo alternativo ma equivalente all'uso della finestra di ritardo è quello di riconoscere che la trasformata di Fourier del prodotto di \hat{R}_{xx} e $\omega(\tau)$ è la convoluzione di \hat{S}_{xx} e $W(f)$. Pertanto la procedura usata è quella di calcolare \hat{S}_{xx} trasformando \hat{R}_{xx} e poi "lisciando" questa trasformata in frequenza mediante convoluzione.

Il metodo, comunque, che è diventato più popolare per il calcolo della densità spettrale (soprattutto a causa della velocità di calcolo), è quello delle medie segmentate. Il procedimento di base è spiegato con riferimento alla seguente figura:



Una registrazione di lunghezza T è divisa in q segmenti di lunghezza T_r (con o senza sovrapposizione dei segmenti) e viene calcolata la densità spettrale bruta di ogni segmento, cioè:

$$\hat{S}_{xx_i}(f) = \frac{1}{T_r} |X_{T_r_i}(f)|^2 \quad \text{per } i=1,2,\dots,q$$

Per ridurre le fluttuazioni viene quindi effettuata la media:

$$\tilde{S}_{xx}(f) = \frac{1}{q} \sum_{i=1}^q \hat{S}_{xx_i}(f)$$

Poiché ogni stima bruta è una variabile random con distribuzione χ^2_2 , allora la quantità $2\tilde{S}_{xx}q/S_{xx}$ è una distribuzione χ^2_{2q} . Segue allora che:

$$\sigma^2[\tilde{S}_{xx}] = \frac{S_{xx}^2}{q}$$

e, poiché la banda di risoluzione è circa $1/T_r$, si ha:

$$\sigma^2[\tilde{S}_{xx}] = \frac{S_{xx}^2}{BT}$$

I segmenti si possono o meno sovrapporre parzialmente uno con l'altro. L'uso di dati finestrati in segmenti, prima dell'effettuazione della trasformata di Fourier, riduce il leakage ma ha lo svantaggio di ridurre anche l'energia totale di ogni segmento (a seconda del tipo di finestra che si utilizza): ciò deve essere pertanto compensato. Si calcola quindi una stima bruta "modificata" per ogni segmento nel seguente modo:

$$\hat{S}_{xx}(f) = \frac{\frac{1}{T_r} \left| \int_{\text{interv. } i\text{-esimo}} x(t)\omega(t)e^{-j2\pi ft} dt \right|^2}{\frac{1}{T_r} \int_{-T_r/2}^{T_r/2} \omega^2(t) dt}$$

Dove il denominatore compensa la riduzione di energia.

Come ad esempio, nel caso di sovrapposizione del 50%, risulta:

$$\sigma^2[\tilde{S}_{xx}(f)] \cong 1.2 \frac{S_{xx}^2(f)}{2q}$$

7.18 Bibliografia

Sestieri A., *Dispense del corso di Meccanica delle Vibrazioni*, Università "La Sapienza" Roma, 1990.

Hammond J.K., *Signal Processing and analysis*, ISVR, Southampton (appunti del corso), 1989.

Brigham E.O., *The Fast Fourier Transform and its Applications*, Prentice-Hall, 1988.

Kolmogorov A. and S. Fomin, *Introductory Real Analysis*, Dover, New York, 1975

Bendat J.S., Persol A.G., *Random Data - Analysis and Measurement Procedures*, 2nd Edition, John Wiley & Sons, New York, 1986.

Oppenheim A.V. Willsky A.S. *Signals & Systems*, Prentice Hall, 1997