ROMA
TRE
Università degli studi di Roma Tre
Facoltà di Ingegneria

Appunti dalle lezioni del corso di

Teoria dei Segnali

TEORIA DELLE PROBABILITA’

Prof. Alessandro NERI

Roma, novembre 2001

INDICE

Teoria delle probabilità

INTRODUZIONE

Il seguente capitolo è dedicato all'introduzione dei concetti fondamentali della teoria della probabilità. Tale teoria rappresenta lo strumento analitico utilizzato per studiare quei fenomeni in cui si presentano elementi di incertezza " a priori" che rendono imprevedibili, anche se solo parzialmente, le determinazioni di grandezze fisiche o attributi qualitativi di un dato fenomeno prima che esso si sia verificato, e che prendono il nome di fenomeni aleatori.
Il capitolo è strutturato come segue:

- nel paragrafo 2 viene introdotto, su base intuitiva, il concetto di probabilità
- nel paragrafo 3 viene definita la terminologia di base
- nel paragrafo 4 viene definito il concetto di spazio di probabilità.

IL CONCETTO DI PROBABILITÀ

Come accennato nell'introduzione la teoria delle probabilità nasce dall'esigenza di trattare in un qualche modo i fenomeni aleatori, ovvero quei fenomeni che presentano caratteristiche di imprevedibilità. E' da notare che per sua natura, l'aleatorietà non costituisce una caratteristica intrinseca del fenomeno, ma è legata al nostro grado di conoscenza a priori circa i fattori che condizionano il fenomeno, nonché alla sua presentazione all'osservatore. Cause di aleatorietà possono essere ad esempio:

- l'ignoranza completa o parziale delle leggi che regolano il fenomeno

- l’ignoranza completa o parziale delle condizioni iniziali e degli ingressi
- difficoltà pratica nel misurare correttamente i parametri d’interesse anche quando sono note le leggi e le condizioni iniziali
- impossibilità pratica di individuare completamente un fenomeno di cui si osserva solo una manifestazione.

La Teoria della probabilità è nata per riuscire a descrivere comportamenti medi di fenomeni di massa, che possono presentarsi sia sequenzialmente che simultaneamente, per i quali sia possibile o non sia conveniente definire un modello matematico completo ed esatto come ad esempio:

- l’emissione degli elettroni
- le chiamate telefoniche
- le rivelazioni di un radar
- i guasti di un sistema
- il controllo di qualità
- la meccanica statistica
- le turbolenze
- il rumore
- le code
- …..

Il punto di partenza di tale teoria è costituito dall’osservazione(sperimentale) che, qualora si ha a che fare con fenomeni di massa, alcuni valori medi tendono a valori costanti all’aumentare del numero delle osservazioni e tali valori rimangono gli stessi anche se vengono calcolati su una qualche sottosequenza ottenuta da quella originale con modalità specificate, però, prima dell’esperimento. Ad esempio, se si considera il lancio di moneta truccata, la percentuale di volte in cui il risultato è “testa” tende a 0.5 anche se si considerano i risultati ogni 10 lanci.

Lo scopo della teoria delle probabilità è quello di descrivere e di predire tali valori medi in termini di probabilità di eventi.

Tale teoria potrebbe essere sviluppata su base intuitiva mettendo in relazione il concetto di probabilità con la frequenza di occorrenza relativa. In tal caso la probabilità di un evento può essere definita come il limite:

$$\Pr\{E\} = \lim_{N \to \infty} \frac{N_\varepsilon}{N} \tag{0.1}$$

essendo N_ε il numero di occorrenze dell'evento E ed N_ε il numero totale di priove (impostazione frequentistica). Anche se semplice concettualmente questo approccio non consente una trattazione rigorosa di molti problemi. Inoltre, ammesso che il limite esista, in tale definizione non c'è distinzione precisa tra il concetto di probabilità ed il modo in cui tale grandezza deve essere misurata. E' come se definissimo la resistenza R come il limite:

$$R = \lim_{N \to \infty} \frac{V(t)}{i_N(t)}, \tag{0.2}$$

dove V(t) è la tensione d'ingresso e $i_N(t)$ le correnti di una sequenza di resistori che tendono in un qualche modo al bipolo ideale.

Per tali ragioni la moderna teoria delle probabilità si basa su una impostazione più rigorosa di tipo assiomatico. Assumendo quindi che le probabilità soddisfino determinati assiomi, tale impostazione consente di determinare con un ragionamento deduttivo, a partire dalle probabilità $Pr\{A_1\}$ di certi eventi A_1, supposte note, le probabilità $Pr\{B_1\}$di altri eventi B_1.

In tale contesto la probabilità di un evento rappresenta un numero assegnato a detto evento come la massa è assegnata ad un corpo e la resistenza ad un resistore. Il collegamento tra la teoria e l'evidenza sperimentale é garantito da una serie di teoremi che costituiscono le *leggi dei grandi numeri* che definiscono sotto quali condizioni e in che termini la frequenza relativa, osservata sperimentalmente , converga verso la probabilità dell'evento.

DEFINIZIONI

Prima di introdurre in modo formale la teoria assiomatica delle probabilità è utile richiamare la terminologia a cui si farà riferimento nel seguito.

Esperimento	descrizione delle modalità di attuazione di un fenomeno
Prova	Attuazione di un esperimento
Determinazione	valore che può essere assunto da una grandezza fisica o da un attributo qualitativo a seguito di una prova
Risultato	n-upla ordinata delle determinazioni assunte dalle grandezze fisiche o degli attributi qualitativi descriventi il fenomeno a seguito di una prova
Serie aleatoria	insieme di tutti i possibili diversi risultati
Risultati Incompatibili	due risultati si dicono incompatibili se il verificarsi di uno esclude il verificarsi dell'altro
Evento	attributo (logico) di un risultato (ad es. "il numero estratto è un numero pari"). L'evento si verifica se il risultato soddisfa l'attributo
Risultati Favorevoli	Risultati per cui un evento si verifica
Evento Semplice	evento a cui è favorevole un solo risultato
Evento Composto	evento a cui sono favorevoli più risultati

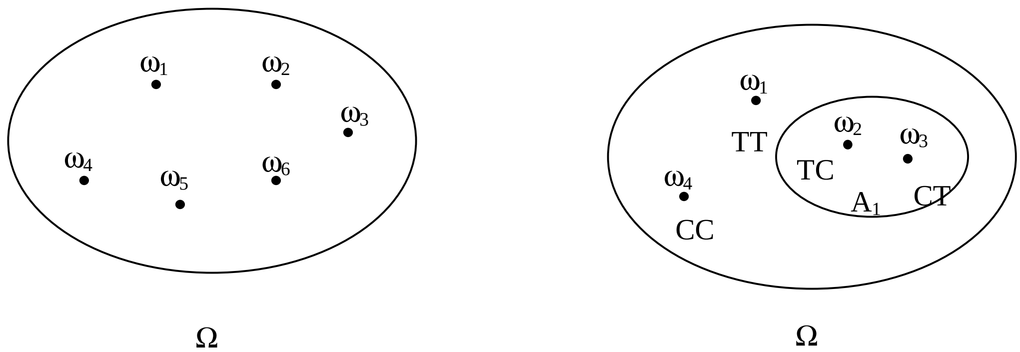
Eventi Compatibili	eventi che hanno in comune almeno un risultato (ad es. "il numero che compare a seguito del lancio di un dado è un numero pari"e "il numero è un multiplo di tre")
Eventi Incompatibili o Mutuamente Escludentesi	eventi che non hanno in comune alcun risultato (ad es. "il numero che compare a seguito del lancio di un dado è inferiore a tre")
Evento Certo	evento che si verifica sempre (ad es. "a seguito del lancio di una moneta esce o testa o croce")
Evento Impossibile	evento che non si verifica mai

TEORIA ASSIOMATICA

La teoria delle probabilità secondo questa impostazione altri non è che una applicazione particolare della teoria della misura secondo Lebesgue su insiemi di punti di uno spazio astratto. La procedura per l'individuazione degli enti matematici di base è la seguente.

Iniziamo con il considerare il caso semplice in cui il numero dei possibili risultati è finito, ad esempio pari ad N.

Ad ognuno dei risultati possibili possiamo associare, con una corrispondenza biunivoca, un punto di uno spazio astratto Ω. Ad esempio nel caso del lancio di un dado possiamo considerare una situazione come quella di fig.1.1.a.



Lo spazio Ω prende il nome di SPAZIO BASE o SPAZIO DEI RISULTATI. Dato un evento *E*, ad esso corrisponderà l'insieme E dei punti di Ω corrispondenti ai risultati favorevoli all'evento *E*:

$$\mathit{E} \Leftrightarrow \mathrm{E}$$

Ad esempio all'evento"su due lanci di una moneta esce una sola volta testa"corrisponde l'insieme A_1 di fig. 1.b.

Poiché all'evento certo sono favorevoli tutti i risultati, occorre definire quale sia la classe degli per cui si può definire una probabilità. Nel caso di numero finito dei possibili risultati, tale classe è costituita dagli eventi corrispondenti a tutti i possibili sottoinsiemi di Ω, che risultano pari a 2^N. L'insieme vuoto $\varnothing$ corrisponde l'insieme impossibile.

Ora, com'è noto, sugli insiemi di punti di Ω si possono definire le tre operazioni:

1) complementazione
2) unione
3) intersezione

Tali operazioni possono essere messe in relazione con le operazioni elementari che consentono di definire eventi a partire da altri eventi come segue:

1) complementazione

Dato un insieme A corrispondente all'evento *A*, al suo complementare $\overline{A}$:

$$\overline{A} = \{\omega : \omega \in \Omega, \omega \notin A\}$$

corrisponde l'evento (*non A*):

$$\{nonA\} \Leftrightarrow \overline{\mathrm{A}}$$

che si verifica quando non si verifica *A*.

2) unione

Dati due insiemi A e B corrispondenti agli eventi *A* e *B* alla loro unione:

$$A \bigcup B = \{\omega : \omega \in \Omega, \omega \in A \; o \; \omega \in B\}$$

corrisponde l'evento (*A o B*):

$$(AoB) \Leftrightarrow \mathrm{A} \bigcup \mathrm{B}$$

che si verifica quando si verifica almeno uno dei due eventi *A* o *B*.

3) *intersezione*

Dati due insiemi *A* e B corrispondenti agli eventi *A* e *B* alla loro intersezione:

$$A \bigcap B = \{\omega : \omega \in \Omega, \omega \in A \text{ e } \omega \in B\}$$

corrisponde l'evento (*A e B*):

$$(AeB) \Leftrightarrow \mathrm{A} \bigcap \mathrm{B}$$

che si verifica quando si verificano contemporaneamente i due eventi *A* e *B*.

Ad eventi mutuamente escludentesi corrispondono insiemi disgiunti e pertanto:

$$A \bigcap B = \emptyset$$

Dato quindi uno spazio base Ω costituito da un numero finito di punti ed indicata con F la classe dei sottoinsiemi costituiti dai punti di Ω, si ha che essa è rappresentativa di tutti i possibili eventi e quindi la probabilità *P* può essere definita come una funzione d'insieme avente come dominio la classe F. Affinché essa possa essere messa in relazione in un qualche modo (ad es. attraverso un procedimento di misura) con la frequenza relativa di un evento è inoltre ragionevole richiedere che il dominio della funzione sia l'intervallo [0,1]:

$$P : F \rightarrow [0,1]$$

A tale condizione vanno inoltre aggiunte due ulteriori condizioni di consistenza : la prima è che la probabilità dell'evento certo sia pari ad 1, mentre la seconda definisce le modalità secondo cui le probabilità di eventi possano essere calcolate a partire dalle probabilità dell'evento (*A* o *B*), in cui *A* e *B* sono due eventi incompatibili, sia la somma delle probabilità dei due eventi *A* e *B* .
Quanto precede può essere riassunto dicendo che data una funzione *P* definita su *F*, essa può essere considerata una funzione di probabilità se soddisfa le seguenti condizioni:

I) P(E) >0 $\quad \forall E \in F \qquad$ (1,8)

II) $P(E_1 \bigcup E_2) = P(E_1) + P(E_2) \quad \forall E_1, E_2 \in F : E_1 \bigcap E_2 = \varnothing \qquad$ (1,9)

III) $P(\bigcup_{i=1}^{M} E_i) = \sum_{i=1}^{M} P(E_i), \quad \forall E_1, E_2, \ldots E_M \in F : E_1 \bigcap E_2 = \varnothing \qquad$ (1,10)

IV) P(Ω)=1 $\qquad$ (1,11)

Ovvero la classe di insiemi *F* costituisce un' <u>algebra</u>

Esempio.

Con riferimento al lancio dei dadi a cui corrisponde lo spazio base di fig.1°a si ha che una funzione P che assegni a w_1 w_2 w_3 w_4 w_5 e w_6 rispettivamente i valori (0.2, 0.1, 0.1, 0.2, 0.2, 0.2) è una possibile funzione di probabilità mentre non lo sono (-0.2, -0.1, 0.1, 0.2, 0.2, 0.8) e (0.2, -0.2, 0.2, 0.2, 0.2, 0.2). Infatti la prima non è una funzione non negativa mentre la seconda non soddisfa la condizione di normalizzazione in quanto il suo valore calcolato su tutto Ω è pari a 1.2.

Nel caso in cui il numero di punti di Ω non sia finito, come ad esempio nel caso di sequenze di eventi (ad es. lancio ripetuto di un dado) o di eventi relativi a grandezze di tipo continuo (ad es. la tensione ai capi di un resistore) si potrebbe pensare di estendere quanto sopra considerando funzioni che siano state definite per tutti i sottoinsiemi di Ω e che soddisfino la condizione III estesa ad una successione qualsiasi di insiemi disgiunti costituita da 8un numero eventualmente infinito numerabile di elementi.

Sfortunatamente è noto dalla teoria della misura che una funzione reale che soddisfi tutte queste proprietà non esiste [De Julio, 1973 pag.21] , pertanto occorre o rinunciare a considerare tutti i sottoinsiemi di Ω o alla condizione III estesa come sopra. La strada scelta nella teoria assiomatica delle probabilità è quella di rinunciare a definire una probabilità per tutti i possibili sottoinsiemi di c e di estendere la condizione III solo ad un'infinità numerabile di elementi. Affinché però tale condizione abbia senso occorre che la classe di insiemi per cui si definisce una probabilità sia chiusa rispetto all'unione di una infinità numerabile di sottoinsiemi, indicata pertanto con F tale classe su di essa devono essere definite le 3 operazioni Complementazione ($\bar{\ }$), Unione ($\bigcup$), Intersezione ($\bigcap$) ed inoltre:

1) $Q \in F$ (1,16)

2) $se\ \mathrm{A} \in F \Rightarrow \overline{\mathrm{A}} \in F$ (1,17)

3) $se\ \mathrm{A}, \mathrm{B} \in F \Rightarrow (A \bigcup B) \in F$ (1,18)

4) $se\ \mathrm{A}, \mathrm{B} \in F \Rightarrow (A \bigcap B) \in F$ (1,19)

5) $se\ \mathrm{A}_{\mathrm{n}} \in F \Rightarrow \bigcup_{n=1}^{\infty} \mathrm{A}_{\mathrm{n}} \in F (\bigcap_{n=1}^{\infty} \mathrm{A}_{\mathrm{n}} \in F)$ (1,20)

Ovvero la classe di insiemi *F* costituisce una σ -algebra.

(Si noti che non tutte le condizioni precedenti sono necessarie per definire una σ -algebra in quanto alcune di esse sono deducibili dalle altre).

Se quindi *F* è una σ -algebra, è possibile definire su *F* una funzione d'insieme reale non negativa completamente additiva ovvero che gode delle seguenti proprietà:

I) P(E) >0 $\forall E \in F$ (1,21)

II) $P(E_1 \cup E_2) = P(E_1) + P(E_2)$ $\forall E_1, E_2 \in F : E_1 \bigcap E_2 = \varnothing$ (1,22)

III) $P(\bigcup_{i=1}^{M} E_i) = \sum_{i=1}^{M} P(E_i),$ $\forall E_1, E_2, ...E_M \in F : E_1 \bigcap E_2 = \varnothing$ (1,23)

IV) P(Ω)=1 (1,24)

Tale funzione prende il nome di *MISURA DI PROBABILITA'*.

Pertanto nella teoria delle probabilità un fenomeno aleatorio è descritto per mezzo dei seguenti concetti:

- uno spazio dei risultati Ω
- una classe di eventi F costituenti una σ-algebra
- le una misura di probabilità

mentre la proprietà (21), (22), (23) e (24) costituiscono gli *ASSIOMI* di tale teoria. Ciò viene sintetizzato dicendo che ad un fenomeno aleatorio è associato uno spazio di probabilità costituito dalla terna (Ω, *F*, *P*).

Si dirà *MISURABILE* ogni elemento di *F* e *AMMISSIBILE* l'evento corrispondente.

Per chiarire meglio il caso di utilizzo più frequente che corrisponde ad assumere lo spazio euclideo come spazio base. In particolare consideriamo il caso in cui $\Omega = R$. In tal caso si assume in generale come σ-algebra su cui costruire la misura la classe di insiemi generata a partire dagli intervalli illimitati inferiormente $(-\infty, a]$ tramite le operazioni di complementazione, unione e intersezione. Tale σ-algebra è così rilevante da avere acquistato un proprio nome: CAMPO DI BOREL e verrà pertanto indicata nel seguito con $F_{o.}$

E' facile verificare che la classe degli intervalli illimitati inferiormente non è chiusa rispetto all'operazione di complementazione, infatti $\overline{(-\infty, a]} = (a, \infty)$ e quindi di per sé non costituisce una σ-algebra. Utilizzando le operazioni di complementazione, unione e intersezione, a partire da tali intervalli è possibile generare altri intervalli aperti, chiusi, semiaperti nonché punti. Infatti posto:

$$A_1 = (-\infty, a_1] \ , \ A_2 = (-\infty, a_2] \qquad \text{con } a_1 < a_2$$

si ha:

1) $\overline{A_1} = \overline{(-\infty,\ a_1]} = (a_1, \infty\ )$
2) $\overline{A_1} \cap A = (a_1, \infty) \cap (-\infty, a_2] = (a_1, a_2\ \ ]$

3) $\bigcap_{k=1}^{\infty}(b-1/K,b]=[b,b]$ Cioè il punto b

4) $(b,c]\bigcup[b,b]=[b,c]$

5) $(b,\infty)\bigcup[b,b]=[b,\infty)$

Per definizione il *CAMPO DI BOREL* F_B è la più piccola σ-algebra che contiene la classe degli intervalli illimitati inferiormente. Quindi il campo di Borel è composto da tutti quei sottoinsiemi dello spazio euclideo che possono avere un qualche interesse nel descrivere un problema di probabilità con Ω.

<u>Teoremi fondamentali</u>

<u>Teorema 1.1</u> (teorema della probabilità dell'evento complementare)
Dato un evento *E* ammissibile con probabilità *Pr*{*E*}, la probabilità *Pr*{*non E*} dell'evento complementare (*non E*) è pari a:

Pr{*non E*}= 1- *Pr*{*E*} (1,25)

Prova- Poiché è misurabile $E \Leftrightarrow \mathrm{E}$ lo è anche $E \Leftrightarrow non\,\mathrm{E}$ ed inoltre poiché:

$E \bigcup \overline{E} = \Omega$ (1,26)

$E \bigcap \overline{E} = \varnothing$ (1,26)

In base al II e al III assioma si ha:

$P(\Omega) = P(E \bigcup \overline{E}) = P(E) + P(\overline{E}) = 1$ (1,28)

Ovvero

$$P(\overline{E}) = 1 - P(E) \qquad (1,29)$$

Che è appunto la (25) c.v.d..

<u>*Corollario*</u>1.1 La probabilità dell'evento impossibile è nulla:

$$P(\varnothing) = 0 \qquad (1,30)$$

<u>*Prova*</u> – Poiché Ω=$\varnothing$ e $P(\Omega) = 1$ dalla (25) segue immediatamente la (30)c.v.d.

<u>*Teorema*</u> 1.2 (Teorema delle probabilità totali)

Dati due eventi *A* e *B* ammissibili non necessariamente mutuamente escludentesi, l'evento (*A o B*) è un evento ammissibile e la sua probabilità vale:

$$Pr\{A\ o\ B\}=Pr\{A\}+ Pr\{B\}- Pr\{AeB\} \qquad (1,31a)$$

Ovvero in termini di insiemi

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \qquad (1,31b)$$

<u>*Prova*</u>- Come è facile verificare con l'aiuto della Fig.2 l'insieme $A \cup B$ può essere espresso come segue:

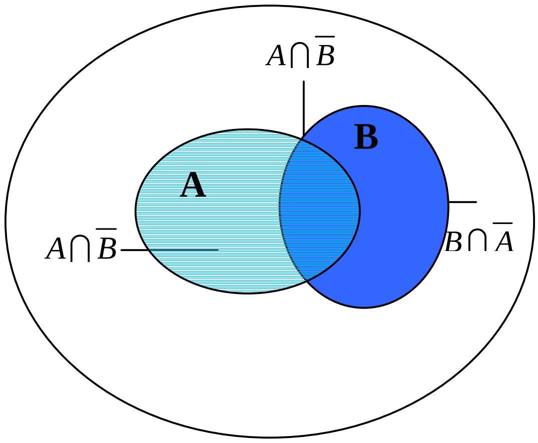


Fig.2

Per cui essendo i tre insiemi disgiunti si ha :

$$P(A\bigcup B)=P(A\bigcap\overline{B})+P(\overline{A}\bigcap B)+P(A\bigcap B) \quad (1,31b)$$

Ma in virtù del fatto che

$$A=\left(A\bigcap B\right)\bigcup\left(A\bigcap\overline{B}\right)e \quad B=\left(A\bigcap B\right)\bigcup\left(\overline{B}\bigcap A\right) \quad (1,31b)$$

Per il II assioma si ha immediatamente:

$$P(A)=P(A\bigcap B)+P(A\bigcap\overline{B}) \quad (1,34a)$$

$$P(B)=P(A\bigcap B)+P(\overline{A}\bigcap B) \quad (1,34b)$$

Da cui si ha:

$$P(A\bigcap\overline{B})=P(A)-P(A\bigcap B) \quad (1,35a)$$

$$P(\overline{A}\bigcap B)=P(B)-P(A\bigcap B) \quad (1,35b)$$

E sostituendo le (35) nelle (33) si ottiene la (31) c.v.d..

Commento - Il teorema asserisce che la misura di probabilità dell'unione dei due insiemi A e B è pari alla somma delle misura di A e di B diminuita della misura dell'insieme intersezione che, come è facile verificare dalla Fig.2 verrebbe altrimenti portato in conto due volte.

EVENTO CONDIZIONATO

Dato un certo fenomeno aleatorio sia (Ω_F, F_F, P_F) lo spazio di probabilità ad esso associato. Consideriamo ora un evento ammissibile Γ corrispondente ad un insieme $C \in F_F$ e definiamo un nuovo fenomeno aleatorio ottenuto da quello di partenza scartando i casi in cui l'evento Γ non si verifica. Per esempio nel caso del lancio di un dado Γ potrebbe essere l'evento "esce un numero pari"; in tal caso i risultati per cui tale evento si verifica sono quelli dello spazio Ω_C{2, 4, 6} e quindi nello svolgimento di un esperimento tutte le prove che diano luogo a risultati vanno considerate come non effettuate.

Il nuovo fenomeno aleatorio prende il nome di *fenomeno aleatorio condizionato all'evento Γ*.

Se ora si considera un evento $E \Leftrightarrow \mathrm{E}$ poiché occorre scartare le prove che danno luogo a risultati per cui Γ non si verifica, rispetto al fenomeno aleatorio condizionato esso sarà rappresentato dall'insieme $\mathrm{E} \cap \mathrm{C}$.

Pertanto assumendo $\Omega_c = C$ come spazio base la σ -algebra rispetto a cui definire la misura di probabilità del nuovo fenomeno aleatorio potrà essere quella ottenuta da F_F tramite l'intersezione con l'insieme C e che indicheremo con $F_{F/\Gamma}$. A questo punto per individuare completamente lo spazio di probabilità $\left(\Omega, F_{F/\Gamma}, P_{F/\Gamma}\right)$ associato al fenomeno aleatorio condizionato a *Γ*, occorre definire la misura di probabilità $P_{F/\Gamma}$ che chiameremo <u>misura di probabilità condizionata a *Γ*</u>.

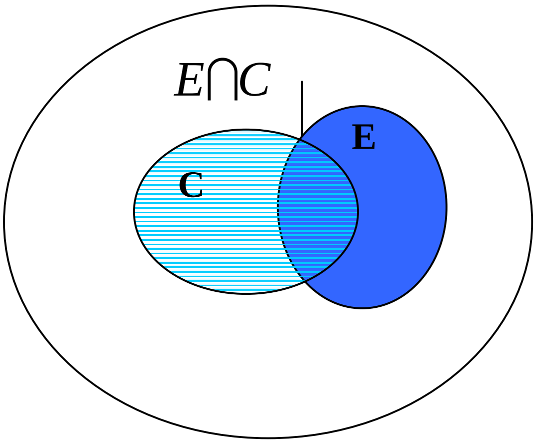


Fig.4

Poiché $F_{F/\Gamma} \subseteq F_F$ si potrebbe pensare di prendere $F_{F/\Gamma}\left(\Omega_C\right) = P_F\left(E \cap C\right)$, così facendo però vengono soddisfatti solo i primi tre assiomi ma non il IV, infatti, poiché lo spazio base in tal caso coincide con $\Omega_c = C$, si avrebbe:

$$F_{F/\Gamma}\left(\Omega_C\right) = P_F\left(C\right) < 1 \quad \forall C \subset \Omega, \qquad (1,36)$$

D'altronde nel caso in cui si prenda $C = \Omega$ la misura di probabilità condizionata $F_{F/\Gamma}$ deve ridursi a P_F. Quanto precede suggerisce di prendere $P_{F/\Gamma}$ non coincidente ma proporzionale a P_F (in tal caso i primi tre assiomi sono ancora rispettati):

$$P_{F/\Gamma}\left(E\right) = kP_F\left(E \cap C\right); \qquad (1,37)$$

Da cui segue immediatamente che

$$k = \frac{1}{P_F(C)}$$

per cui:

$$P_{F/\Gamma}\left(E\right) = \frac{P_F\left(E \cap C\right)}{P_F(C)},$$

affinché però tale relazione abbia sempre senso occorre imporre la condizione aggiuntiva. $P_F(c) > 0$

In base a quanto precede la misura di probabilità condizionata può essere definita come segue.

Def. 1.1 Dato un evento ammissibile *Γ* con $P_\Gamma(\Gamma) > 0$, si definisce misura di probabilità condizionata di un evento *E* rispetto a *Γ* il rapporto :

$$\Pr\left(E/\Gamma\right) = \frac{P_r\left(Ee\Gamma\right)}{P_r(\Gamma)} \qquad (1,41a)$$

ovvero in termini di insiemi:

$$P_{F/\Gamma}\left(E\right) = \frac{P_F\left(E \cap C\right)}{P_F(C)},$$

Si noti che in base a tale definizione $P_{F/F}(E)$ si riduce a $P_F(E)$ come richiesto in precedenza.

Dal punto di vista formale, essendo soddisfatti i 4 assiomi, la definizione è soddisfacente; verifichiamo ora cosa succede dal punto di vista operativo. Se si indica con $N_{E\,e\,\Gamma}$ il numero di volte in cui si verificano contemporaneamente l'evento *E* e l'evento *Γ*, con N_r il numero di volte in cui si verifica l'evento *Γ* e con *N* il numero di prove totale, si ha :

$$\frac{N_{E\,e\,\Gamma}}{N_r} = \frac{\frac{N_{E\,e\,\Gamma}}{N}}{\frac{N_r}{N}} \qquad (1,42)$$

Poiché come anticipato, sotto opportune condizioni che verranno chiarite nel seguito, per la legge dei grandi numeri, i rapporti che compaiono nella (42) tendono al crescere di *N* alle probabilità che compaiono nella (41), la definizione introdotta è in perfetto accordo con l'evidenza sperimentale.

<u>*Esempio*</u>- Consideriamo il caso del lancio di un dado e consideriamo l'evento *Γ*: "sortita di una determinazione ≤ 5". Se ora si prende in esame l'evento *E:* "sortita di un numero pari" si ottiene in base alla definizione (vedi fig. 4):

$$\Pr\left(E/\Gamma\right) = \frac{P_r\left(Ee\Gamma\right)}{P_r(\Gamma)} = \frac{2/6}{5/6} = \frac{2}{5} = 0.4$$

Poiché in questo caso particolare si ha $P_r(E/\Gamma) < P_r(E)$ possiamo dire che il sapere che si è verificato *Γ* cambia il grado di conoscenza a priori posseduta rispetto a *E.*

In generale il sapere che si è verificato l'evento *Γ* altera il grado di conoscenza a priori circa il verificarsi dell'evento *E* in quanto in generale $P_{F/F}(E)$ è diverso da $P_F(E)$.Qualora ciò si verifichi, la conoscenza che si sia verificato l'evento *Γ* non porta alcuna informazione statistica circa il contemporaneo verificarsi dell'evento *E*, e pertanto l'evento *E* si dice statisticamente indipendente da *Γ*.

<u>*Def. 1.2*</u> Un evento *E* si dice statisticamente indipendente dall'evento Γ se e solo se :

$$P_r\left\{E/\Gamma\right\} = P_r\left\{E\right\} \qquad (1,44)$$

Si osservi che in base alla (41) dalla precedente definizione ne consegue che E è statisticamente indipendente da Γ si ha:

$$P_r\{E\ e\ \Gamma\} = P_r\{E\}P_r\{\Gamma\} \tag{1,45}$$

D'altronde è facile verificare che se è anche valida la (45) risulta valida anche la (44). La relazione (45) mette in luce come la proprietà d'indipendenza statistica sia in realtà una proprietà reciproca, infatti dalla (45) segue anche che

$$P_r\{\Gamma/E\} = \frac{P_r\{E\ e\ \Gamma\}}{P_r\{E\}} = P_r\{\Gamma\} \tag{1,46}$$

Per cui si può affermare quanto segue:

Def. 1.2' Due eventi E e Γ si dicono (mutuamente) statisticamente indipendenti se e solo se:

$$P_r\{E\ e\ \Gamma\} = P_r\{E\}P_r\{\Gamma\}$$

$$P_r\{E/\Gamma\} = P_r\{E\}$$

$$P_r\{\Gamma/E\} = P_r\{\Gamma\}$$

Teorema di Bayes

Nell'osservare un fenomeno aleatorio complesso spesso capita di prendere in esame due elementi con fisionomie distinte, di cui solo uno direttamente osservabile, individuanti due fenomeni semplici componenti il fenomeno aleatorio complessivo interconnessi più o meno rigidamente tra loro. Consideriamo ad esempio un sistema di trasmissione come quello in fig. 5, in cui il canale introduce un errore di trasmissione.

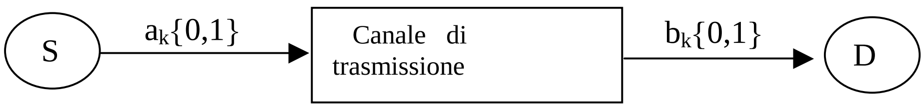


Fig.5

Un primo fenomeno aleatorio è costituito dalla emissione da parte della sorgente S di una sequenza $\{a_k\}$ di N simboli di un alfabeto che possiamo supporre, senza perdita di generalità, binario. Il secondo fenomeno aleatorio è costituito dalla ricezione da parte del destinatario D di una sequenza $\{b_k\}$ di N simboli che, a causa degli errori introdotti dal canale non coincide necessariamente con $\{a_k\}$. Ai due fenomeni semplici possiamo associare due spazi base Ω_1 e Ω_2, mentre al fenomeno aleatorio complessivo possiamo associare uno spazio base $\Omega = \Omega_1 x \Omega_2$ che chiameremo <u>spazio congiunto</u> .

In fig. 6a e fig. 6b sono riportati gli spazi base associabili al fenomeno aleatorio congiunto per i casi di N=1 e N=2.

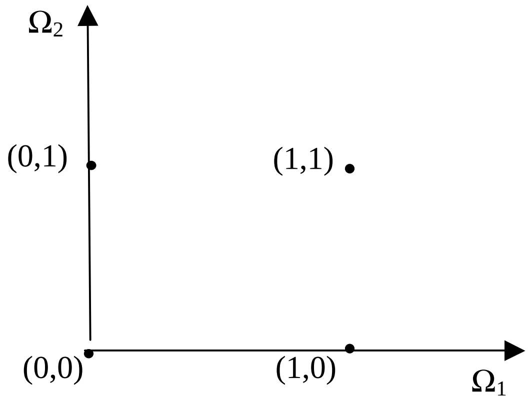


Fig.6a

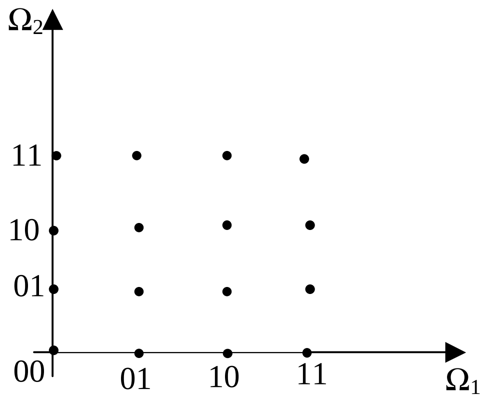


Fig.6b

Esaminiamo per semplicità il caso di N=1. Ora la conoscenza "a priori" posseduta dal destinatario circa il comportamento statistico della sorgente si traduce nell'esplicazione delle probabilità dei due eventi:

$$\Pr\{a_k = 0\},\ \ \Pr\{a_k = 1\} \tag{48}$$

(Si supponga ad esempio che sia $\Pr\{a_k = 0\} = P_0$, $\Pr\{a_k = 1\} = P_1 = 1 - P_0$) mentre la conoscenza sui meccanismi fisici che governano il processo di trasmissione dalla sorgente al destinatario può tradursi in una agevole determinazione delle probabilità condizionate :

$$\Pr\{b_k = 0 / a_k = 0\} \tag{49a}$$
$$\Pr\{b_k = 1 / a_k = 0\} \tag{49b}$$
$$\Pr\{b_k = 0 / a_k = 1\} \tag{49c}$$
$$\Pr\{b_k = 1 / a_k = 1\} \tag{49d}$$

Ciò posto, ricevendo un certo simbolo b_k uguale, ad esempio, a 1 ci si può chiedere se in base alle informazioni possedute, si può determinare quale sia la probabilità che la sorgente abbia inviato proprio il valore 1.

A tale quesito risponde il teorema di Bayes (1763).

Teorema (teorema di Bayes).

Date due partizioni $E_1, E_2, \ldots, E_M$ e $C_1, C_2, \ldots, C_N$ dello spazio base Ω, ovvero

$$\bigcup_{j=1}^{M} E_j = \Omega \ , \ E_i \cap E_j = \emptyset, \quad i \neq j \tag{50a}$$

$$\bigcup_{h=1}^{N} C_h = \Omega \ , \ C_h \cap C_k = \emptyset, \quad h \neq k \tag{50b}$$

corrispondenti agli eventi $\Gamma_h \Leftrightarrow C_h$ e $\theta_j \Leftrightarrow E_j$, note le probabilità degli eventi Γ_h:

$$Pr\{\Gamma_h\}, \quad h=1,\ldots,N \tag{51}$$

e le probabilità condizionate degli eventi θ_j rispetto a Γ_h:

$$Pr\{\theta_j / \Gamma_h\}, \quad j=1,\ldots,M \ , h=1,\ldots,N \tag{52}$$

le probabilità condizionate degli eventi Γ_h rispetto agli eventi θ_j sono date da:

$$Pr\{\Gamma_h / \theta_j\} = \frac{Pr\{\theta_j / \Gamma_h\} \cdot Pr\{\Gamma_h\}}{\sum_{k=1}^{N} Pr\{\theta_j / \Gamma_k\} \cdot Pr\{\Gamma_k\}} \tag{53}$$

Prova In base alla definizione di probabilità condizionata si ha che:

$$Pr\{\Gamma_h / \theta_j\} = \frac{Pr\{\Gamma_h \ e \ \theta_j\}}{Pr\{\theta_j\}} \tag{54}$$

ma poiché è anche:

$$Pr\{\Gamma_h \ e \ \theta_j\} = Pr\{\theta_j / \Gamma_h\} \cdot Pr\{\Gamma_h \tag{55}$$

si ha:

$$Pr\{\Gamma_h / \theta_j\} = \frac{Pr\{\theta_j / \Gamma_h\} \cdot Pr\{\Gamma_h}{Pr\{\theta_j\}} \tag{56}$$

A questo punto per poter esprimere il denominatore in funzione delle qualità note si può utilizzare il fatto che $C_1, C_2, \dots, C_N$ costituisce una partizione di Ω e quindi:

$$E_j = E_j \cap \Omega = E_j \cap \left(\bigcup_{k=1}^{N} C_k\right) = \bigcup_{k=1}^{N} (E_j \cap C_k) \tag{57}$$

Poiché gli insiemi C_k, k= 1, …, N sono disgiunti per ipotesi, lo sono anche gli insiemi $E_j \cap C_k$, k= 1, …, N (vedi fig. 7) e pertanto può essere applicato il III° assioma ottenendo:

$$Pr\{\theta_j\} = \sum_{k=1}^{N} Pr\{\theta_j \ e \ \Gamma_h\} = \sum_{k=1}^{N} Pr\{\theta_j / \Gamma_h\} \cdot Pr\{\Gamma_h\} \tag{58}$$

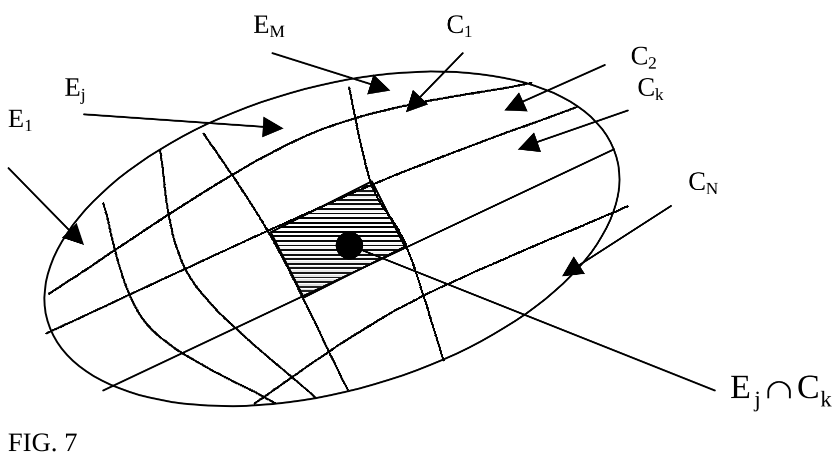


FIG. 7

Che sostituita nella (56) da direttamente la (53). C. v. d.

Osservazione Si noti come nella dimostrazione del teorema sia in realtà utilizzata una sola partizione, ciò implica che il teorema continua a valere qualora si consideri un unico evento $\theta \Leftrightarrow \mathrm{E}$ ed una partizione C_1, C_2, …, C_N , e pertanto esso può anche essere formulato come segue.

Teorema (teorema di Bayes). Date una partizione C_1, C_2, …, C_N dello spazio base Ω (ovvero

$$\bigcup_{h=1}^{N} \mathrm{C_h} = \Omega \ , \ \mathrm{C_h} \cap \mathrm{C_k} = \emptyset, \quad \mathrm{h} \neq \mathrm{k} \tag{59}$$

corrispondenti agli eventi $\Gamma_h \Leftrightarrow \mathrm{C_h}$ e l'evento $\theta \Leftrightarrow \mathrm{E}$, note le probabilità degli eventi Γ_h:

$$Pr\{\Gamma_h\}, \ \ \mathrm{h} = 1,\dots,\mathrm{N} \tag{60}$$

e le probabilità condizionate dell'evento θ rispetto a Γ_h:

$$Pr\{\theta/\Gamma_h\}, \quad \mathrm{h} = 1,\dots,\mathrm{N} \tag{61}$$

le probabilità condizionate degli eventi Γ_h rispetto all'evento θ sono date da:

$$Pr\{\Gamma_h/\theta\} = \frac{Pr\{\theta/\Gamma_h\} \cdot Pr\{\Gamma_h\}}{\sum_{k=1}^{N} Pr\{\theta/\Gamma_k\} \cdot Pr\{\Gamma_k\}} \tag{62}$$