

Appunti di Teoria dei Segnali

Capitolo 8 - Introduzione alla probabilità

Concetti preliminari di probabilità	2
Introduzione alla probabilità.....	2
<i>Definizione di spazio degli eventi</i>	2
<i>Definizione di evento</i>	3
Esempio	3
<i>Metodi di combinazione degli eventi</i>	4
La funzione probabilità.....	4
Definizione.....	4
Proprietà fondamentali della funzione probabilità	5
Le probabilità condizionate	5
<i>Esempio</i>	6
<i>Osservazione: probabilità condizionate per eventi disgiunti</i>	6
Eventi indipendenti	7
<i>Esempio: dipendenza causale tra due eventi</i>	7
Partizione di uno spazio di eventi S.....	8
Teorema sulle probabilità totali	9
<i>Esempio</i>	10
Teorema di Bayes	10
Processo di Bernoulli	11
Definizione.....	11
Sistema di trasmissione binaria.....	13
Introduzione: canale binario simmetrico.....	13
Metodi di prevenzione e correzione degli errori	15
<i>Codice a ripetizione</i>	15
<i>Codice a controllo di parità</i>	18
Modello uniforme di probabilità in spazi campione finiti.....	22
Esempio	22
Variabili aleatorie	23
Definizione.....	23
<i>Variabile aleatoria discreta</i>	24
Continuità di una variabile aleatoria e funzione densità.....	25
<i>Il significato matematico della densità</i>	25
<i>Osservazione: peso nullo dei singoli punti</i>	26
Densità di una variabile aleatoria continua distribuita uniformemente su un intervallo	27
Distribuzione cumulativa di una variabile aleatoria continua.....	27
<i>Esempio: variabile aleatoria continua</i>	29
<i>Esempio: distribuzione uniforme su un intervallo</i>	30
<i>Esempio: distribuzione esponenziale</i>	30
Funzioni di variabili aleatorie.....	32
<i>Esempio: distribuzione cumulativa di X^2 con X generica</i>	32
<i>Esempio: distribuzione cumulativa di X^2 con X distribuita uniformemente</i>	34
Teorema: calcolo di F_Y nell'ipotesi di invertibilità di $Y=H(X)$	34

Esempio: $H(X)=3X+1$	36
Aspettazione di una variabile aleatoria.....	37
Definizione.....	37
Proprietà della media di una variabile aleatoria	37
<i>Teorema</i>	38
Varianza e deviazione standard di una variabile aleatoria	38
Definizione.....	38
<i>Proprietà della varianza di una variabile aleatoria</i>	38
Cenni sulle variabili aleatorie bidimensionali	39
Funzione caratteristica di una variabile aleatoria.....	41
Definizione nel caso continuo	41
Applicazione	42
Proprietà: momento del k° ordine	42
Esempio: distribuzione esponenziale	43
Funzione caratteristica di una variabile aleatoria discreta	44
Densità di probabilità condizionata	45
Variabili aleatorie con distribuzione gaussiana	47
Definizione di distribuzione gaussiana	47
Funzione caratteristica.....	47
Funzione densità di probabilità congiunta	48
Indipendenza di due variabili aleatorie con distribuzione gaussiana	49
Funzione densità di probabilità condizionata.....	49
Combinazione lineare di variabili aleatorie con distribuzione gaussiana... 49	
<i>Conseguenza</i>	50
Esempio	50

Concetti preliminari di probabilità

INTRODUZIONE ALLA PROBABILITÀ

E' esperienza comune che vi sono molti fenomeni empirici la cui osservazione, anche se effettuata nelle stesse condizioni, porta a risultati differenti: per esempio, quando si osserva il lancio di un dado, si nota che può presentarsi, in modo del tutto imprevedibile, una qualunque delle 6 facce.

*Siamo allora indotti a ricercare un modello matematico adatto allo studio di quei fenomeni che non manifestano una regolarità deterministica, in quanto la loro osservazione, relativamente ad un dato insieme di circostanze, non porta sempre allo stesso risultato: il modello che cerchiamo è il cosiddetto **modello probabilistico**.*

Definizione di spazio degli eventi

Supponiamo di osservare un certo *fenomeno* quale può essere il lancio di una moneta, il lancio di un dado, l'estrazione di una pallina da un'urna e cose di questo tipo. Definiamo **spazio degli eventi**

o anche **spazio dei campioni**, indicandolo con **S**, l'insieme di tutti i possibili risultati (o, meglio, i possibili **eventi**) osservati per il dato fenomeno.

Per esempio, considerato il fenomeno del lancio di una moneta, lo spazio dei campioni non è altro che $S = \{\text{testa, croce}\}$, in quanto gli unici possibili risultati del lancio possono essere che esca testa o che esca croce.

In modo del tutto analogo, dato il fenomeno del lancio di un classico dado a 6 facce, lo spazio degli eventi sarà $S = \{1,2,3,4,5,6\}$.

Quando si ha a che fare con lo spazio degli eventi relativi ad un dato fenomeno, è molto importante conoscere il numero degli eventi che appartengono a tale spazio; trattandosi di un insieme, esso può essere di tre tipi diversi:

- può essere *finito*, per cui noi conosciamo esattamente il numero degli eventi che gli appartengono;
- può essere *infinito ma numerabile*, cioè....
- può essere *infinito e non numerabile*, ossia

Consideriamo ad esempio il fenomeno consistente nell'estrarre una carta da un mazzo di 52 carte. Lo spazio campione è costituito da 52 punti, tanti quanti sono i VALORI che la carta estratta può assumere; siamo perciò in presenza di uno spazio campione finito.

Al contrario, se il fenomeno consiste nella scelta di un numero intero maggiore di 100, lo spazio campione è $\{101,102,103,\dots\}$ ed è perciò infinito numerabile.

Infine, se la scelta del numero può essere fatta nel campo dei numerici reali, allora lo spazio campione è infinito non numerabile.

Definizione di evento

C'è da chiarire bene cosa intendiamo per **evento**. Dato un certo spazio di eventi S relativo ad un dato fenomeno, un evento A è semplicemente un insieme di possibili risultati osservati per il fenomeno: si tratta perciò di un sottoinsieme di S . Chiaramente, potrà essere di 2 tipi: potrà includere più risultati oppure potrà consistere in un risultato solo.

Per esempio, se noi consideriamo il lancio di un dado, noi possiamo indicare come evento A l'evento per cui il risultato del lancio è 1, ma possiamo indicare come evento B quello per cui il risultato del lancio è un numero pari: possiamo perciò scrivere che

$$S = \{1,2,3,4,5,6\}$$

$$A = \{1\}$$

$$B = \{2,4,6\}$$

Esempio

Supponiamo di scegliere a caso 10 persone tra gli abitanti di una città, di domandare loro se usano un certo prodotto e di contare il numero di *sì* dati come risposta. Anche questo è un esperimento casuale, i cui possibili risultati sono 11: nessun *sì*, 1 *sì*, 2 *sì* e così via. Indichiamo perciò lo spazio degli eventi con $S = \{0,1,2,\dots,10\}$.

Se consideriamo l'evento A corrispondente a dire che le risposte affermative sono più di otto, è chiaro che questo evento si verifica quando 9 o 10 persone hanno risposto *sì*, per cui $A = \{9,10\}$.

Metodi di combinazione degli eventi

Supponiamo di osservare un dato fenomeno (o di effettuare un certo esperimento) per il quale S sia lo spazio degli eventi; consideriamo 2 qualsiasi eventi A e B relativi a tale fenomeno; questi 2 eventi possono essere combinati in vari modi per formare ulteriori eventi:

- l'evento $C=A\vee B$ (unione) è l'evento che accade quando almeno uno dei due eventi accade: se accade solo A o solo B , noi potremo dire che è accaduto anche C ;
- l'evento $D=A\wedge B$ (intersezione) è l'evento che accade solo quando entrambi gli eventi sono accaduti; se uno solo di essi non si verifica, diremo che non si è verificato neanche D ;
- l'evento A^* è l'evento opposto di A , ossia si ritiene verificato quando A non si è verificato;

Facciamo anche qui un esempio riferito al lancio di un dado a 6 facce, il cui spazio dei campioni è $S = \{1,2,3,4,5,6\}$. Consideriamo i seguenti eventi:

$A =$ esce un numero pari

$B =$ esce un numero minore di 4

E' chiaro allora che l'evento $C=A\vee B$ racchiude i seguenti risultati:

$$C = A \cup B = \{1,2,3,4,6\}$$

Al contrario, invece, l'evento $D=A\wedge B$ è

$$D = A \cap B = \{2\}$$

La funzione probabilità

DEFINIZIONE

Consideriamo un generico spazio degli eventi S : come sappiamo, si tratta, dato un certo fenomeno, di tutti i possibili eventi, relativi a tale fenomeno, che si possono verificare. Noi allora definiamo **funzione probabilità** la funzione reale così definita:

$$P : S \longrightarrow \mathbb{R}$$

Essa cioè, dato un qualsiasi evento preso dallo spazio S , vi associa un valore reale che indica la probabilità che tale evento si verifichi.

Inoltre, noi diremo che la coppia (S,P) costituita dallo spazio degli eventi S e dalla funzione probabilità P in esso definita è uno **spazio di probabilità**.

PROPRIETÀ FONDAMENTALI DELLA FUNZIONE PROBABILITÀ

Vediamo alcune proprietà fondamentali di cui gode una funzione probabilità P :

- 1) $P(\emptyset)=0$: la probabilità che accada l' **evento nullo** è ovviamente nulla;
- 2) $P(S)=1$: la probabilità che accada uno qualsiasi degli eventi contenuti in S è ovviamente pari al massimo, cioè all'unità, in quanto S contiene tutti gli eventi possibili e quindi almeno uno si deve verificare;
- 3) $0 \leq P(A) \leq 1$: la probabilità che accada un certo evento A è chiaramente positiva e minore di 1;
- 4) se A e B sono **disgiunti**, o anche **mutuamente esclusivi**, cioè non hanno risultati in comune $\otimes P(A \cup B) = P(A) + P(B)$: se A e B non hanno alcun evento in comune (ossia se, accaduto A , non si possa dire assolutamente che è accaduto anche B e viceversa), la probabilità che accada prima uno e poi l'altro, o viceversa, è ovviamente pari alla somma delle probabilità che ciascuno accada singolarmente¹;
- 5) se $A \subseteq B \otimes P(A) \leq P(B)$: dire che A è contenuto in B significa dire che B contiene gli stessi eventi di A ed anche degli altri, il che implica che la probabilità che si verifichi A è al più uguale a quella che si verifichi B ;
- 6) Se $A \subseteq B \otimes P(B-A) = P(B) - P(A)$: l'evento complessivo $B-A$ non è altro che l'insieme degli eventi di B che non fanno parte di A ;
- 7) Posto $A^* = S-A$, si ha che $P(A^*) = P(S-A) = P(S) - P(A) = 1 - P(A)$: l'evento A^* non è altro che l'insieme degli eventi di S che non appartengono ad A , per cui questa proprietà scaturisce direttamente dalla precedente;

LE PROBABILITÀ CONDIZIONATE

Sia dato lo spazio di probabilità (S,P) . Supponiamo che ci sia un certo evento $B \subseteq S$ che abbia probabilità non nulla di accadere, ossia tale che $P(B) \neq 0$. Supponiamo che effettivamente l'evento B si verifichi. Prendiamo poi un altro evento $A \subseteq S$. Allora, noi diremo che la probabilità che si verifichi A , una volta che si è verificato B , è data dalla formula

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

dove $P(A|B)$ si legge appunto **probabilità di A dato B** .

Possiamo fare qualche calcolo a partire dalla relazione appena ricavata e sfruttando le proprietà fondamentali di una funzione probabilità: infatti, da lì si ricava che

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

¹ Un'altra definizione di *eventi disgiunti* è la seguente: due eventi A e B sono disgiunti se non possono presentarsi insieme, ossia se l'evento intersezione è l'evento impossibile.

Dato che $P(A \cap B) = P(B \cap A)$, possiamo anche scrivere che

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

per cui, eguagliando i secondi membri, noi otteniamo

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

Esempio

Supponiamo di avere un ufficio con 100 macchine calcolatrici. Alcune di esse sono elettriche (E) ed altre manuali (M); inoltre, alcune di esse sono nuove (N) ed altre vecchie (U).

Supponiamo che queste siano le esatte quantità:

- 40 macchine elettriche nuove;
- 20 macchine elettriche usate;
- 30 macchine manuali nuove;
- 10 macchine manuali usate.

Una persona entra nell'ufficio e sceglie una macchina a caso. Se contrassegniamo ciascuna macchina con un numero (che andrà da 1 a 100), è ovvio che la probabilità che la persona scelga la k° macchina è pari a $1/100$ ed è ovviamente costante per ciascuna macchina.

Supponiamo che la persona, dopo aver scelto la macchina, si accorga che essa sia nuova. Vogliamo la probabilità che sia anche elettrica. Possiamo esprimerci dicendo che vogliamo $P(E|N)$.

Il fatto che sia accaduto N restringe le possibilità a 70 macchine; su queste 70, ce ne sono 40 elettriche, per cui

$$P(E|N) = \frac{40}{70}$$

Un secondo metodo è l'applicazione della formula:

$$P(E|N) = \frac{P(E \cap N)}{P(N)}$$

L'evento $E \cap N$ è quello per cui la macchina scelta è elettrica e nuova e ce ne sono in tutto 40; l'evento N è quello per cui la macchina è nuova e ce ne sono 70, per cui il risultato è quello di prima.

Osservazione: probabilità condizionate per eventi disgiunti

Facciamo notare che, se A e B sono 2 eventi disgiunti, cioè tali che $A \cap B = \emptyset$, è ovvio che

$$P(A|B) = P(B|A) = 0$$

Il motivo è sia intuitivo sia matematico: da un punto di vista intuitivo, il fatto che siano disgiunti comporta che il verificarsi di B (o di A) escluda la possibilità che si verifichi A (o B); da un punto di vista matematico, invece, si ha evidentemente che

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(\emptyset)}{P(B)} = 0$$

N.B. Facciamo anche osservare l'estremo opposto: infatti, se $B \supset A$, allora $P(B|A)=1$.

EVENTI INDIPENDENTI

Dalla relazione $P(A|B) = \frac{P(A \cap B)}{P(B)}$ si ricava evidentemente che $P(A|B)$ può essere maggiore, minore o uguale a $P(A)$. Allora, se accade che sono uguali, significa una cosa molto importante: il fatto che B si sia verificato non influisce in alcun modo sulla probabilità che subito dopo si verifichi A. In questo caso, si dice che i 2 eventi A e B sono **indipendenti** tra loro.

Una importante conseguenza della indipendenza di 2 eventi è dunque che

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

per cui noi possiamo dire che *2 eventi A e B sono indipendenti se verificano la relazione appena scritta*.

Nel caso gli eventi siano 3 (A, B e C) non cambia molto: noi diremo che sono indipendenti se sono a 2 a 2 indipendenti se verificano la relazione

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

Esempio: dipendenza causale tra due eventi

Consideriamo un'urna con 5 palline rosse e 4 bianche: supponendo di fare 2 estrazioni con restituzione, abbiamo 4 possibilità:

- 1) rossa, rossa
- 2) rossa, bianca
- 3) bianca, rossa
- 4) bianca, bianca

Consideriamo l'evento A per cui esce prima una pallina rossa e poi una bianca. Vogliamo calcolare $P(A)$, cioè la probabilità che tale evento si verifichi.

Per prima cosa, possiamo scrivere formalmente che

$$P(A) = P((R1, B2)) = P(R1 \wedge B2)$$

dove con $R1$ indichiamo l'evento per cui la prima pallina estratta sia rossa e con $B2$ l'evento per cui la seconda pallina sia bianca. Avendo supposto di fare la restituzione dopo la prima estrazione, è lecito

aspettarsi che l'uscita della pallina rossa e successivamente di quella bianca siano due eventi indipendenti, per cui possiamo applicare la relazione che caratterizza due eventi indipendenti per il calcolo di $P(A)$:

$$P(A) = P(R1 \wedge B2) = P(R1)P(B2)$$

A questo punto, dato che $P(R1)=5/9$ e $P(B2)=4/9$, abbiamo che

$$P(A) = 20/81$$

Vediamo adesso cosa cambia se non facessimo la restituzione dopo la prima estrazione. A prescindere dalla non restituzione, noi possiamo sempre scrivere che

$$P(A) = P(R1 \wedge B2)$$

Il problema sorge dal fatto che i 2 eventi non sono più indipendenti adesso, per cui non possiamo applicare la relazione caratteristica degli eventi indipendenti. Abbiamo allora 2 possibilità per procedere:

- la prima è di considerare la probabilità che si verifichi $B2$ dato $R1$, ossia

$$P(A) = P(B2 | R1) * P(R1)$$

- la seconda è di considerare la probabilità che si verifichi $R1$ dato $B2$, ossia

$$P(A) = P(R1 | B2) * P(B2)$$

Dovendo scegliere, conviene la prima strada, in quanto è più facile valutare $P(R1)$ e poi $P(B2 | R1)$ che non valutare prima $P(B2)$ e poi $P(R1 | B2)$. Infatti

$$\begin{aligned} P(R1) &= 5/9 \\ P(B2 | R1) &= 4/8 = 1/2 \end{aligned}$$

per cui

$$P(A) = (5/9) * (1/2) = 5/18$$

Prima di passare avanti, ricordiamo che in una situazione in cui i 2 eventi non sono indipendenti, ossia in una situazione in cui il primo evento condiziona necessariamente il secondo, si dice che c'è una **dipendenza causale** tra i due eventi (o anche una **dipendenza causa-effetto**).

PARTIZIONE DI UNO SPAZIO DI EVENTI S

Consideriamo uno spazio di eventi S ; consideriamo in particolare gli eventi B_1, B_2, \dots, B_k ; noi diremo che essi costituiscono una **partizione di S** se godono delle seguenti 3 proprietà:

- sono a due a due disgiunti;
- la loro unione dà S ;
- la probabilità che accada ciascun di essi è maggiore di 0, ossia sono eventi effettivamente osservabili.

Facciamo un esempio evidente di partizione: considerato il fenomeno del lancio di un dado, il cui spazio degli eventi è $S=\{1,2,3,4,5,6\}$, una partizione è costituita dagli eventi

$$\begin{aligned} B_1 &= \{1,2\} \\ B_2 &= \{3,4,5\} \\ B_3 &= \{6\} \end{aligned}$$

mentre non è una partizione quella costituita dagli eventi

$$\begin{aligned} C_1 &= \{1,2,3,4\} \\ C_2 &= \{4,5,6\} \end{aligned}$$

TEOREMA SULLE PROBABILITÀ TOTALI

Supponiamo di avere uno spazio di probabilità (S,P) , ossia uno spazio degli eventi S ed una funzione reale P che associa ad ogni evento la probabilità che esso avvenga. Consideriamo una partizione di S , ossia, come visto poco fa, una successione $(B_n)_{n \in \mathbb{N}}$ di insiemi a 2 a 2 disgiunti e tali che la loro unione dia proprio S . Fissiamo un certo evento A . Possiamo evidentemente scrivere quanto segue:

$$A = (A \cap B_1) \cup (A \cap B_2) \cup \dots \cup (A \cap B_N)$$

Dato che gli eventi della partizione sono disgiunti per definizione, anche gli eventi tra parentesi sono disgiunti, per cui possiamo applicare il principio di addizione e scrivere che

$$P(A) = P(A \cap B_1) + P(A \cap B_2) + \dots + P(A \cap B_N)$$

per cui

$$P(A) = \sum_{i=1}^N P(A \cap B_i)$$

Il teorema dice cioè questo: *la probabilità che si verifichi l'evento A si può calcolare come somma delle probabilità che si verifichino gli eventi $A \cap B_n$, con $n \in \mathbb{N}$.*

Questo teorema dice anche un'altra cosa: infatti, in base a quanto abbiamo visto circa le probabilità condizionate, noi possiamo scrivere che

$$P(A \cap B_i) = P(A|B_i)P(B_i) \quad i = 1, 2, \dots, N$$

per cui la tesi diventa

$$P(A) = \sum_{i=1}^N P(A|B_i)P(B_i)$$

L'utilità di questa relazione è evidente: dovendo valutare $P(A)$, spesso è difficile farlo direttamente; al contrario, noi possiamo calcolarla come somma di opportune probabilità condizionate assumendo che siano accaduti gli eventi di una opportuna partizione e le probabilità condizionate sono spesso facili da determinare.

Esempio

Supponiamo di avere un lotto di 100 componenti meccanici di cui solo 20 difettosi e gli altri 80 integri. Supponiamo di scegliere a caso 2 componenti (senza restituire il primo) e supponiamo di voler la probabilità che il secondo sia difettoso.

Possiamo porre

A = il primo componente è difettoso

B = il secondo componente è difettoso

Possiamo allora scrivere che

$$P(B) = P(B|A)P(A) + P(B|A^*)P(A^*)$$

Infatti, abbiamo considerato come partizione quella composta dagli eventi A e A*. Sostituendo i rispettivi valori numerici abbiamo

$$P(B) = (19/99)*(20/100) + (20/99)*(80/100) = 1/5$$

TEOREMA DI BAYES

Questo teorema è una diretta conseguenza del teorema sulle probabilità totali. Ci mettiamo anche nelle stesse ipotesi del teorema precedente: abbiamo perciò lo spazio di probabilità (S,P) ed abbiamo una partizione $(B_n)_{n \in N}$ dello spazio degli eventi S. Consideriamo infine un evento generico A.

Vogliamo calcolare la probabilità che si verifichi un generico evento B_i della partizione una volta che si sia verificato A, ossia vogliamo calcolare $P(B_i|A)$. La prima cosa che possiamo fare è applicare la formula vista in precedenza per le probabilità condizionate: abbiamo perciò che

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

A questo punto, per il calcolo di P(A) noi possiamo servirci del teorema sulle probabilità totali, in particolare ci possiamo servire della 2° tesi. Abbiamo perciò il seguente risultato:

$$P(B_i|A) = \frac{P(A|B_i)P(B_i)}{\sum_{i=1}^N P(A|B_i)P(B_i)}$$

e questa è la tesi del teorema di Bayes.

Processo di Bernoulli

DEFINIZIONE

Sia dato un certo esperimento il cui spazio dei campioni S è costituito solo da due possibili risultati: $S = \{0,1\}$. Supponiamo di ripetere n volte questo esperimento e supponiamo anche che ogni ripetizione sia indipendente dalle altre, ossia non influenzi le altre e non venga influenzata dalle altre.

Vogliamo calcolare la probabilità che, sulle n prove, il risultato 1 venga fuori K volte. Se indichiamo con X una variabile che tiene conto di quante volte viene fuori il risultato 1 sulle n prove, è chiaro che noi vogliamo valutare il valore di $P(X = K)$.

Supponiamo che p sia la probabilità che, nella generica prova, venga fuori 1; ovviamente, allora, $1-p$ sarà la probabilità che, sempre nella generica prova, venga fuori 0.

Si può dimostrare analiticamente che la probabilità che noi cerchiamo è

$$P(X = K) = \binom{n}{k} p^k (1-p)^{n-k}$$

L'insieme delle n ripetizioni del nostro esperimento con tutte le proprietà elencate, inclusa $P(X=K)$, prende il nome di **processo di Bernoulli**.

Vediamo di esaminare un po' più nel dettaglio tale processo. Intanto, il fatto che lo spazio dei campioni abbia solo 2 elementi implica che, su n ripetizioni, i risultati possibili siano 2^n . Per esempio, se supponiamo di effettuare $n=3$ ripetizioni, i risultati possibili sono i seguenti:

- 1) 0 0 0
- 2) 1 0 0
- 3) 0 1 0
- 4) 0 0 1
- 5) 1 1 0
- 6) 0 1 1
- 7) 1 0 1
- 8) 1 1 1

Sulla base di questa tabella di risultati, diamo una giustificazione intuitiva della formula data prima per $P(X=K)$. Supponiamo ad esempio di volere $P(X=1)$, ossia la probabilità che, sulle 3 ripetizioni, venga fuori 1 una sola volta. E' intanto evidente che i casi a noi favorevoli sono solo 3 e precisamente i casi (2), (3) e (4): possiamo perciò cominciare a scrivere che

$$P(X = 1) = P(0,0,1 \cup 0,1,0 \cup 1,0,0)$$

Gli 8 eventi sono tra loro disgiunti (o mutuamente esclusivi), in quanto è ovvio che non possono accadere insieme: se esce la sequenza 111, certamente non potrà uscire nessun'altra delle altre sequenze. Di conseguenza, la probabilità che accada almeno uno dei 3 è la somma delle probabilità che ciascuno accada singolarmente: quindi

$$P(X = 1) = P(0,0,1) + P(0,1,0) + P(1,0,0)$$

Consideriamo $P(0,0,1)$: questa è la probabilità che alla prima e alla seconda ripetizione venga fuori 0 ed alla terza venga fuori 1. Avendo supposto all'inizio che ciascuna prova sia indipendente dalle altre, possiamo scrivere che

$$P(0,0,1) = P(1^{\circ}0)P(2^{\circ}0)P(3^{\circ}1)$$

ossia che la probabilità che i tre eventi accadano in sequenza è pari al prodotto delle singole proprietà, in base alla definizione di eventi indipendenti.

In modo analogo, per le altre due, ossia

$$P(0,1,0) = P(1^{\circ}0)P(2^{\circ}1)P(3^{\circ}0)$$

$$P(1,0,0) = P(1^{\circ}1)P(2^{\circ}0)P(3^{\circ}0)$$

Avendo infine detto che la probabilità che, alla generica prova, venga fuori 1 è pari a p e quella che venga fuori 0 è pari a $1-p$, possiamo scrivere che

$$P(0,0,1) = P(1^{\circ}0)P(2^{\circ}0)P(3^{\circ}1) = (1-p)(1-p)p = (1-p)^2 p$$

$$P(0,1,0) = P(1^{\circ}0)P(2^{\circ}1)P(3^{\circ}0) = (1-p)p(1-p) = (1-p)^2 p$$

$$P(1,0,0) = P(1^{\circ}1)P(2^{\circ}0)P(3^{\circ}0) = p(1-p)(1-p) = (1-p)^2 p$$

e quindi possiamo concludere che

$$P(X = 1) = P(0,0,1) + P(0,1,0) + P(1,0,0) = 3(1-p)^2 p$$

e questa si può anche scrivere nella forma

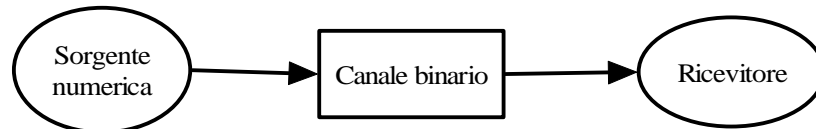
$$P(X = 1) = \binom{3}{1} (1-p)^{3-1} p^1$$

che è in perfetto accordo con la formula generale.

Sistema di trasmissione binaria

INTRODUZIONE: CANALE BINARIO SIMMETRICO

Supponiamo di dover trasmettere un segnale binario, ossia una sequenza di bit, da una certa sorgente ad un certo ricevitore. Possiamo schematizzare l'apparato di trasmissione nel modo seguente:



La *sorgente numerica* è quella che genera il segnale binario da trasmettere, ossia una sequenza di simboli 1 e 0; il **canale binario** rappresenta invece tutti quei dispositivi necessari per la trasmissione del segnale dalla sorgente al *ricevitore*.

I problemi, in uno schema di trasmissione di questo tipo, vengono proprio dal *canale binario*: infatti, trattandosi di un insieme di cavi e dispositivi fisici, è possibile che esso commetta degli errori nella trasmissione del segnale: in parole povere, è possibile che, nonostante la sorgente abbia emesso il simbolo 1, il ricevitore riceva il simbolo 0 e viceversa.

Vogliamo allora studiare, da un punto di vista probabilistico, la possibilità che le informazioni che arrivano al ricevitore siano diverse da quelle emesse dalla sorgente.

A tal fine, possiamo caratterizzare il canale binario mediante dei *concetti probabilistici*; in particolare, è evidente che il canale commette un errore in due casi:

- il primo caso è che, in corrispondenza della trasmissione di un 1, esso faccia ricevere uno 0;
- il secondo caso, invece, è ovviamente che, in corrispondenza della trasmissione di uno 0, esso faccia ricevere un 1.

Possiamo associare a questi due casi due eventi:

evento A : 1 ricevuto in corrispondenza di uno 0 trasmesso

evento B : 0 ricevuto in corrispondenza di un 1 trasmesso

Ancora più sinteticamente, possiamo usare il concetto di probabilità condizionate e scrivere che

evento A : $1R|0T$

evento B : $0R|1T$

Prendono allora il nome di **probabilità di transizione** la probabilità che si verifichi l'evento A e la probabilità che si verifichi l'evento B:

$$\text{probabilità di transizione : } \begin{cases} p_0 = P(1R|0T) \\ p_1 = P(0R|1T) \end{cases}$$

Chiaramente, quindi, p_0 indica la probabilità che sia ricevuto un 1 quando invece la sorgente ha emesso uno 0, mentre p_1 indica la probabilità che sia ricevuto uno 0 quando invece la sorgente ha emesso un 1.

Noi diremo che il canale binario considerato è **simmetrico** quando $p_0=p_1$: ciò significa dire che la probabilità di sbagliare simbolo è identica sia quando viene trasmesso uno 0 sia quando viene trasmesso un 1.

Mettiamoci dunque in questa ipotesi e poniamo semplicemente $p=p_0=p_1$: quindi p rappresenta la probabilità che il canale faccia giungere al ricevitore un simbolo diverso da quello che è stato emesso dalla sorgente.

Quindi, quando si verifica un errore? L'evento errore, che indichiamo con E , si verifica quando il bit ricevuto è diverso dal bit trasmesso. Vogliamo $P(E)$.

Per calcolare $P(E)$ intendiamo applicare il teorema delle probabilità totali, del quale ripetiamo perciò l'enunciato: sia dato lo spazio dei campioni S relativo ad un certo fenomeno; consideriamo una partizione di S , ossia una successione $(B_n)_{n \in N}$ di insiemi a 2 a 2 disgiunti e tali che la loro unione dia proprio S ; fissiamo un certo evento A ; la probabilità che A si verifichi può essere calcolata mediante la formula

$$P(A) = \sum_{i=1}^N P(A \cap B_i)$$

Vediamo come questo teorema ci è di aiuto nel nostro caso. Intanto, lo spazio degli eventi relativo al fenomeno della generazione, da parte della sorgente, del segnale binario è semplicemente $S = \{0T, 1T\}$, in quanto la sorgente può trasmettere o 0 o 1. Allora, come partizione di S noi possiamo prendere gli eventi

$$\begin{aligned} B_1 &= \{1T\} \\ B_2 &= \{0T\} \end{aligned}$$

In tal modo, possiamo valutare la probabilità che si verifichi un errore mediante la formula

$$P(E) = P(E \cap 0T) + P(E \cap 1T)$$

Usando adesso le probabilità condizionate, possiamo esprimere le probabilità assolute in termini appunto di probabilità condizionate, ottenendo che

$$P(E) = P(E|0T)P(0T) + P(E|1T)P(1T)$$

Adesso riflettiamo su quanto abbiamo scritto:

- $(E|T)$

parole, è la probabilità che sia ricevuto 1 avendo trasmesso 0, ossia è la prima probabilità di transizione $p = p_0 = P(1R|0T)$;

- in modo analogo, $P(E|1)$ è la probabilità che avvenga un errore nell'ipotesi di aver trasmesso uno 1, ossia la probabilità che sia ricevuto 0 avendo trasmesso 1, ossia è la seconda probabilità di transizione $p = p_1 = P(0R|1T)$;

Quindi abbiamo che

$$P(E) = pP(0T) + pP(1T) = p[P(0T) + P(1T)]$$

Inoltre, è chiaro che il termine tra parentesi quadre è la probabilità dell'evento $P(0T \vee 1T)$: infatti, l'evento $0T$ è disgiunto dall'evento $1T$ (in quanto la sorgente o trasmette 1 o trasmette 0), per cui la somma delle probabilità di due eventi disgiunti è pari alla probabilità dell'evento unione. Tuttavia, l'evento $0T \vee 1T$ è l'evento certo, in quanto la sorgente o trasmette 1 o trasmette 0, per cui $P(0T \vee 1T) = 1$.

Possiamo perciò concludere che, *per ogni simbolo emesso dalla sorgente, la probabilità che si verifichi un errore, ossia la probabilità che il ricevitore riceva il simbolo sbagliato, è pari proprio a p , cioè alla probabilità che sia il canale a sbagliare il simbolo.*

Questo, ovviamente, nell'ipotesi di canale simmetrico.

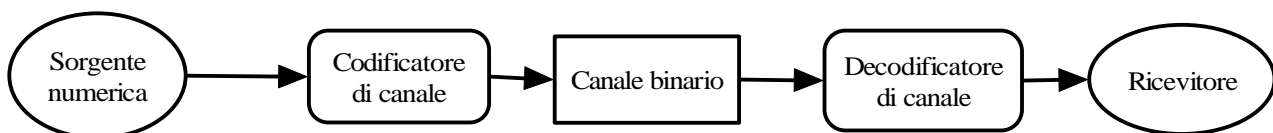
METODI DI PREVENZIONE E CORREZIONE DEGLI ERRORI

Codice a ripetizione

Per motivi fisici, la probabilità p che il canale sbaglia non sempre è trascurabile, per cui è opportuno ridurre al minimo la probabilità di errore: è cioè necessario trovare un metodo che riesca a recuperare gli errori del canale, ossia a far sì che, nonostante il canale commetta degli errori, il ricevitore riceva comunque il segnale corretto.

I metodi possibili sono diversi: il primo che consideriamo è il cosiddetto **codice a ripetizione**. La caratteristica fondamentale di questo metodo è quella per cui, dato il segnale² emesso dalla sorgente, esso, prima di essere inviato al canale per la trasmissione, viene arricchito di un numero prefissato di bit, posti ovviamente in posizione opportuna. Vediamo di spiegarci meglio.

Intanto, dobbiamo modificare lo schema del sistema di trasmissione nel modo seguente:



Sono stati dunque aggiunti due nuovi apparati, denominati **codificatore di canale** e **decodificatore di canale**. Vediamo quali sono le loro funzioni a partire dal primo.

Supponiamo che la sorgente emetta la seguente sequenza di simboli binari:

0 1 0 0 1 1 1 0 0 0

² cioè la sequenza di bit

Come si vede dallo schema, questa sequenza, prima di arrivare al canale per la trasmissione, va in input al **codificatore di canale**: il compito di questo organo è quello di modificare il segnale binario nel modo seguente: per ogni simbolo binario in ingresso, il codificatore ne genera un numero DISPARI di copie.

Nel nostro caso, in corrispondenza di quel segnale, il segnale generato dal canale, nell'ipotesi di 3 copie ogni bit, sarà il seguente:

000 111 000 000 111 111 111 000 000 000

Questo è il segnale che va nel canale e da questo deve essere trasmesso verso il ricevitore e questo è anche il motivo per cui si parla di **codice a ripetizione**.

Nell'ipotesi che il canale non commetta nessun errore durante la trasmissione, quello è anche il segnale che arriva in ingresso al **decodificatore di canale**: il compito di questo organo è quello di ricostruire, a partire da questo segnale in ingresso, il segnale originale emesso dalla sorgente e di inviarlo quindi al ricevitore.

Come avviene la ricostruzione? Si dice che il decodificatore *agisce a maggioranza*: per ogni terna di simboli che arriva in ingresso, il decodificatore trasmette al ricevitore il simbolo che, tra i 3, si ripete più volte. Nel caso di 3 ripetizioni, è ovvio che trasmetterà il simbolo che si ripete almeno 2 volte; nel caso generale di $2n+1$ ripetizioni, il decodificatore trasmette il simbolo che si ripete almeno $n+1$ volte.

Vediamo allora per quale motivo questo metodo di trasmissione consente di ridurre la probabilità di errore, ossia la probabilità che il ricevitore riceva un simbolo diverso da quello effettivamente inviato dalla sorgente.

Sappiamo che il problema deriva dal canale, il quale, in modo del tutto casuale, può modificare qualcuno dei bit che riceve in ingresso dal codificatore. Per esempio, supponiamo che, anziché trasmettere il segnale prima indicato, il canale trasmetta quest'altro:

0 1 0 1 0 0 0 0 0 0 0 0 1 1 0 1 1 1 0 1 1 0 0 0 0 0 0 0 0 0
 ↑ ↑ ↑ ↑ ↑ ↑ ↑ ↑ ↑ ↑

Le frecce indicano i bit in corrispondenza dei quali si è verificato un errore rispetto al segnale corretto. Vediamo come il decodificatore ricostruisce il segnale da inviare al ricevitore:

$\underbrace{010}_0$ $\underbrace{100}_0$ $\underbrace{000}_0$ $\underbrace{000}_0$ $\underbrace{110}_1$ $\underbrace{111}_1$ $\underbrace{011}_1$ $\underbrace{000}_0$ $\underbrace{000}_0$ $\underbrace{000}_0$

Confrontando questo con il segnale emesso dalla sorgente, osserviamo che si è verificato 1 solo errore (nella seconda terna), a fronte di ben 5 errori commessi da parte del canale. Quindi, dei 5 errori, ben 4 sono stati recuperati.

Anche se tra un attimo lo vedremo meglio dal punto di vista matematico, è intuitivo accorgersi che la probabilità che il ricevitore riceva un simbolo sbagliato diventa senz'altro piccola: infatti, è chiaro che questa eventualità si verifica solo se il canale commette almeno 2 errori in 1 sola terna, il che è abbastanza difficile. Non solo, ma vedremo anche che la probabilità di errore si riduce di parecchio aumentando il numero di ripetizioni, ossia associando, a ciascun bit emesso dalla sorgente, un numero (dispari) ancora maggiore di copie.

Vediamo allora quali sono le osservazioni che possiamo fare da un punto di vista probabilistico. Intanto, facciamo due ipotesi di partenza:

- la prima è che il canale sia **simmetrico**: abbiamo visto prima che questo significa che la probabilità di sbagliare simbolo è identica sia quando viene trasmesso uno 0 sia quando viene

trasmesso un 1 ed abbiamo anche visto quali conseguenze ci siano per la valutazione della probabilità di errore;

- la seconda ipotesi è che un eventuale errore, da parte del canale, su un bit non sia influenzata da altri errori e non influenzi altri errori; detto in altre parole, gli eventuali errori sono tutti eventi indipendenti tra loro.

Consideriamo il seguente evento:

$$E_s : \text{errore di sistema}$$

Si tratta cioè dell'evento per cui il ricevitore riceve un bit diverso da quello inviato dalla sorgente. Noi vogliamo calcolare $P(E_s)$, con l'accortezza di non confondere questa probabilità con p , che è invece la probabilità di errore del canale sul generico bit.

Mettiamoci nel caso generale in cui il codificatore di canale, per ciascun bit, ne invia al canale $2n+1$ copie. L'errore si verifica quando, su questi $2n+1$ bit da trasmettere, il canale compie almeno $n+1$ errori (e al più $2n+1$): quindi, possiamo scrivere che

$$P(E_s) = P(n+1 \text{ bit errati} \vee n+2 \text{ bit errati} \vee \dots \vee 2n+1 \text{ bit errati} \vee \text{altri bit corretti})$$

E' subito ovvio che l'ultimo evento si può eliminare per definizione di unione di due o più eventi: quindi

$$P(E_s) = P(n+1 \text{ bit errati} \vee n+2 \text{ bit errati} \vee \dots \vee 2n+1 \text{ bit errati})$$

Gli eventi rimanenti sono certamente disgiunti tra loro, in quanto non si possono verificare contemporaneamente (o si verificano $n+1$ bit errati, o se ne verificano $n+2$ e così via), per cui possiamo scrivere che

$$P(E_s) = P(n+1 \text{ bit errati}) + P(n+2 \text{ bit errati}) + \dots + P(2n+1 \text{ bit errati})$$

Concentriamoci su queste probabilità: dato k generico (compreso tra $n+1$ e $2n+1$), $P(k \text{ bit errati})$ è la probabilità che, su $2n+1$ bit, il canale commetta k errori; essendo ciascun errore indipendente dagli altri ed essendo p la probabilità di errore sia per il bit 0 sia per il bit 1, è ovvio che siamo nelle stesse ipotesi del processo di Bernoulli, per cui possiamo scrivere che

$$P(k \text{ bit errati}) = \binom{2n+1}{k} p^k (1-p)^{2n+1-k}$$

e quindi

$$P(E_s) = \sum_{k=n+1}^{2n+1} \binom{2n+1}{k} p^k (1-p)^{2n+1-k}$$

Trattandosi di una sommatoria finita di termini, una volta scelto n , ossia scelto il numero di ripetizioni (che è $2n+1$), è chiaro che possiamo valutare numericamente $P(E_s)$. Sostituendo i valori numerici, supponendo $p=10^{-2}$, si ottiene quanto segue:

n	$P(E_s)$
1	$3 \cdot 10^{-4}$
2	10^{-5}
3	$3.5 \cdot 10^{-7}$
4	$1.28 \cdot 10^{-8}$
...	...

Questa tabella evidenzia come la probabilità di errore decresca notevolmente all'aumentare di n: per esempio essa ci dice che, per 5 ripetizioni (ossia $n=4$) la probabilità di errore vale $1.28 \cdot 10^{-8}$, ossia, in altre parole, il ricevitore riceve un simbolo sbagliato approssimativamente ogni 10^8 simboli.

Naturalmente, ci sono anche degli svantaggi nel codice a ripetizione: infatti, è chiaro che quanti più simboli devono essere inviati, tanto maggiore è il tempo necessario alla trasmissione, per cui l'abbassamento della probabilità di errore, mediante il metodo del codice a ripetizione, comporta un innalzamento del tempo di trasmissione.

Codice a controllo di parità

Così come il codice a ripetizione esaminato nei paragrafi precedenti, il cosiddetto **codice a controllo di parità** è un'altra tecnica utilizzata per prevenire gli errori di trasmissione dovuti al canale. Vediamo di che si tratta.

Supponiamo che la sequenza di bit che il canale deve trasmettere sia la seguente:

0 0 1 0 1 1 1 0 0 1 0 1 0 1 1

La prima cosa che viene effettuata, da parte del codificatore di canale, consiste nel raggruppare i bit della sequenza da trasmettere in gruppi di $2n-1$ bit, dove ovviamente il valore di n dipende dal progetto. Per esempio, prendiamo $n=2$, per cui i gruppi che otteniamo sono composti da 3 bit e, nel nostro esempio, sono i seguenti:

0 0 1 0 1 1 1 0 0 1 0 1 0 1 1

A ciascuno di questi gruppi viene aggiunto un ulteriore bit a destra: questo bit ha lo scopo di rendere PARI (da cui il nome appunto di controllo di parità) il numero di bit a 1 contenuti in ciascun gruppo. Evidentemente, ciò comporta che il bit aggiuntivo, detto **bit di controllo**, venga posto a 1 quando il numero di bit a 1 è dispari, mentre, in caso contrario, venga posto a 0.

Nel nostro esempio, la sequenza viene allora modificata nel modo seguente:

0 0 1 1 0 1 1 0 1 0 0 1 1 0 1 0 0 1 1 0

Questa è la sequenza che viene dunque inviata dal codificatore di canale sul canale stesso per la trasmissione. Vediamo allora cosa accade in fase di ricezione; in particolare, supponiamo che si verifichino degli errori durante la trasmissione e che quindi la sequenza che giunge al decodificatore di canale sia fatta nel modo seguente:

0 0 1 1 0 0 1 0 1 1 1 1 0 0 1 0 0 1 1 0

↓ ↓ ↓ ↓
 (0 0 1 1) (0 0 1 0) (1 1 1 1) (0 0 1 0) (0 1 1 0)

Le freccette nel disegno indicato evidentemente gli errori che si sono avuti durante la trasmissione.

Il principio di fondo è quello per cui il decodificatore, per controllare se ci sono stati o meno degli errori, controlla che, per ciascun gruppo di n bit (dove 2n-1 bit sono i bit di informazione, mentre l'n° bit è quello di parità), il numero di bit ad 1 sia pari:

- quando il numero di bit ad 1 è pari, il decodificatore prende per buono il gruppo, elimina il bit di parità e trasmette il gruppo al ricevitore;
- quando il numero di bit ad 1 è dispari, il decodificatore deduce che c'è stato un errore e quindi chiede la *ritrasmissione* della stessa sequenza; naturalmente, in base a come è stato progettato il decodificatore di canale, il numero di ritrasmissione possibili può essere anche infinito (nel senso che il decodificatore chiede la ritrasmissione finché non riceve una sequenza corretta), ma può anche essere finito (nel senso che, pur non avendo ancora ricevuto una sequenza corretta, esso non chiede più la ritrasmissione dopo che ne sono state già effettuate un certo numero fisso).

Da quanto detto si deduce che si possono verificare, in ricezione, tre diverse situazioni:

- ricezione corretta : quando, sul generico gruppo di n bit, non si è verificato alcun errore, è chiaro che il gruppo è corretto, per cui esso viene inviato al ricevitore (privato del bit di parità);
- ricezione sbagliata o errore non rilevato : quando, sul generico gruppo di n bit, si è verificato un numero PARI di errori, il gruppo contiene un numero pari di bit ad 1, per cui il decodificatore accetta il gruppo e lo trasmette al ricevitore; la differenza con il caso precedente è che la sequenza è in questo caso sbagliata, ossia è diversa da quella emessa dalla sorgente;
- errore rilevato : quando, sul generico gruppo di n bit, si è verificato un numero DISPARI di errori, il gruppo non contiene certamente un numero pari di bit ad 1, per cui il decodificatore deduce la presenza di almeno un errore e quindi chiede la ritrasmissione del gruppo.

Analizziamo allora da un punto di vista probabilistico queste eventualità, facendo sempre l'ipotesi di avere a disposizione un **CBS** (*canale binario simmetrico*) con probabilità di errore pari a p.

La ricezione corretta si ha quando, su n bit, non si verifica nessun errore, per cui, applicando la *formula di Bernoulli*, possiamo scrivere che

$$P\left(\begin{matrix} \text{ricezione} \\ \text{corretta} \end{matrix}\right) = (1-p)^{2n}$$

L'errore non rilevato si ha invece quando il numero di errori, su n bit, è pari:

$$P\left(\begin{matrix} \text{errore non} \\ \text{rilevato} \end{matrix}\right) = \sum_{k=0}^n \binom{2n}{2k} p^{2k} (1-p)^{2n-2k}$$

In modo analogo, l'errore rilevato si verifica quando il numero di errori, su n bit, è dispari:

$$P\left(\begin{matrix} \text{errore} \\ \text{rilevato} \end{matrix}\right) = \sum_{k=0}^n \binom{2n}{2k-1} p^{2k-1} (1-p)^{2n-2k+1}$$

Le tre probabilità appena calcolate sono relative al generico gruppo di bit. Il discorso più generale da fare è più complesso; consideriamo ad esempio la *probabilità di ricezione corretta*: è chiaro che la ricezione corretta può avvenire subito, nel senso che, sul gruppo di n bit considerato, non si verifica alcun errore, ma può avvenire anche dopo una ritrasmissione, oppure dopo due ritrasmissioni e così via fino, teoricamente, a ∞ ritrasmissioni. Di conseguenza, possiamo scrivere in generale che

$$P\left(\begin{array}{c} \text{ricezione} \\ \text{corretta} \end{array}\right) = (1-p)^{2n} + P(R)(1-p)^{2n} + P(R)P(R)(1-p)^{2n} + \dots = (1-p)^{2n} \sum_{k=0}^{\infty} P^k(R)$$

dove ovviamente abbiamo indicato con $P(R)$ la **probabilità di ritrasmissione**, ossia la probabilità di errore rilevato calcolata prima. Ricordando poi che quella somma non è altro che quella della serie geometrica, possiamo concludere che

$$\boxed{P\left(\begin{array}{c} \text{ricezione} \\ \text{corretta} \end{array}\right) = \frac{(1-p)^{2n}}{1-P(R)}}$$

In modo del tutto analogo, l'evento dell'errore non rilevato può verificarsi dopo 0 ritrasmissioni, dopo 1 ritrasmissione, dopo 2 ritrasmissioni fino a ∞ , per cui, se noi poniamo

$$P(\text{ENR}) = \sum_{k=0}^n \binom{2n}{2k} p^{2k} (1-p)^{2n-2k}$$

dobbiamo scrivere che

$$\boxed{P\left(\begin{array}{c} \text{errore non} \\ \text{rilevato} \end{array}\right) = P(\text{ENR}) \sum_{k=0}^{\infty} P^k(R) = \frac{P(\text{ENR})}{1-P(R)}}$$

L'ultima cosa su cui ci concentriamo è cosiddetto **numero medio di ritrasmissioni**.

Indichiamo intanto con n_R la variabile aleatoria che ci quantifica il numero di ritrasmissioni per il generico gruppo di n bit. Il numero medio di ritrasmissioni non è altro che la media di questa variabile aleatoria: per calcolarlo, ci basta perciò applicare la definizione, ossia ci basta scrivere che

$$n_R = \sum_{k=0}^{\infty} k P(n_T = k)$$

Dobbiamo dunque calcolare $P(n_T = k)$, ossia la probabilità che, sul generico gruppo di n bit, ci siano k ritrasmissioni.

Cominciamo dal caso in cui $k=0$: dire che $k=0$, ossia che non si verifica nessuna ritrasmissione, equivale a dire che il decodificatore ha accettato il gruppo; questo accade sia quando non si è verificato alcun errore sul gruppo sia quando il numero di errori è pari: indichiamo allora con $P(C)$ la probabilità che non ci sia nessun errore e con $P(NR)$ la probabilità che ci sia stato un numero pari di errori (entrambe queste probabilità sono state calcolate all'inizio), possiamo scrivere che

$$P(n_T = 0) = P(C) + P(NR)$$

Vediamo adesso il caso in cui $k=1$: dire che $k=1$ significa che la prima volta il decodificatore ha rilevato l'errore e ha richiesto la ritrasmissione e che la seconda volta, ossia dopo la ritrasmissione, esso non ha rilevato alcun errore. Possiamo perciò scrivere che

$$P(n_T = 1) = (P(C) + P(NR))P(R)$$

dove $P(R)$ è la probabilità di ritrasmissione calcolata in precedenza.

Chiaramente, per k generico, noi avremo che

$$P(n_T = k) = (P(C) + P(NR))P^k(R)$$

per cui abbiamo che

$$n_R = (P(C) + P(NR)) \sum_{k=0}^{\infty} kP^k(R)$$

Questa relazione può essere facilmente semplificata; in particolare, riguardo quella sommatoria possiamo fare i seguenti passaggi al fine di ricondurci ancora una volta alla serie geometrica:

$$\begin{aligned} \sum_{k=0}^{\infty} kP^k(R) &= P(R) \sum_{k=0}^{\infty} kP^{k-1}(R) = P(R) \sum_{k=0}^{\infty} \frac{d}{dP(R)} (P^k(R)) = P(R) \frac{d}{dP(R)} \sum_{k=0}^{\infty} P^k(R) = \\ &= P(R) \frac{d}{dP(R)} \left(\frac{1}{1-P(R)} \right) = P(R) \frac{1}{(1-P(R))^2} \end{aligned}$$

Possiamo dunque scrivere che

$$n_R = (P(C) + P(NR)) \frac{P(R)}{(1-P(R))^2}$$

Non è ancora finita, in quanto è evidente che vale la relazione

$$P(C) + P(NR) + P(R) = 1$$

Infatti, sul generico gruppo di bit, le situazioni possibili sono solo la ricezione corretta (C), la ricezione sbagliata (NR) o la ritrasmissione (R), per cui la somma delle rispettive probabilità deve essere uguale ad 1. Da quella relazione si deduce che

$$P(C) + P(NR) = 1 - P(R)$$

per cui possiamo concludere che il *numero medio di ritrasmissioni* per il generico gruppo di bit è

$$\boxed{n_R = \frac{P(R)}{1 - P(R)}}$$

Modello uniforme di probabilità in spazi campione finiti

Il caso più semplice di *modello di probabilità* su un fenomeno (o esperimento) avente un numero finito N di risultati, si ottiene quando sono tutte uguali tra loro le probabilità degli eventi elementari. Indicato con A_k il generico evento elementare del fenomeno, ciò significa che

$$P(A_k) = \frac{1}{N}$$

Un modello siffatto prende il nome di **modello uniforme**.

Considerato allora un qualsiasi evento A dell'esperimento, è ovvio che la probabilità che esso si verifichi è data dal rapporto tra il numero di eventi elementari favorevoli ad A ed il numero totale di eventi elementari possibili: possiamo cioè scrivere che $P(A) = \frac{k}{N}$, dove k è appunto il numero di eventi elementari favorevoli ad A .

L'esempio classico è quello del lancio di un dado a 6 facce. Assumendo tutte uguali ad $1/6$ le probabilità elementari, è ovvio che la probabilità che venga fuori un numero pari è $3/6$, in quanto 6 sono gli eventi elementari possibili e 3 di essi (2,4,6) sono quelli favorevoli all'evento.

ESEMPIO

Supponiamo di avere a disposizione 10 *diodi* identici tra loro e di sapere che 3 sono guasti. Vogliamo la probabilità che, scegliendone 5 a caso, se ne trovino 2 guasti.

La prima cosa da fare è individuare lo spazio dei campioni sul quale fare i nostri ragionamenti. Il fenomeno (o esperimento) che stiamo considerando consiste nello scegliere 5 diodi su 10: i risultati possibili sono dunque in numero di $\binom{10}{5}$, per cui questa è la dimensione dello spazio degli eventi, ossia il numero di eventi elementari.

Adesso, il fatto che i diodi siano identici ci consente di utilizzare il *modello uniforme* introdotto prima. Noi vogliamo la probabilità che, dei 5 diodi presi, 2 siano guasti e gli altri 3 funzionanti. Detto in altre parole, noi vogliamo la probabilità dell'evento A per cui, prendendo 5 diodi, ce ne siano 2 guasti e 3 sani.

Dato che, nei 10 diodi, ce ne sono 3 guasti, noi abbiamo $\binom{3}{2}$ possibilità di scegliere 2 diodi guasti.

In modo analogo, dato che, nei 10 diodi, ce ne sono 7 sani, noi abbiamo $\binom{7}{3}$ possibilità di scegliere 3 diodi sani. Possiamo allora scrivere che

$$P(A) = \frac{\binom{3}{2} \binom{7}{3}}{\binom{10}{5}}$$

Variabili aleatorie

DEFINIZIONE

Fino ad ora, abbiamo considerato spazi di probabilità in cui gli eventi sono insiemi astratti non necessariamente numerici: basti pensare allo spazio di probabilità del lancio di un dado o della estrazione di una carta da un mazzo o di una pallina da un'urna. E' spesso conveniente associare, ad ogni risultato dell'esperimento casuale considerato, un valore numerico: tale valore può essere scelto in modo del tutto convenzionale oppure può essere introdotto al solo scopo di classificare in modo più preciso i risultati o può anche rappresentare il valore assunto da una grandezza fisica in corrispondenza del risultato ottenuto.

Per esempio, nella produzione di un certo tipo di oggetti, è possibile associare convenzionalmente il valore logico 0 a ciascun pezzo difettoso ed il valore logico 1 a ciascun pezzo sano. Oppure, se l'esperimento consiste nella produzione in serie di un certo tipo di resistori, viene immediato pensare di caratterizzare il resistore prodotto (cioè il risultato dell'esperimento) con il valore della sua resistenza.

Consideriamo perciò un esperimento casuale il cui spazio di probabilità sia (S,P) , dove S è lo spazio di eventi, ossia l'insieme di tutti i possibili risultati osservabili per il fenomeno, e P è la funzione probabilità, che a ciascun evento di S associa la probabilità che esso si verifichi. Consideriamo inoltre una funzione così definita:

$$X: S \rightarrow \mathbb{R}^k$$

Si tratta cioè di una funzione che a ciascun evento contenuto in S associa un certo valore numerico, che sarà un valore reale nel caso che $k=1$ oppure un vettore di elementi reali se $k>1$. Da notare che non si tratta di una funzione probabilità: si tratta semplicemente di una legge analitica che fa corrispondere, a ciascun evento, un certo valore determinato in un modo prestabilito.

Una funzione del tipo di X prende il nome di **variabile aleatoria (o anche casuale) k -dimensionale**.

Nel seguito noi ci limiteremo a considerare i casi in cui $k=1$ e $k=2$.

Consideriamo alcuni semplici esempi di variabili aleatorie

Prendiamo l'esperimento casuale consistente nel lancio di una moneta: lo spazio campione è perciò $S = \{\text{testa}, \text{croce}\}$ e la funzione probabilità comprende i valori $P(\text{croce}) = 0.5$ e $P(\text{testa}) = 0.5$. Un esempio di variabile aleatoria può essere la funzione X che associa il valore 0 all'evento croce ed il valore 1 all'evento testa: possiamo perciò scrivere che

$$P(X = 1) = P(\text{testa}) = 0.5$$

$$P(X = 0) = P(\text{croce}) = 0.5$$

Consideriamo adesso l'esperimento casuale consistente nel lancio di un dado a 6 facce: lo spazio dei campioni è evidentemente $S = \{f_1, f_2, f_3, f_4, f_5, f_6\}$ e la funzione probabilità è

$$P(f_i) = \frac{1}{6} \quad i = 1, 2, \dots, 6$$

Un esempio di variabile aleatoria per questo esperimento è quella che, a ciascuna faccia, associa il suo numero: possiamo cioè definirla tale variabile aleatoria con la notazione $X(f_i) = i$.

Un altro esempio di ancora di spazio di campioni può essere quello di tutti i segnali ad energia finita. Indichiamo perciò con $S = \{s_1(t), s_2(t), \dots\}$ tale spazio, che è evidentemente uno spazio infinito non numerabile. Una variabile aleatoria per questo spazio può essere quella che, a ciascun segnale, associa la sua energia: ciò significa che essa è definita mediante la relazione

$$X(s_k(t)) = \int_{-\infty}^{+\infty} |s_k(t)|^2 dt$$

Supponiamo ora di considerare un generico spazio dei campioni S e di considerare una variabile aleatoria X altrettanto generica. Supponiamo inoltre di volere la probabilità che $X=a$, con a costante reale: questo significa che ci interessa la probabilità che si verifichi uno qualsiasi degli $s \in S$ per i quali $X(s)=a$. Indicato con A l'insieme che racchiude tutti questi eventi, lo possiamo rappresentare nel modo seguente:

$$A = \{s \in S | X(s) = a\}$$

Calcolare $P(X=s)$ equivale dunque a calcolare la probabilità che si verifichi l'evento A .

In modo analogo, se vogliamo calcolare la probabilità che si verifichi la condizione $a \leq X \leq b$, noi dobbiamo calcolare la probabilità che si verifichi l'evento

$$B = \{s \in S | a \leq X(s) \leq b\}$$

Diamo adesso una importante definizione:

Def. Avendo detto che una variabile aleatoria associa ad ogni evento di S un certo valore numerico (scalare o vettoriale), noi diremo **rango della variabile X** il codominio di X , cioè l'insieme $X(S)$ dei valori osservabili per la funzione. Lo indicheremo con R_X .

Variabile aleatoria discreta

Def. Noi diremo che una variabile aleatoria è **discreta** se il rango di X è un insieme finito o comunque numerabile.

Quindi, dire che una certa variabile X è discreta significa dire che il suo rango è del tipo

$$R_X = \{x_1, x_2, \dots, x_n, \dots\}$$

dove $x_1, x_2, \dots, x_n, \dots$ (con n finito nel caso di R_X finito oppure indeterminato se R_X è infinito numerabile) sono degli scalari se X è monodimensionale oppure dei vettori a k componenti se X è k -dimensionale.

Allora, noi saremo sempre interessati a valutare la probabilità che la variabile aleatoria X assuma quei valori, ossia saremo interessati a valutare

$$P(X=x_1) \quad P(X=x_2) \quad \dots\dots\dots$$

Questi numeri, che spesso indicheremo con le lettere minuscole p_1, p_2, \dots , godono ovviamente di 2 proprietà:

- trattandosi di valori di probabilità relativi ad uno stesso fenomeno, essi sono tutti maggiori o al più uguali a zero;
- inoltre, per lo stesso motivo, la loro somma, per n finito o infinito, è sempre pari ad 1.

La funzione P che associa ad ogni valore x_i del rango di X la probabilità che X assuma quel valore prende il nome di **funzione di probabilità di X** . L'insieme delle coppie (x_i, p_i) , con $i=1,2,\dots$, costituisce invece ciò che si chiama **distribuzione di probabilità di X** .

Ad ogni modo, noi ci occuperemo poco di variabili aleatorie discrete, mentre ci concentreremo maggiormente su di un'altra classe, che è quella che viene introdotta nel prossimo paragrafo.

CONTINUITÀ DI UNA VARIABILE ALEATORIA E FUNZIONE DENSITÀ

Consideriamo una generica variabile aleatoria

$$X: S \rightarrow \mathbb{R}^k$$

con $k=1$ o $k=2$. Noi diremo che si tratta di una variabile aleatoria **continua** se esiste una funzione reale $f: \mathbb{R}^k \rightarrow \mathbb{R}$ che gode di **3 proprietà caratteristiche**:

1. $f(x) \geq 0 \quad \forall x \in \mathbb{R}^k$

2. $\int_{-\infty}^{+\infty} f(x) dx = 1$

3. $\forall A \subset \mathbb{R}^k: P(X \in A) = \int_A f(x) dx$: questa proprietà dice che, preso un qualsiasi sottoinsieme A di \mathbb{R}^k , la probabilità che i valori riscontrati per X facciano parte di A è data da quell'integrale

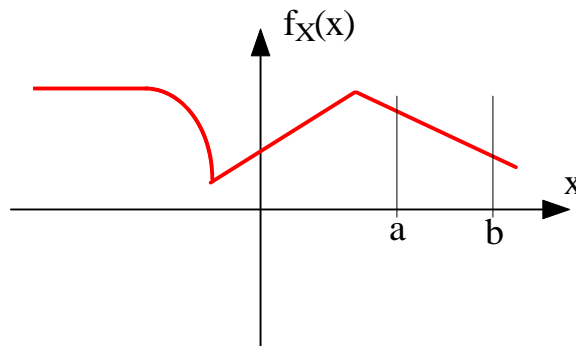
Una funzione f che goda delle prime 2 proprietà si dice che è una **densità**; la terza proprietà fa' invece sì che ad f si dia il nome di **densità di probabilità**.

Come sarà chiaro tra poco e come si deduce proprio della terza proprietà, una funzione densità di probabilità serve a indicare il peso di singoli punti o di interi sottoinsiemi di \mathbb{R}^k .

Per indicare che la funzione $f(x)$ rappresenta una densità di probabilità relativa ad una variabile aleatoria X , si usa indicarla con $f_X(x)$.

Il significato matematico della densità

Riguardo la terza proprietà della funzione densità f_X relativa ad una certa variabile aleatoria X , possiamo visualizzare il suo significato nel modo che segue: la funzione densità può essere rappresentata su di un grafico cartesiano in funzione di x ed avrà un certo andamento:



Fissiamo adesso un intervallo $[a,b]$ generico, con a e b che potrebbero anche essere infiniti; se noi andiamo a calcolare, mediante la *proprietà 3*, il valore di $P(a < X < b)$, non facciamo altro che misurare l'area della regione di piano compresa tra la curva di f_X , l'asse delle ascisse ed i punti a e b . Questo dal punto di vista matematico; dal punto di vista probabilistico, con questo calcolo misuriamo il PESO dell'intervallo $[a,b]$ relativamente alla variabile aleatoria X presa in esame.

Osservazione: peso nullo dei singoli punti

Consideriamo una variabile aleatoria X continua e consideriamo un qualsiasi punto x di \mathbb{R}^k . Se X è continua, essa avrà una certa densità f_X . Vogliamo calcolare la probabilità che per X si osservi il valore x , ossia vogliamo $P(X=x)$. Utilizzando la terza proprietà di f_X , noi abbiamo che

$$P(X = x) = \int_{A=\{x\}} f_X(x) dx$$

dove abbiamo cioè posto l'insieme A pari semplicemente al valore x . Tuttavia, proprio per il fatto di essere costituito da 1 solo punto, è noto che ad A si assegna misura nulla ed un integrale esteso ad un insieme di misura nulla è sempre 0, quale che sia la funzione integranda. Quindi $P(X=x) = 0$.

Questa è una proprietà fondamentale delle variabili aleatorie continue: *data una variabile aleatoria, ogni singolo punto del rango (cioè del codominio), che può essere \mathbb{R}^k o un suo sottoinsieme, ha peso 0.*

Apparentemente questo fatto può sembrare poco intuitivo: tuttavia, dobbiamo sempre tenere presente che, se assumiamo che X possa assumere tutti i valori di un determinato intervallo, dobbiamo da un lato dire che la probabilità che assuma un determinato valore sia 0, ma, contemporaneamente, dobbiamo anche ammettere che la probabilità uguale a zero non equivalga alla impossibilità.

In altre parole, nel caso di variabili aleatorie continue, il fatto che $P(A)=0$ NON implica necessariamente che A sia un insieme vuoto.

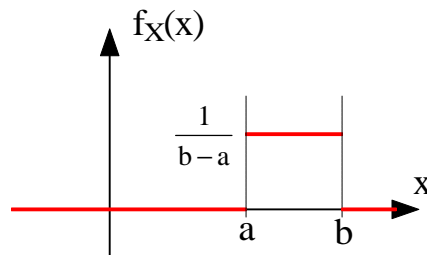
DENSITÀ DI UNA VARIABILE ALEATORIA CONTINUA DISTRIBUITA UNIFORMEMENTE SU UN INTERVALLO

Fissiamo un certo intervallo $[a,b]$ e consideriamo una variabile aleatoria monodimensionale X il cui rango R_X coincida proprio con tale intervallo.

Consideriamo adesso la funzione reale

$$f_X(x) = \begin{cases} \frac{1}{b-a} & \text{se } x \in [a, b] \\ 0 & \text{altrimenti} \end{cases}$$

il cui andamento in funzione x è il seguente:



Vogliamo intanto vedere se si tratta di una funzione densità, ossia se verifica le prime 2 proprietà: assume sempre valori positivi o al più nulli ed inoltre, calcolandone l'integrale esteso ad $[a,b]$, esso è pari ad 1. Quindi si tratta di una funzione densità.

Prendiamo adesso un sottoinsieme A di R e, sfruttando la terza proprietà, calcoliamo la probabilità che il rango di X sia contenuto in A , ossia calcoliamo $P(X \in A)$:

$$P(X \in A) = \int_A f(x) dx$$

Dato che la funzione f assume valori nulli all'esterno di $[a,b]$, anche il suo integrale esteso a punti esterni a tale intervallo sarà nullo, per cui noi possiamo scrivere

$$P(X \in A) = \int_a^b \frac{1}{b-a} dx = 1$$

Si dice allora che la variabile aleatoria X è **distribuita uniformemente su $[a,b]$** .

DISTRIBUZIONE CUMULATIVA DI UNA VARIABILE ALEATORIA CONTINUA

Supponiamo di avere una variabile aleatoria X continua e supponiamo che sia f_X la sua densità. Dato un qualsiasi punto x del codominio di X , calcoliamo la probabilità che la variabile X assuma un valore minore o al più uguale ad x : sfruttando la funzione densità, possiamo scrivere che

$$P(X \leq x) = \int_{-\infty}^{+x} f_X(x) dx$$

Si pone allora

$$F_X(x) = P(X \leq x) = \int_{-\infty}^x f_X(x) dx$$

Questa funzione $F_X(x)$ è una funzione reale di variabile reale che non pesa ogni singolo punto ma pesa da $-\infty$ fino a quel punto. Essa gode della proprietà fondamentale per cui

$$f_X(x) = \frac{dF_X(x)}{dx}$$

ossia è una primitiva della funzione $f_X(x)$. Ma la funzione F_X gode anche di altre proprietà, che sono le seguenti:

1. $\lim_{x \rightarrow \infty} F_X(x) = 1$: infatti, è certo che $X \leq \infty$
2. $\lim_{x \rightarrow -\infty} F_X(x) = 0$: infatti è impossibile che $X \leq -\infty$
3. F_X è una funzione continua
4. F_X è strettamente monotona (crescente o decrescente) sul rango di X , mentre al di fuori del rango, potrebbe anche essere costante.

Questa funzione F_X prende il nome di **distribuzione cumulativa** o **funzione di distribuzione** della variabile aleatoria X . E' evidente che la conoscenza di F_X equivale alla conoscenza di f_X : infatti, in base alla relazione vista prima, basta derivare F_X per conoscere f_X .

Vediamo una serie di altre proprietà di F_X .

La prima è la seguente:

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

La dimostrazione è immediata: infatti, dato che

$$P(X \leq x_2) = P(X \leq x_1 \cup x_1 \leq X \leq x_2) = P(X \leq x_1) + P(x_1 \leq X \leq x_2)$$

è chiaro che

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

Chiaramente, una conseguenza immediata di quest'ultima proprietà e della definizione di densità di probabilità è che

$$P(x_1 \leq X \leq x_2) = \int_{x_1}^{x_2} f_X(x) dx$$

Sempre sulla falsa riga di quest'ultima e ricordando che $F_X(-\infty)=0$, è chiaro che

$$P(X \leq x_1) = \int_{-\infty}^{x_1} f_X(x) dx = F_X(x_1)$$

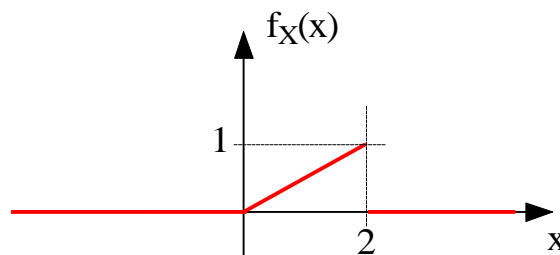
Un'altra proprietà è la seguente:

$$\text{Se } x_1 \leq x_2 \Rightarrow \begin{cases} F_X(x) \text{ monotona crescente} \longrightarrow F_X(x_2) \leq F_X(x_1) \\ F_X(x) \text{ monotona decrescente} \longrightarrow F_X(x_2) \geq F_X(x_1) \end{cases}$$

Esempio: variabile aleatoria continua

Consideriamo una variabile aleatoria X continua, la cui densità sia

$$f_X(x) = \begin{cases} \frac{x}{2} & \text{se } x \in [0,2] \\ 0 & \text{altrimenti} \end{cases}$$



Vogliamo la distribuzione cumulativa di X sulla base della definizione: abbiamo

$$F_X(x) = P(X \leq x) = \int_{-\infty}^x f_X(\tau) d\tau$$

Data la particolare struttura di $f_X(x)$, possiamo spezzare l'integrale in 2 parti:

$$F_X(x) = \int_{-\infty}^0 f_X(\tau) d\tau + \int_0^x f_X(\tau) d\tau$$

Il primo integrale è sicuramente nullo in quanto la funzione integranda, $f_X(\tau)$, è nulla prima di 0: quindi

$$F_X(x) = \int_0^x f_X(\tau) d\tau$$

Questo secondo integrale assume due valori diversi a seconda del valore di x , dato che il valore di $f_X(\tau)$ dipende proprio da x :

- il primo caso è quello in cui $x \in [0,2]$: in questo caso, risulta $f_X(\tau) = \tau/2$, per cui scriviamo che

$$x \in [0,2] \longrightarrow F_X(x) = \int_0^x \frac{\tau}{2} d\tau = \frac{1}{2} \left[\frac{\tau^2}{2} \right]_0^x = \frac{x^2}{4}$$

- il secondo caso è quello in cui x è esterno all'intervallo $[0,2]$: in questo caso, per le ipotesi fatte, risulta $f_X(\tau) = 0$ e quindi anche l'integrale risulta nullo.

Possiamo perciò concludere che

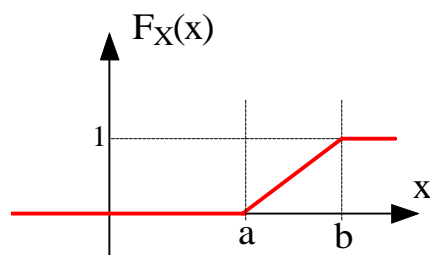
$$F_X(x) = \begin{cases} \frac{x^2}{4} & \text{se } x \in [0,2] \\ 0 & \text{altrimenti} \end{cases}$$

Esempio: distribuzione uniforme su un intervallo

Consideriamo una variabile aleatoria X la cui funzione densità sia

$$f_X(x) = \begin{cases} \frac{1}{b-a} & \text{se } x \in [a, b] \\ 0 & \text{altrimenti} \end{cases}$$

Sappiamo che una X siffatta si dice *uniformemente distribuita* sull'intervallo $[a,b]$. E' immediato comprendere come la sua distribuzione cumulativa $F_X(x)$ sia la seguente:



Esempio: distribuzione esponenziale

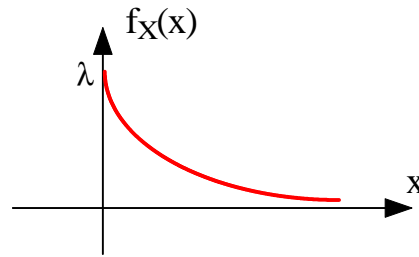
Consideriamo una variabile aleatoria X continua, la cui distribuzione cumulativa sia

$$F_X(x) = \begin{cases} 1 - e^{-\lambda x} & \text{se } x > 0 \\ 0 & \text{se } x \leq 0 \end{cases}$$

Vogliamo la densità di X : ci basta effettuare una derivazione rispetto ad x , dalla quale otteniamo

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & \text{se } x > 0 \\ 0 & \text{se } x \leq 0 \end{cases}$$

L'andamento di questa funzione è il seguente:



Una variabile aleatoria X che abbia una funzione densità di probabilità così fatta si dice che ha una **distribuzione esponenziale con parametro λ** .

Una caratteristica molto importante di una variabile aleatoria X che abbia questa distribuzione è quella di essere **senza memoria**: ciò significa che vale la relazione

$$\boxed{P(X > T + s | X > T) = P(X > s)}$$

La verifica di questa proprietà è immediata: intanto, usando la definizione di probabilità condizionate, abbiamo che

$$P(X > T + s | X > T) = \frac{P(X > T + s \cap X > T)}{P(X > T)}$$

Osservando il numeratore di quella frazione, è evidente che la probabilità che sia $X > T + s$ e $X > T$ è semplicemente la probabilità che $X > T + s$, per cui

$$P(X > T + s | X > T) = \frac{P(X > T + s)}{P(X > T)}$$

Questa può anche essere espressa nella forma

$$P(X > T + s | X > T) = \frac{1 - P(X \leq T + s)}{1 - P(X \leq T)}$$

Adesso, per definizione di distribuzione cumulativa di una variabile aleatoria continua, possiamo scrivere che

$$P(X > T + s | X > T) = \frac{1 - F_X(T + s)}{1 - F_X(T)}$$

Fin qui, il discorso è del tutto generale. Adesso subentra la particolare struttura di $F_X(x)$: infatti, sostituendo l'espressione della distribuzione cumulativa, abbiamo che

$$P(X > T + s | X > T) = \frac{1 - (1 - e^{-\lambda(T+s)})}{1 - (1 - e^{-\lambda T})} = \frac{e^{-\lambda(T+s)}}{e^{-\lambda T}} = \frac{e^{-\lambda T} e^{-\lambda s}}{e^{-\lambda T}} = e^{-\lambda s} = F_X(s) = P(X > s)$$

FUNZIONI DI VARIABILI ALEATORIE

Supponiamo di avere un certo fenomeno il cui spazio degli eventi sia S e supponiamo anche di avere una generica variabile aleatoria $X: S \rightarrow \mathbb{R}^k$, con $k=1$ o $k=2$. Non ci interessano eventuali proprietà di cui possa godere X . Consideriamo invece un'altra funzione H così definita: $H: \mathbb{R}^k \rightarrow \mathbb{R}$.

Si tratta di una funzione che associa ad elementi di \mathbb{R}^k dei precisi valori reali. A questo punto, componiamo le funzione X e H nel modo seguente:

$$S \xrightarrow{X} \mathbb{R}^k \xrightarrow{H} \mathbb{R}$$

Abbiamo cioè costruito la funzione

$$H(X): S \rightarrow \mathbb{R}$$

che evidentemente è una variabile aleatoria monodimensionale. Da notare che il fatto che si tratti di una variabile aleatoria non dipende dalla struttura della funzione H , ma solo dal fatto che X sia una variabile aleatoria.

Vediamo allora se è possibile ricavare le caratteristiche statistiche della variabile aleatoria $H(x)$, note che siano quelle di X , ossia, in definitiva, conoscendo $f_X(x)$ e/o $F_X(x)$.

Intanto, posto per comodità $Y=H(X)$, abbiamo, per definizione di distribuzione cumulativa, che

$$F_Y(y) = P(Y \leq y) = P(H(X) \leq y)$$

Se noi poniamo allora $A = \{x \in \mathbb{R} | H(x) \leq y\}$, è chiaro che

$$F_Y(y) = P(H(X) \leq y) = \int_A f_X(x) dx$$

Esempio: distribuzione cumulativa di X^2 con X generica

Supponiamo di avere una variabile aleatoria X della quale conosciamo la funzione densità $f_X(x)$. Vogliamo trovare la distribuzione dell'altra variabile aleatoria $H(X) = X^2$, che sarà evidentemente anch'essa una variabile aleatoria continua.

Per comodità, poniamo $Y = H(X) = X^2$, per cui il nostro scopo diventa quello di conoscere $F_Y(y)$: infatti, come abbiamo detto prima, se troviamo F_Y , ci basterà derivare per ottenere la distribuzione f_Y della variabile $Y=X^2$.

Possiamo subito applicare la definizione circa la funzione F_Y : si ha che $F_Y(y) = P(Y \leq y)$. Poiché $Y=H(X)$ abbiamo che

$$F_Y(y) = P(H(x) \leq y)$$

Il calcolo di $P(H(x) \leq y)$ si effettua sfruttando la funzione densità f_X (che è nota) della variabile aleatoria X : infatti, si ha che

$$P(H(X) \leq y) = \int_{\{x \in \mathbb{R} | H(x) \leq y\}} f_X(x) dx$$

Questa formula deriva dalla seguente considerazione: l'evento per cui $H(X) \leq y$, dal punto di vista della probabilità di accadere, è equivalente all'evento per cui X assuma quei valori tali che la relazione $H(x) \leq y$ sia verificata; di conseguenza, la probabilità di quell'evento è pari al peso di tali punti per la X .

Poniamo per comodità $A = \{x \in \mathbb{R} | H(x) \leq y\}$, per cui abbiamo che

$$F_Y(y) = P(H(X) \leq y) = \int_A f_X(x) dx$$

A questo punto, si tratta di capire come è fatto questo insieme A sul quale noi effettuiamo la nostra integrazione. I casi possibili sono 2:

- se $y < 0$, è ovvio che A è l'insieme vuoto, in quanto non potrà mai essere verificata la relazione $x^2 \leq y$;
- se, invece $y > 0$, l'insieme A diventa evidentemente

$$A = \{x \in \mathbb{R} | x^2 \leq y\} = \{x \in \mathbb{R} | -\sqrt{y} \leq x \leq \sqrt{y}\}$$

per cui

$$F_Y(y) = \int_{\{-\sqrt{y}, \sqrt{y}\}} f_X(x) dx = \int_{-\sqrt{y}}^{\sqrt{y}} f_X(x) dx$$

Poiché la funzione $F_X(x)$ è per definizione una primitiva di $f_X(x)$, abbiamo che

$$F_Y(y) = F_X(\sqrt{y}) - F_X(-\sqrt{y})$$

Derivando (mediante il teorema di derivazione delle funzioni composte) otteniamo

$$\begin{aligned} f_Y(y) &= \frac{dF_Y(y)}{dy} = \frac{d}{dy} [F_X(\sqrt{y}) - F_X(-\sqrt{y})] = \frac{dF_X(\sqrt{y})}{dx} \frac{d}{dy}(\sqrt{y}) - \frac{dF_X(-\sqrt{y})}{dx} \frac{d}{dy}(\sqrt{y}) = \\ &= f_X(\sqrt{y}) \frac{d}{dy}(\sqrt{y}) - f_X(-\sqrt{y}) \frac{d}{dy}(\sqrt{y}) = \frac{f_X(\sqrt{y})}{2\sqrt{y}} - \frac{f_X(-\sqrt{y})}{2\sqrt{y}} \end{aligned}$$

In conclusione, quindi, la densità di $Y=X^2$ è

$$f_Y(y) = \frac{f_X(\sqrt{y})}{2\sqrt{y}} + \frac{f_X(-\sqrt{y})}{2\sqrt{y}}$$

Esempio: distribuzione cumulativa di X^2 con X distribuita uniformemente

Supponiamo di avere una variabile aleatoria X distribuita uniformemente e continua nell'intervallo $[0,10]$. Troviamo ancora una volta la distribuzione della variabile aleatoria $Y=H(X) = X^2$.

Possiamo procedere sia applicando la formula ricavata poco fa, sia ripetendo il ragionamento. La prima via è immediata, considerando che

$$f_X(x) = \begin{cases} \frac{1}{10} & \text{se } x \in [0,10] \\ 0 & \text{altrimenti} \end{cases}$$

per cui vediamo la seconda.

Usando la definizione della funzione F_X , abbiamo che

$$F_X(x) = \int_{-\infty}^x f_X(\tau) d\tau$$

Possiamo subito applicare la definizione circa la funzione F_Y : si ha che

$$F_Y(y) = P(Y \leq y) = P(H(x) \leq y) = \int_{A=\{x \in R | H(x) \leq y\}} f_X(x) dx$$

A questo punto, dato che la funzione f_X assume valori non nulli solo nel rango di X , cioè nell'intervallo $[0,10]$, non ha senso calcolare quell'integrale su tutto A , ma solo sull'intersezione con R_X , nella quale, tra l'altro, la funzione f_X vale semplicemente $1/10$: quindi

$$F_Y(y) = \int_{A \wedge R_X} \frac{1}{10} dx = \frac{1}{10} \int_{A \wedge R_X} dx$$

A questo punto, l'unico problema sta nel determinare la misura dell'insieme $A \wedge R_X$. Per ricavare questa misura, la prima cosa da fare è individuare come è fatto l'insieme A per cui, ricordandoci di quello a cui corrisponde, è necessario risolvere la disequazione $H(x) = x^2 \leq y$. Questa disequazione dà questo risultato:

$$-\sqrt{y} \leq x \leq \sqrt{y} \quad \forall y \geq 0$$

L'intersezione con $R_X=[0,10]$ può allora dare 2 diversi risultati a seconda del valore scelto per la variabile (reale) y : infatti, l'estremo sinistro dell'intersezione è comunque 0; l'estremo destro è 10 oppure $y^{1/2}$ a seconda di quale dei 2 valori sia minore.

TEOREMA: CALCOLO DI F_Y NELL'IPOTESI DI INVERTIBILITÀ DI $Y=H(X)$

Una sicura facilitazione per i calcoli del tipo di quelli svolti nell'esercizio precedente risulta nel caso particolare in cui la variabile aleatoria $H(X)$ è invertibile sul rango di X . Infatti, sotto questa ipotesi, esiste la funzione inversa di H e ciò consente di facilitare il calcolo di $P(H(X) \leq y)$.

Ricordandoci che l'invertibilità di una funzione su un certo intervallo implica la stretta monotonia della funzione sullo stesso intervallo, si possono allora presentare 2 differenti casi:

- se $H(X)$ è strettamente crescente, si ha che

$$P(H(X) \leq y) = P(X \leq H^{-1}(y))$$

- se $H(X)$ è strettamente decrescente, si ha invece che

$$P(H(X) \leq y) = P(X \geq H^{-1}(y))$$

Quindi possiamo scrivere che

$$F_Y(y) = P(H(X) \leq y) = \begin{cases} P(X \leq H^{-1}(y)) & \text{per } H(x) \text{ crescente} \\ P(X \geq H^{-1}(y)) & \text{per } H(x) \text{ decrescente} \end{cases}$$

Fissato y , è ovvio che $H^{-1}(y)$ è un numero, per cui, in base alla definizione di distribuzione cumulativa di una variabile aleatoria, quella relazione diventa

$$F_Y(y) = P(H(X) \leq y) = \begin{cases} F_X(H^{-1}(y)) & \text{per } H(x) \text{ crescente} \\ 1 - F_X(H^{-1}(y)) & \text{per } H(x) \text{ decrescente} \end{cases}$$

A questo punto, ricordando che $f_Y(y) = \frac{dF_Y(y)}{dy}$, possiamo scrivere che

$$f_Y(y) = \begin{cases} f_X(H^{-1}(y)) \frac{dH^{-1}(y)}{dy} & \text{per } H(x) \text{ crescente} \\ -f_X(H^{-1}(y)) \left(-\frac{dH^{-1}(y)}{dy} \right) & \text{per } H(x) \text{ decrescente} \end{cases}$$

e quindi possiamo compattare queste due relazioni nell'unica relazione

$$\boxed{f_Y(y) = f_X(H^{-1}(y)) \frac{dH^{-1}(y)}{dy}}$$

Per esempio, se fosse $Y=aX+b$, avremmo

$$f_Y(y) = f_X\left(\frac{y-b}{a}\right) \frac{1}{|a|}$$

Esempio: $H(X)=3X+1$

Supponiamo di avere una variabile aleatoria X la cui densità sia

$$f_X(x) = \begin{cases} \frac{x}{2} & \text{se } x \in [0,2] \\ 0 & \text{altrimenti} \end{cases}$$

Consideriamo adesso la funzione $H(x)=3x+1$ mediante la quale costruiamo la nuova variabile aleatoria $Y=H(X)=3X+1$. Vogliamo la distribuzione di probabilità di Y . Usiamo anche qui il concetto di distribuzione cumulativa: si ha intanto che

$$F_Y(y) = P(H(X) \leq y) = P(3X+1 \leq y) = P\left(X \leq \frac{y-1}{3}\right) = \int_{-\infty}^{\frac{y-1}{3}} f_X(x) dx$$

Ora, data la struttura di $f_X(x)$, è ovvio che

$$F_Y(y) = \int_{-\infty}^{\frac{y-1}{3}} f_X(x) dx = \int_0^{\frac{y-1}{3}} f_X(x) dx$$

Inoltre, a questo punto bisogna distinguere due casi, a seconda che la quantità $\frac{y-1}{3}$ sia maggiore o minore di 1:

$$F_Y(y) = \int_0^{\frac{y-1}{3}} f_X(x) dx = \begin{cases} \int_0^2 \frac{x}{2} dx & \frac{y-1}{3} > 2 \\ \int_0^{\frac{y-1}{3}} \frac{x}{2} dx & \text{altrimenti} \end{cases} = \begin{cases} \frac{1}{2} \cdot 2 & \frac{y-1}{3} > 2 \\ \frac{1}{2} \cdot \frac{1}{2} \left(\frac{y-1}{3}\right)^2 & \text{altrimenti} \end{cases} = \begin{cases} 1 & \frac{y-1}{3} > 2 \\ \frac{1}{4} \left(\frac{y-1}{3}\right)^2 & \text{altrimenti} \end{cases}$$

Nota F_Y , ci basta derivare per ottenere f_Y : quindi

$$f_Y(y) = \begin{cases} 0 & \frac{y-1}{3} > 2 \\ \frac{1}{2} \left(\frac{y-1}{3}\right) & \text{altrimenti} \end{cases}$$

Aspettazione di una variabile aleatoria

DEFINIZIONE

Consideriamo una variabile aleatoria X . La funzione di distribuzione oppure la densità di probabilità (nel caso X sia continua) forniscono una descrizione statistica completa di X in quanto consentono di determinare la probabilità di ogni evento $x_1 < X < x_2$. Tuttavia, *in molte applicazioni si ha interesse a conoscere non tanto la legge di distribuzione, quanto alcuni indici caratteristici di una variabile aleatoria che riassumano gli aspetti principali della legge di distribuzione*. Il nostro scopo è perciò quello di introdurre alcuni parametri che forniscono delle informazioni aggiuntive circa una variabile aleatoria.

Sia data perciò una generica variabile aleatoria continua X e sia $f_X(x)$ la sua densità: si definisce **aspettazione** o **media** o **momento del 1° ordine** o **valor medio** di X il numero reale

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

Si definisce, in generale, **momento del k° ordine** di X il numero reale

$$E(X^k) = \int_{-\infty}^{+\infty} x^k f_X(x) dx$$

dove $k=1,2,3,\dots$

Facciamo osservare come la *media* di X esista ed è finita solo nell'ipotesi che dia un valore finito il seguente integrale:

$$\int_{-\infty}^{+\infty} |x| f_X(x) dx$$

Quale può essere una interpretazione fisica, concreta del concetto di media di una variabile aleatoria? Possiamo dire che la media fornisce una indicazione sulla posizione attorno alla quale si raggruppano i valori della X : in questo senso, essa è un indice di posizione della variabile.

PROPRIETÀ DELLA MEDIA DI UNA VARIABILE ALEATORIA

Vediamo velocemente alcune proprietà della media:

1. Se $X \geq 0 \Rightarrow E(X) \geq 0$
2. $E(X+Y) = E(X) + E(Y)$
3. $E(cX) = cE(X) \quad \forall c \in \mathbb{R}$
4. $E(c) = c \quad \forall c \in \mathbb{R}$
5. Se $X \leq Y \Rightarrow E(X) \leq E(Y)$

Teorema

Sia data la variabile aleatoria X della quale siano note tutte le caratteristiche statistiche. Sia data poi un'altra variabile aleatoria $Y=H(X)$ costruita a partire da X . Si dimostra che il generico momento del k° ordine di Y si può calcolare mediante la relazione

$$E(Y^k) = \int_{-\infty}^{+\infty} y^k f_Y(y) dy = \int_{-\infty}^{+\infty} H^k(x) f_X(x) dx$$

Varianza e deviazione standard di una variabile aleatoria

DEFINIZIONE

Sia data la variabile aleatoria continua X . Si definisce **varianza di X** il numero reale

$$\sigma_X^2 = \text{Var}(X) = E[(X - E(X))^2]$$

Quindi, data la variabile aleatoria X , si calcola la sua media, la si sottrae ad X , la si eleva al quadrato e si calcola ancora una volta la media.

Si definisce invece **deviazione standard di X** il numero reale

$$\sigma_X = \sqrt{\text{Var}(X)}$$

In analogia a quanto vale e a come viene indicata la deviazione standard, talvolta la varianza di X si indica con il simbolo σ_X^2 .

L'interpretazione fisica concreta della deviazione standard di una variabile aleatoria, in accordo a quanto detto circa la media della stessa variabile, è la seguente: se $E(X)$ è valore intorno al quale di dispongono i valori della X , $\text{Var}(X)$ indica di quanto tali valori si discostano da tale valore medio.

La varianza, invece, offre una indicazione circa l'addensamento dei valori della variabile attorno al valor medio: in questo senso si dice che essa costituisce una indice di dispersione.

Proprietà della varianza di una variabile aleatoria

Vediamo anche qui velocemente qualche proprietà della varianza:

1. $\text{Var}(X) \geq 0$
2. $\text{Var}(X) = E(X^2) - (E(X))^2$
3. $\text{Var}(cX) = c^2 \text{Var}(X) \quad \forall c \in \mathbb{R}$
4. $\text{Var}(X) = \text{Var}(-X)$
5. $\text{Var}(c) = 0 \quad \forall c \in \mathbb{R}$

Cenni sulle variabili aleatorie bidimensionali

Siano X ed Y due variabili aleatorie monodimensionali continue, le cui rispettive distribuzioni cumulative siano $F_X(x)$ e $F_Y(y)$. La coppia (X, Y) costituisce a sua volta una variabile aleatoria continua, questa volta bi-dimensionale. Per definizione, la distribuzione cumulativa di questa variabile aleatoria è la funzione

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

mentre la funzione densità di probabilità è la funzione

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

Vista al contrario, questa relazione può anche essere scritta nella forma

$$F_{XY}(x, y) = \int_{-\infty}^y \int_{-\infty}^x f_{XY}(x, y) dx dy$$

In questa espressione, è bene distinguere le variabili di integrazione dagli estremi di integrazione: indicando allora le variabili di integrazione con τ e t , possiamo riscrivere lo stesso integrale doppio nella forma

$$F_{XY}(x, y) = \int_{-\infty}^y \int_{-\infty}^x f_{XY}(t, \tau) dt d\tau$$

Diremo che le due variabili aleatorie X ed Y sono **indipendenti** nel caso in cui sussiste la relazione

$$F_{XY}(x, y) = F_X(x)F_Y(y)$$

Si dimostra inoltre che sussiste il seguente risultato:

$$P(x_1 \leq X \leq x_2 \cap -\infty \leq Y \leq y) = \int_{x_1}^{x_2} \int_{-\infty}^y f_{XY}(x, y) dx dy$$

Prendono il nome di **funzioni di distribuzione marginale** rispettivamente di X e di Y le seguenti due funzioni:

$$f_X(x) = \int_{-\infty}^{+\infty} f_{XY}(x, y) dy \qquad f_Y(y) = \int_{-\infty}^{+\infty} f_{XY}(x, y) dx$$

Supponiamo adesso che Z sia una variabile aleatoria ottenuta come funzione delle variabili aleatorie X ed Y : in particolare, supponiamo che sia $Z = g(X, Y)$. Sussiste il seguente risultato circa il calcolo della media di Z :

$$E[Z] = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} g(x, y) f_{XY}(x, y) dx dy$$

Diamo adesso alcune importanti definizioni:

- si definisce **momento congiunto, di ordine k, delle variabili X ed Y** la quantità $E[X^k \cdot Y^k]$, ossia la media del prodotto delle variabili aleatorie X ed Y ciascuna elevata ad un esponente di valore k; è intuitivo comprendere che tale momento congiunto vale

$$E[X^k \cdot Y^k] = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} x^k y^k f_{XY}(x, y) dx dy$$

- si definisce **correlazione di X ed Y** la quantità $E[XY]$; nel caso in cui X ed Y sono indipendenti, si dimostra che tale correlazione è data da

$$E[XY] = E[X]E[Y]$$

- si definisce **momento centrale di ordine k-n** delle variabili X ed Y la quantità

$$E[(X - E(X))^k (Y - E(Y))^n]$$

Nel caso in cui $k=n=1$, quella quantità diventa evidentemente

$$E[(X - E(X))(Y - E(Y))] = E[XY] - E[X]E[Y]$$

e prende il nome di **covarianza di X ed Y** (si indica con **cov(X,Y)**)

- si definisce **coefficiente di correlazione di X ed Y** (o anche **covarianza normalizzata**) la quantità

$$\rho_{XY} = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y} = \frac{E[XY] - E[X]E[Y]}{\sqrt{E[(X - E(X))^2] E[(Y - E(Y))^2]}}$$

E' facile verificare che, se X ed Y sono indipendenti, il loro coefficiente di correlazione risulta uguale a 0. In generale, invece, $\rho_{XY} \leq 1$

Funzione caratteristica di una variabile aleatoria

DEFINIZIONE NEL CASO CONTINUO

Consideriamo una generica variabile aleatoria X che sia continua: questo significa che essa possiede una *funzione densità di probabilità*, che per comodità indichiamo con $g_X(x)$. Sussiste la seguente definizione:

Def. Si definisce **funzione caratteristica di X** la funzione

$$\Phi_X(v) = E[e^{jvX}]$$

Si tratta cioè della media della variabile aleatoria e^{jvX} : dato che questa è ottenuta come funzione della variabile aleatoria di partenza X , possiamo senz'altro applicare il noto teorema sulla media di una funzione di una variabile aleatoria e scrivere che

$$\Phi_X(v) = E[e^{jvX}] = \int_{-\infty}^{+\infty} e^{jvx} g_X(x) dx$$

In quest'ultima relazione viene dunque indicato un preciso legame tra la funzione caratteristica di X e la funzione densità di X stessa. Possiamo caratterizzare ancora meglio questo legame, facendo uso del concetto di *trasformata di Fourier*. Infatti, in quell'integrale nessuno ci vieta di effettuare un cambio di variabile; in particolare, possiamo porre $x=f$: otteniamo

$$\Phi_X(v) = \int_{-\infty}^{+\infty} e^{jvf} g_X(f) df$$

Possiamo anche porre anche $v=2\pi t$, per cui

$$\Phi_X(2\pi t) = \int_{-\infty}^{+\infty} e^{j2\pi ft} g_X(f) df$$

Questa relazione indica evidentemente che

$$\Phi_X(2\pi t) \xleftrightarrow{\text{Fourier}} g_X(f)$$

ossia che *la funzione caratteristica di una variabile aleatoria può essere vista come l'antitrasformata di Fourier della densità di probabilità della variabile aleatoria stessa.*

APPLICAZIONE

Questo risultato è importante in quanto ci consente di utilizzare tutte le proprietà che noi conosciamo circa la trasformata di Fourier. Per esempio, in precedenza noi abbiamo dimostrato che, considerate due variabili aleatorie X_1 e X_2 indipendenti tra loro e considerata la loro somma $X = X_1 + X_2$, la densità di probabilità di X si ottiene mediante la formula

$$g_X(x) = g_{X_1}(x) * g_{X_2}(x)$$

ossia come prodotto di convoluzione della densità delle due variabili aleatorie di partenza.

La dimostrazione di questo risultato non è stata molto agevole, mentre invece possiamo farla in modo estremamente rapido sfruttando proprio la trasformata di Fourier e la definizione di funzione caratteristica.

Infatti, la funzione caratteristica di X è data per definizione da

$$\Phi_X(v) = E[e^{jvX}] = E[e^{jv(X_1+X_2)}] = E[e^{jvX_1} e^{jvX_2}]$$

Dato che X_1 e X_2 sono per ipotesi indipendenti tra loro, lo saranno anche le variabili e^{jvX_1} ed e^{jvX_2} . Allora, dato che la media del prodotto è pari al prodotto delle medie quando le due variabili sono indipendenti, possiamo scrivere che

$$\Phi_X(v) = E[e^{jvX_1}] \cdot E[e^{jvX_2}] = \Phi_{X_1}(v) \Phi_{X_2}(v)$$

A questo punto, possiamo trasformare secondo Fourier questa relazione: sfruttando la proprietà di convoluzione in frequenza (in base alla quale un prodotto nel dominio dei tempi equivale ad una convoluzione nel dominio trasformato), noi abbiamo che

$$\text{Fourier} [\Phi_X(v)] = \text{Fourier} [\Phi_{X_1}(v)] * \text{Fourier} [\Phi_{X_2}(v)]$$

e da qui scaturisce evidentemente che

$$g_X(x) = g_{X_1}(x) * g_{X_2}(x)$$

PROPRIETÀ: MOMENTO DEL K° ORDINE

Sia data sempre la generica variabile aleatoria X con densità di probabilità $g_X(x)$ e con funzione caratteristica $\Phi_X(v)$: è possibile dimostrare la relazione

$$E[X^k] = \frac{1}{j^k} \left. \frac{d^k \Phi_X(v)}{dv^k} \right|_{v=0}$$

Questa formula dice che, nota la funzione caratteristica, mediante una semplice operazione di doppia derivazione noi siamo in grado di calcolare i momenti di qualsiasi ordine della nostra variabile. Tra l'altro, in base a quella relazione, tali momenti differiscono solo del termine $1/j^k$.

Anziché verificare la relazione nel caso generale, facciamo nel caso semplice di $k=1$:

$$\begin{aligned} \left. \frac{1}{j} \frac{d\Phi_X(v)}{dv} \right|_{v=0} &= \frac{1}{j} \left[\frac{d}{dv} \left(\int_{-\infty}^{+\infty} e^{jvf} g_X(f) df \right) \right]_{v=0} = \frac{1}{j} \left[\left(\int_{-\infty}^{+\infty} \frac{d}{dv} (e^{jvf} g_X(f)) df \right) \right]_{v=0} = \\ &= \left[\int_{-\infty}^{+\infty} f e^{jvf} g_X(f) df \right]_{v=0} = \int_{-\infty}^{+\infty} f g_X(f) df = E[X] \end{aligned}$$

ESEMPIO: DISTRIBUZIONE ESPONENZIALE

Sia X una variabile aleatoria con distribuzione esponenziale: sappiamo che questo significa che la funzione densità di probabilità di X è

$$g_X(x) = \lambda e^{-\lambda x} \quad x > 0$$

Calcoliamo la funzione caratteristica di X : in base alla definizione, possiamo scrivere che

$$\Phi_X(v) = E[e^{jvX}] = \int_{-\infty}^{+\infty} e^{jvx} g_X(x) dx = \lambda \int_0^{+\infty} e^{jvx} e^{-\lambda x} dx = \lambda \int_0^{+\infty} e^{-x(\lambda - jv)} dx = \frac{\lambda}{\lambda - jv} [e^{-x(\lambda - jv)}]_0^{+\infty} = \frac{\lambda}{\lambda - jv}$$

Adesso calcoliamo il generico momento del k° ordine: per fare questo calcolo possiamo sfruttare il risultato trovato al paragrafo precedente, ma possiamo anche sfruttare la particolare forma della funzione caratteristica appena trovata. Vediamo cosa succede con questa seconda strada.

La funzione caratteristica può essere scritta nel modo seguente:

$$\Phi_X(v) = \frac{1}{1 - j \frac{v}{\lambda}}$$

Nell'ipotesi che $|v/\lambda| < 1$, quella è la somma della serie geometrica, per cui possiamo scrivere che

$$\Phi_X(v) = \sum_{k=0}^{\infty} \left(\frac{jv}{\lambda} \right)^k$$

D'altro canto, essendo $\Phi_X(v) = E[e^{jvX}]$, noi possiamo sviluppare in serie quell'esponenziale e scrivere quanto segue:

$$\Phi_X(v) = E[e^{jvX}] = E \left[1 + jvX + \dots + \frac{j^n}{n!} (vX)^n + \dots \right]$$

Usando la linearità della media, quella diventa

$$\Phi_X(v) = E[1] + E[jvX] + \dots + E \left[\frac{j^n}{n!} (vX)^n \right] + \dots = 1 + jvE[X] + \dots + \frac{j^n}{n!} v^n E[X^n] + \dots$$

e possiamo quindi scrivere che

$$\Phi_X(v) = \sum_{k=0}^{\infty} \frac{j^k}{k!} v^k E[X^k]$$

Uguagliando adesso

$$\Phi_X(v) = \sum_{k=0}^{\infty} \left(\frac{jv}{\lambda} \right)^k = \sum_{k=0}^{\infty} \frac{j^k}{k!} v^k E[X^k]$$

si deduce evidentemente che

$$\left(\frac{jv}{\lambda} \right)^k = \frac{j^k}{k!} v^k E[X^k]$$

da cui concludiamo che $E[X^k] = \frac{k!}{\lambda^k}$.

FUNZIONE CARATTERISTICA DI UNA VARIABILE ALEATORIA DISCRETA

La funzione caratteristica di una variabile aleatoria X è stata definita mediante la relazione

$$\Phi_X(v) = E[e^{jvX}]$$

Questa definizione vale sia per quando X è continua sia quando è discreta. Naturalmente, nel caso di X continua, quella relazione, in base alla definizione di media, diventa

$$\Phi_X(v) = \int_{-\infty}^{+\infty} e^{jvf} g_X(t) dt$$

mentre, invece, se X è discreta, abbiamo che

$$\Phi_X(v) = \sum_i e^{jvx_i} P(x_i)$$

Densità di probabilità condizionata

Consideriamo un generico spazio di eventi S relativo ad un certo fenomeno e in particolare consideriamo due eventi A e B contenuti in S : sappiamo già che si definisce probabilità condizionata di A dato B la probabilità che si verifichi A dopo che si è verificato B e sappiamo che tale probabilità è valutabile mediante la formula

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

Adesso, consideriamo un generica variabile aleatoria X continua e con densità $f_X(x)$. Supponiamo anche che l'evento generico A corrisponda all'evento per cui $X \leq x$, dove x è un qualsiasi numero reale. La probabilità condizionata di prima diventa allora

$$F_X(x|B) = \frac{P(X \leq x \cap B)}{P(B)}$$

La funzione $F_X(x|B)$ prende il nome di **funzione di distribuzione condizionata** della variabile aleatoria X . Evidentemente, a seconda di quello a cui corrisponde l'evento B , tale funzione può assumere espressioni diverse.

Supponiamo che anche l'evento B sia legato alla variabile X e in particolare che sia $B = \{X \leq a\}$: abbiamo allora che

$$F_X(x|X \leq a) = F_X(x|a) = \frac{P(X \leq x \cap X \leq a)}{P(X \leq a)} = \frac{P(X \leq \min(x, a))}{P(X \leq a)}$$

Da qui si deduce evidentemente che

$$\begin{aligned} \text{se } a \leq x &\longrightarrow F_X(x|a) = \frac{P(X \leq a)}{P(X \leq a)} = 1 \\ \text{se } a > x &\longrightarrow F_X(x|a) = \frac{P(X \leq x)}{P(X \leq a)} \end{aligned}$$

Consideriamo adesso un'altra variabile aleatoria Y e supponiamo che sia $B = \{Y \leq y\}$: in questo caso, abbiamo che

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

Supponiamo adesso che sia $B = \{Y = y\}$: in questo caso, noi sappiamo che $P(B)=0$ in quanto B si riduce ad un solo punto, per cui non siamo in condizioni di applicare la formula

$$P(x|B) = \frac{P(X \leq x \cap B)}{P(B)}$$

Possiamo tuttavia adottare qualche accorgimento per valutare qualcosa di molto simile; in particolare, si pone per definizione

$$F_{X|Y}(x, y) = F_X(x|Y = y) = \lim_{\Delta y \rightarrow 0} F_X(X \leq x | y \leq Y \leq y + \Delta y)$$

Con questo procedimento al limite, noi possiamo fare una serie di conti che ci portino ad un risultato importante: infatti, sempre mediante la formula delle probabilità condizionate, possiamo scrivere che

$$F_{X|Y}(x, y) = \lim_{\Delta y \rightarrow 0} F_X(X \leq x | y \leq Y \leq y + \Delta y) = \lim_{\Delta y \rightarrow 0} \frac{P(X \leq x \cap y \leq Y \leq y + \Delta y)}{P(y \leq Y \leq y + \Delta y)}$$

Usando ora i concetti di distribuzione congiunta di probabilità e di distribuzione di probabilità di una variabile aleatoria, abbiamo che

$$F_{X|Y}(x, y) = \lim_{\Delta y \rightarrow 0} \frac{\int_{-\infty}^{+\infty} \int_y^{y+\Delta y} f_{X,Y}(x_1, y_1) dx_1 dy_1}{\int_y^{y+\Delta y} f_Y(y_1) dy_1}$$

L'integrale a denominatore, dato che Δy si ritiene piccolissimo, può essere risolto immediatamente nel modo (approssimato) seguente:

$$F_{X|Y}(x, y) = \lim_{\Delta y \rightarrow 0} \frac{\int_{-\infty}^{+\infty} \int_y^{y+\Delta y} f_{X,Y}(x_1, y_1) dx_1 dy_1}{f_Y(y) \Delta y}$$

Le stesse considerazioni si possono fare per risolvere l'integrale più interno dell'integrale doppio a numeratore: abbiamo così che

$$F_{X|Y}(x, y) = \lim_{\Delta y \rightarrow 0} \frac{\int_{-\infty}^{+\infty} f_{X,Y}(x_1, y) \Delta y dx_1}{f_Y(y) \Delta y} = \lim_{\Delta y \rightarrow 0} \frac{\int_{-\infty}^{+\infty} f_{X,Y}(x_1, y) dx_1}{f_Y(y)}$$

Il fatto che il Δy sia scomparso ci consente di eliminare il limite, per cui

$$F_{X|Y}(x, y) = \frac{\int_{-\infty}^{+\infty} f_{X,Y}(x_1, y) dx_1}{f_Y(y)}$$

Ogni eventuale passaggio matematico su questa formula si può fare solo conoscendo come è fatta la funzione $f_{X,Y}(x_1, y)$ di densità congiunta di X e di Y .

Inoltre, se deriviamo quella espressione, otteniamo

$$f_{x|Y}(x, y) = \frac{d}{dx} F_{x|Y}(x, y) = \frac{f_{X,Y}(x, y)}{f_Y(y)}$$

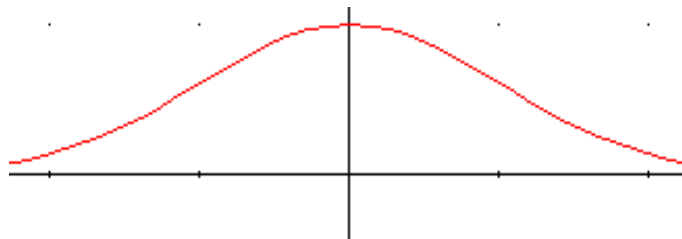
Variabili aleatorie con distribuzione gaussiana

DEFINIZIONE DI DISTRIBUZIONE GAUSSIANA

Si dice che una variabile aleatoria continua X ha una **distribuzione gaussiana con media μ e deviazione standard σ** quando la sua funzione densità di probabilità è

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

L'andamento grafico di questa funzione, supposta $\mu=0$, è il seguente:



Questa specie di campana è tanto più larga quanto maggiore è la varianza.

FUNZIONE CARATTERISTICA

Calcoliamo la funzione caratteristica di X : la definizione di funzione caratteristica, nel caso continuo, dice che

$$\Phi_X(v) = E[e^{jvX}] = \int_{-\infty}^{+\infty} e^{jvx} f_X(x) dx$$

Tuttavia, data la complessità dell'espressione di $f_X(x)$, è consigliabile usare un altro metodo e precisamente il legame che esiste tra quella funzione caratteristica e la trasformata di Fourier: sappiamo infatti che sussiste la relazione

$$\Phi_X(2\pi t) \xleftarrow{\text{Fourier}} f_X(x)$$

per cui il nostro problema diventa il calcolo dell'antitrasformata di $f_X(x)$.

Se applicassimo direttamente la definizione di antitrasformata, otterremmo

$$\begin{aligned} \Phi_X(2\pi t) &= \int_{-\infty}^{+\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} e^{j2\pi xt} dx = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{(x-\mu)^2}{2\sigma^2}} e^{j2\pi xt} dx = \\ &= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{x^2-2\mu x}{2\sigma^2}} e^{-\frac{\mu^2}{2\sigma^2}} e^{j2\pi xt} dx = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{\mu^2}{2\sigma^2}} \int_{-\infty}^{+\infty} e^{-\frac{x^2-2\mu x}{2\sigma^2}} e^{j2\pi xt} dx = \\ &= \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{\mu^2}{2\sigma^2}} \int_{-\infty}^{+\infty} e^{-\frac{x^2}{2\sigma^2}} e^{\left(j2\pi t + \frac{\mu}{\sigma^2}\right)x} dx = \dots \end{aligned}$$

Senza scendere in troppi dettagli, il risultato finale risulta essere

$$\Phi_X(v) = e^{j\mu v - v^2 \frac{\sigma^2}{2}}$$

FUNZIONE DENSITÀ DI PROBABILITÀ CONGIUNTA

Supponiamo adesso che X ed Y siano due variabili aleatorie continue, entrambe con distribuzione gaussiana, la prima con media μ_X e varianza σ_X^2 e la seconda con media μ_Y e varianza σ_Y^2 . E' possibile dimostrare che la funzione densità di probabilità congiunta di tali variabili aleatorie è la seguente:

$$f_{X,Y}(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)} \left[\frac{(x-\mu_X)^2}{\sigma_X^2} + \frac{(y-\mu_Y)^2}{\sigma_Y^2} - \frac{2\rho(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y} \right]}$$

dove

$$\rho = \frac{E[(X - E(X))(Y - E(Y))]}{\sigma_X\sigma_Y}$$

è il **coefficiente di correlazione** di X ed Y.

E' subito ovvio che quella espressione si semplifica notevolmente quando sia X sia Y hanno media nulla e deviazione standard unitaria: infatti, ponendo

$$\begin{cases} \mu_X = \mu_Y = 0 \\ \sigma_X = \sigma_Y = 1 \end{cases}$$

si ottiene

$$f_{X,Y}(x,y) = \frac{1}{2\pi\sqrt{1-\rho^2}} e^{\frac{1}{2(1-\rho^2)}[x^2+y^2-2\rho xy]}$$

INDIPENDENZA DI DUE VARIABILI ALEATORIE CON DISTRIBUZIONE GAUSSIANA

Sappiamo che due generiche variabili aleatorie continue X ed Y , con funzione densità rispettivamente $f_X(x)$ e $f_Y(y)$, si dicono indipendenti quando la funzione di densità congiunta è pari al prodotto delle funzioni densità, ossia quando

$$f_{X,Y}(x,y) = f_X(x)f_Y(y)$$

E' possibile dimostrare che, se X ed Y hanno entrambe distribuzione gaussiana, quella condizione avviene quando $\rho = 0$, ossia quando le due variabili sono *incollegate*.

FUNZIONE DENSITÀ DI PROBABILITÀ CONDIZIONATA

Siano sempre X ed Y due variabili aleatorie entrambe con distribuzione gaussiana. Abbiamo in precedenza definito col nome di funzione densità di probabilità condizionata la funzione

$$f_{X|Y}(x,y) = \frac{d}{dx} F_{X|Y}(x,y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$$

dove

$$F_{X|Y}(x,y) = F_X(X \leq x | Y = y) = \lim_{\Delta y \rightarrow 0} F_X(X \leq x | y \leq Y \leq y + \Delta y)$$

Allora, è possibile dimostrare che anche la variabile aleatoria $Z = X|Y$, la cui funzione densità è $f_{X|Y}(x,y)$, ha distribuzione gaussiana.

Anzi, se μ_X e σ_X^2 sono media e varianza di X e μ_Y e σ_Y^2 sono media e varianza di Y , è possibile dimostrare che la media e la varianza di Z sono le seguenti:

$$\begin{cases} \mu_Z = \mu_X + \rho \frac{\sigma_X}{\sigma_Y} (y - \mu_Y) \\ \sigma_Z = \sigma_X \sqrt{1 - \rho^2} \end{cases}$$

COMBINAZIONE LINEARE DI VARIABILI ALEATORIE CON DISTRIBUZIONE GAUSSIANA

Sia X una variabile aleatoria continua avente distribuzione gaussiana con media μ_X e varianza σ_X^2 . Consideriamo inoltre la variabile aleatoria $Y = aX + b$, dove a e b sono due qualsiasi numeri reali.

Noi abbiamo in precedenza dimostrato che, se Y è definita in quel modo, la sua funzione densità di probabilità, a prescindere dalla natura di X , è

$$f_Y(y) = \frac{1}{|a|} f_X\left(\frac{y-b}{a}\right)$$

In questo caso particolare, questa formula ci dice che, essendo X gaussiana, anche Y risulta essere gaussiana: anzi, sapendo che

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

noi possiamo subito scrivere che

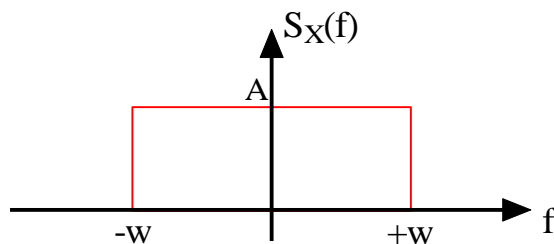
$$f_Y(y) = \frac{1}{|a|} \frac{1}{\sigma_X \sqrt{2\pi}} e^{-\frac{\left(\frac{y-b}{a} - \mu_X\right)^2}{2\sigma_X^2}}$$

Conseguenza

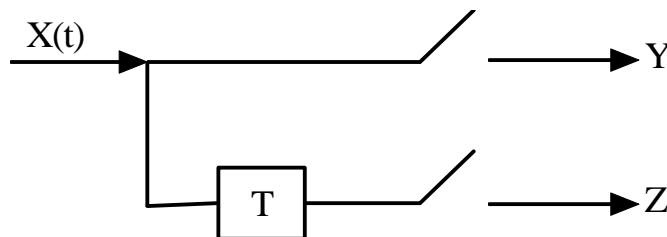
Conseguenza di questa proprietà è che, se X ed Y sono due variabili aleatorie entrambe con distribuzione gaussiana, risulta gaussiana anche la variabile aleatoria $Z = aX + bY$.

ESEMPIO

Sia $X(t)$ una variabile aleatoria con distribuzione gaussiana e con spettro di potenza fatto nel modo seguente:



Supponiamo infine che la $X(t)$ venga posta in ingresso ad un sistema fatto nel modo seguente:



Come funziona questo sistema? Ci sono quei due interruttori che noi supponiamo essere sincroni: ciò significa che essi si chiudono e si aprono nello stesso momento. Allora, supponiamo che essi siano inizialmente aperti e che, in un certo istante t_0 , si chiudano e si riaprono istantaneamente, in modo cioè che Y coincida con $X(t_0)$, mentre Z coincida con $X(t_0+T)$.

In pratica, quindi, il sistema preleva il valore assunto da $X(t)$ in un istante prefissato t_0 , assegnando tale valore a Y , e poi preleva il valore assunto t_0+T , assegnando tale valore a Z .

Per comodità di ragionamento, facciamo due ipotesi semplificative:

- in primo luogo, prendiamo $t_0=0$, per cui

$$\begin{cases} Y = X(0) \\ Z = X(T) \end{cases}$$

- in secondo luogo, supponiamo che il processo da cui X è stata estratta, oltre ad essere gaussiano, sia anche almeno stazionario in senso lato.

Allora, noi vogliamo determinare, se esiste, il valore del ritardo T tale che le variabili di uscita Y e Z siano indipendenti.

E' subito evidente che, essendo $X(t)$ gaussiana, sono anche gaussiane $Z=X(T)$ e $Y=X(0)$. Di conseguenza, possiamo utilizzare un risultato citato in precedenza e dire che Z ed Y risultano indipendenti se risultano incorrelate, ossia se accade che

$$\rho = \frac{E[(Y - E(Y))(Z - E(Z))]}{\sigma_Y \sigma_Z} = 0$$

Dato che $Y=X(0)$ e $Z=X(T)$, il coefficiente di correlazione può anche essere espresso nella forma

$$\rho = \frac{E[(X(0) - E(X(0)))(X(T) - E(X(T)))]}{\sigma_Y \sigma_Z}$$

Indicando con μ_X la media di $X(t)$, possiamo scrivere che

$$\rho = \frac{E[(X(0) - \mu_X)(X(T) - \mu_X)]}{\sigma_Y \sigma_Z}$$

Per comodità facciamo una terza ipotesi semplificativa: supponiamo che $\mu_X=0$, per cui abbiamo che

$$\rho = \frac{E[X(0)X(T)]}{\sigma_Y \sigma_Z}$$

Adesso, il numeratore di quella frazione non è altro che la funzione di autocorrelazione di X calcolata per un intervallo di tempo pari a T : quindi

$$\rho = \frac{R_X(T)}{\sigma_Y \sigma_Z}$$

Dobbiamo trovare il valore di T tale che $\rho=0$ e, in base a quella relazione, ciò equivale a trovare il valore di T per il quale $R_X(T)=0$. Ci serve perciò l'espressione analitica di $R_X(t)$, che possiamo calcolare in quanto sappiamo come è fatta la sua trasformata di Fourier $S_X(f)$: sapendo infatti che

$$S_X(f) = A \operatorname{rect}\left(\frac{f}{2W}\right)$$

deduciamo che

$$R_X(t) = \frac{A}{2W} \operatorname{sinc}(2Wt)$$

Ponendo $t=T$, noi dobbiamo risolvere dunque l'equazione

$$\frac{A}{2W} \operatorname{sinc}(2WT) = 0$$

Le soluzioni di questa equazione sono i valori di T tali che

$$2\pi WT = k\pi$$

per cui concludiamo che i valori di T che ci interessano sono

$$T = \frac{k\pi}{2\pi W} \quad k = 1, 2, \dots, \infty$$

Autore: **SANDRO PETRIZZELLI**
e-mail: sandry@iol.it
sito personale: <http://users.iol.it/sandry>
succursale: <http://digilander.iol.it/sandry1>