LEZIONI DI STATISTICA E

CALCOLO DELLE PROBABILITA'

UMBERTO MAGAGNOLI

Materiale per il Corso di lezioni di

"STATISTICA"

Laurea magistrale in "Matematica"
Facoltà di Scienze Matematiche, Fisiche e Naturali
Università di Ferrara
Anno accademico 2010-11

PARTE SECONDA
"Probabilità e variabili casuali"

1. Ruolo del Calcolo delle probabilità nella metodologia Statistica

A fianco della Statistica descrittiva, che è stata oggetto della parte I di questo testo, esiste un ulteriore livello è quello di considerare la "Statistica" come disciplina avente un metodo proprio in grado di proporre "leggi" e procedure operative, con un continuo sviluppo innovativo, che evidenziano la "concezione scientifica" della disciplina.

Da questo punto di vista la metodologia scientifica della Statistica, indaga le modalità di conduzione delle osservazioni, del campionamento e dei piani sperimentali, indicando l'attendibilità e la validità dei risultati conseguiti. Essa può intendersi come una "interfaccia" per ogni ricerca applicata, indipendentemente dal settore scientifico in cui si svolge.

Per fare ciò si avvale del metodo induttivo: dal particolare trae conoscenze generali; dalle "rilevazioni campionarie" o "parziali" si ricavano conoscenze riguardanti l'interezza del fenomeno, esprimendo informazioni sulle possibili manifestazioni future.

Questa procedura si definisce "inferenza statistica", a essa è associato il concetto di "rischio di decisione errata", data l'incompletezza delle informazioni.

Come misura del grado d'incertezza di ogni evento o decisione ci si avvale del concetto di "probabilità", a cui è affidato il compito di misurare attraverso un numero compreso tra 0 e 1 il rischio di errori decisionali.

Lo strumento logico appropriato per lo studio dei fenomeni sotto il profilo statistico è quello del "modello".

Il modello mette in luce in primo luogo i legami, le leggi o le regolarità che legano le diverse grandezze, avvalendosi di relazioni matematiche, che esprimono le relazioni di causa-effetto: componente "strutturale".

Ogni modello presenta anche la componente "aleatoria", espressa mediante una legge di distribuzione, attraverso i parametri che la definiscono, mettendo in luce la diversità tra le osservazioni anche se sono svolte in condizioni di costanza ambientale.

Il modello, nella sua formulazione matematica, risponde alle esigenze di conoscenza razionale della realtà fenomenica, ne favorisce la comprensione e consente di individuare la scelta operativa più congrua.

Prima di affrontare le tematiche proprie della statistica inferenziale, a cui sarà dedicata la parte III di questo testo verranno qui presentati i principali strumenti e i metodi probabilistici collegati agli eventi che hanno "regolarità statistica" e le distribuzioni delle "variabili casuali" (*v.c.*).

2. Classi e spazi di insiemi

Nelle indagini statistiche inferenziali, che come è stato precisato interessano i così detti "fenomeni ripetitivi", è opportuno definire particolari insiemi e classi di sottoinsiemi.

L'oggetto delle indagini statistiche è lo studio di un fenomeno che presenta delle manifestazioni "sperimentali" e che ha la possibilità di essere "osservato".

Per individuare le manifestazioni del fenomeno si richiede la conduzione di un "esperimento", cioè l'estrinsecazione del fenomeno che avviene mediante una successione di operazioni.

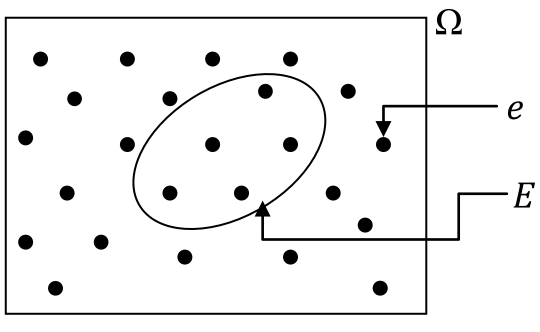
- L'oggetto dell'esperimento casuale è lo studio di alcune proprietà del fenomeno che vengono indicate come "caratteri".

- I caratteri sono posseduti dalle singole “unità osservabili”.
- Stabilita quale sia l’unità osservabile, si eseguono le operazioni di “rilevazione” mediante replicazioni successive.
- La raccolta dei dati rilevati, il loro trattamento e analisi portano all’ottenimento del “risultato”.

Si possono così definire delle distinte “unità sperimentali osservabili” che si possono denominare “unità statistiche” o “eventi elementari”: e.

L’insieme costituito da tutte le unità sperimentali osservabili è detto “spazio degli eventi elementari”, questi ultimi costituiscono gli “elementi” di tale spazio: Ω.

L’interesse dello studioso è rivolto spesso non al singolo evento elementare ma a un raggruppamento di tali eventi elementari, che viene denominato genericamente come “evento”, costituendo un sottoinsieme di Ω, indicato con E. Ogni evento è un raggruppamento di eventi elementari contenuti in Ω. Risulta utile l’impiego della rappresentazione grafica degli insiemi mediante i diagrammi di Eulero-Venn.



Si definisce “campo” o “classe” un particolare insieme costituito da tutti gli sottoinsiemi contenuti in Ω che soddisfano alcune specifiche proprietà e/o regole operative.

Tra le “classi d’insiemi” sono di particolare interesse in ambito probabilistico, le seguenti:

a) Classe di "Boole" o "booleana";
b) Classe di "Borel" o "σ-algebra".

La classe di Borel, presenta proprietà più generali di quella di Boole, quindi è preferibile, o in generale, indispensabile laddove lo spazio degli eventi elementari Ω abbia la potenza del continuo considerare la classe di Borel.

a) La classe di Boole, $\mathcal{B}°$, soddisfa le seguenti proprietà.
   I. Se $E \subset \Omega \in \mathcal{B}°$ allora $E^c \in \mathcal{B}°$, dove $E^c = \Omega - E$ è l'evento complementare di E.
   II. Se $E_1 \subset \Omega \in \mathcal{B}°$ e $E_2 \subset \Omega \in \mathcal{B}°$ allora $E_1 \cup E_2 \in \mathcal{B}°$.

Per la classe di Boole si può verificare che a essa appartengono sia l'intero spazio degli eventi elementari Ω sia l'insieme vuoto $\emptyset$, inoltre la classe è "chiusa" rispetto alle operazioni di unione e intersezione di due o più eventi, in numero finito, e di differenza di due eventi, appartenenti a Ω . Per queste proprietà della classe di Boole essa è detta "classe additiva finita di insiemi".

Se lo spazio Ω è costituito da un numero finito di elementi, o eventi elementari, di numerosità M, esso costituisce un campo o classe di Boole. In tal caso $\mathcal{B}°$ consiste in 2^M eventi (sottoinsiemi di Ω) e tutti gli eventi possono essere elencati, data la cardinalità finita di Ω.

In tal caso $\mathcal{B}°$ è costituito dai seguenti raggruppamenti di eventi elementari:

- $\binom{M}{0} = 1$ contenente nessun elemento – insieme "vuoto" $\emptyset$;
- $\binom{M}{1} = M$ contenente 1 elemento;
- $\binom{M}{2} = \frac{M(M-1)}{2}$ contenente 2 elementi;
- . . . . ;
- $\binom{M}{M-1} = M$ contenente $M-1$ elementi;
- $\binom{M}{M} = 1$ contenente tutti gli M elementi – spazio Ω.

Il numero complessivo dei sottoinsiemi contenuti in $\mathcal{B}°$ è detto "potenza" o anche "cardinalità" di Ω.

b) Se alle proprietà I. e II., valide per la classe di Boole, si aggiunge la seguente terza proprietà si ottiene la classe di Borel $\mathcal{B}$.

III. Se $E_1, E_2, \cdots E_i, \cdots$ è una sequenza infinita, numerabile, di sottoinsiemi di Ω appartenenti a $\mathcal{B}$ allora anche l'unione $\cup_i E_i \in \mathcal{B}$.

Per la classe di Borel si verifica ovviamente quanto detto per la classe di Boole, ossia che a essa appartengono sia l'intero spazio degli eventi elementari Ω che l'insieme vuoto $\emptyset$, inoltre la classe è "chiusa" rispetto alle operazioni di unione e intersezione di sequenze numerabili di eventi, in numero finito ma anche in una infinità numerabile, e di differenza di due eventi, appartenenti a Ω. Per queste proprietà la classe di Borel è detta "classe completamente additiva di insiemi".

Quando lo spazio degli eventi elementari è costituito da: a) una infinità numerabile; b) una infinità non numerabile; i raggruppamenti di tali elementi (eventi) sono possibili sulla base di infinite scelte arbitrarie risultando così non definibili e alcuni non osservabili. Si rende necessario ricorrere a una "classe" o "algebra" che soddisfi una serie di "proprietà formali" quali quelle della classe di Borel.

E' possibile sintetizzate le proprietà formali della classe di Borel o σ-algebra $\mathcal{B}$ relativa allo spazio degli eventi elementari Ω come segue.

Sia data una sequenza di eventi $E_1, E_2, \cdots, E_1, \cdots \in \mathcal{B}$, allora

$$\Omega, \in \mathcal{B} \text{ e } \emptyset, \in \mathcal{B}$$

$$\bigcup_{i=1}^{\infty} E_i \in \mathcal{B}$$

$$\bigcap_{i=1}^{\infty} E_i \in \mathcal{B}$$

Se Ω è costituito da un numero finito di elementi, come si verifica nel caso di valori numerici da un insieme discreto e finito di elementi, è sufficiente considerare la classe di Boole. Se Ω contiene una infinità numerabile di elementi o ha la potenza del continuo, situazione che si concretizza quando si studiano *v.c.* continue aventi come dominio $\Re^d$, allora si richiede l'impiego di una classe di Borel. Si dispone così di due spazi d'insiemi $\{\Omega, \mathcal{B}°\}$ oppure $\{\Omega, \mathcal{B}\ \}$.

Si comprende come per la classe di Borel sia impossibile elencare tutti gli elementi, rispettando le proprietà formali. La classe di Borel, per definizione, risulta "chiusa" rispetto alle operazioni d'insiemi: a) unione; b) intersezione; c) differenza.

Nella situazione che spesso interessa, se lo spazio degli eventi elementari è costituito dall'asse dei numeri reali $\Omega \equiv \Re$ conviene prendere in considerazione una classe "iniziale" di sottoinsiemi di Ω formata da tutti gli intervalli semiaperti $(a, b]$ con $a \leq b$ e $a, b \in \Re$. Il campo di Borel che si ottiene, formato dalla classe iniziale e da tutti gli ulteriori insiemi dati dalle operazioni di unione, intersezione e differenza su componenti finiti o d'infinità numerabile di intervalli, è indicato con $\mathcal{B}_1$. Lo stesso campo di Borel si ottiene anche se come classe iniziale si è considerata quella generata dagli intervalli di tipo: (a, b) o $[a, b)$ o $[a, b]$ di $\Re$.

Si può generalizzare quanto detto considerando come spazio degli eventi elementari quello euclideo a d dimensioni $\Omega \equiv \Re^d$, prendendo in considerazione intervalli a d dimensioni (rettangoli) $(\mathbf{a}, \mathbf{b}]$ con $\mathbf{a} = (a_1, a_2, \cdots, a_d)$ e $\mathbf{b} = (b_1, b_2, \cdots, b_d)$ con $a_i \leq b_i$ per $i = 1, 2, \cdots, d$, ottenendosi, quindi, la classe di Borel $\mathcal{B}_d$.

3. Impostazione frequentistica del calcolo delle probabilità

Si considerino esperimenti replicati che si presentano con risultati aleatori (casuali) statisticamente "stabili".

Un esperimento è detto statisticamente stabile se soddisfa condizioni, spesso vengono dette "legge empirica del caso", che consistono nelle seguenti.

- L'esperimento è ripetibile quante volte si desidera, nelle stesse condizioni, quindi ogni sua replicazione è autonoma rispetto a ogni altra replicazione.
- All'esperimento può associarsi un coppia di spazi $\{\Omega, \mathcal{B}\}$ con le proprietà precisate in precedenza.
- Data una successione di replicazioni, per ogni evento contenuto nella classe di Borel $E \subset \Omega$ e $E \in \mathcal{B}$, sia N il numero di replicazioni dell'esperimento eseguite e $N_E = \#(E)$ il numero di risultati *x* tali che $x \in E$, si ha

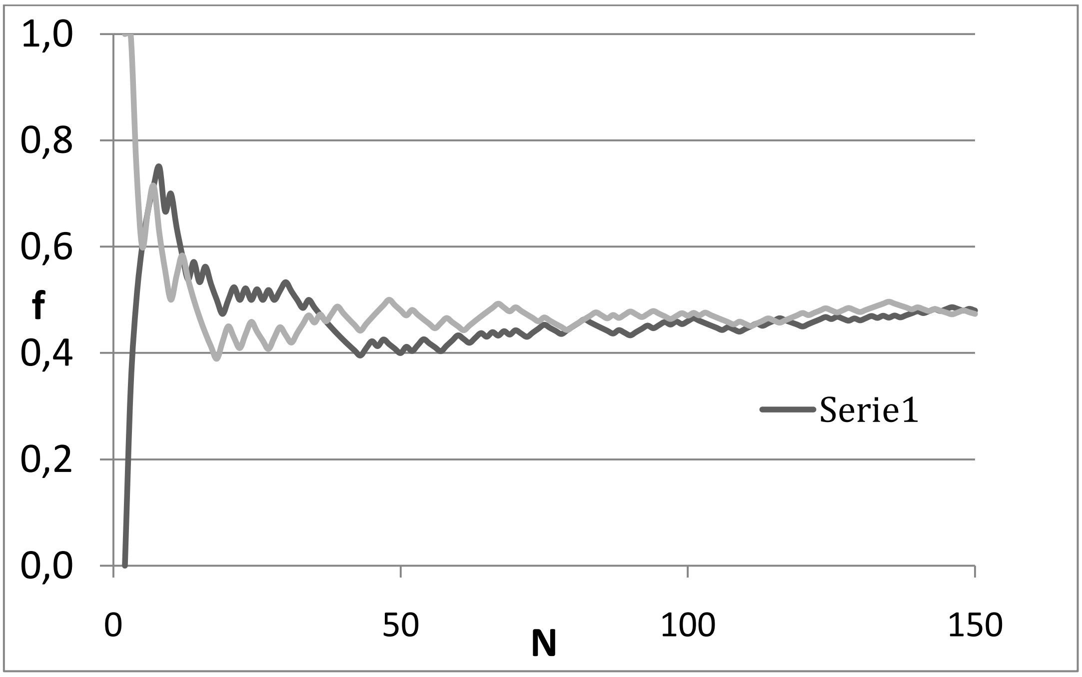
$$\left|\frac{N_E}{N} - P(E)\right| < \varepsilon$$

Dove $\varepsilon > 0$ è un numero scelto a piacere (piccolo) e la relazione è valida per ogni $N > N_0$ essendo N_0 un numero intero che dipende da ε, da E e dalla successione di replicazioni. Si osservi che N_0 è un numero aleatorio e la relazione è valida per *N* sufficientemente grande. L'espressione precedente può intendersi come "limite in probabilità" e il valore $P(E) = cost.$ rappresenta la probabilità di *E*.

A questa modalità di introduzione del concetto di probabilità di un evento viene dato il nome di teoria "frequentistica della probabilità" che, per i caratteri del concetto limite di frequenza relativa di ogni evento, è in sintonia con la metodologia statistica.

Si è fatto ricorso al concetto di regolarità “statistica” che consiste nell’associare a ogni evento della classe $\mathcal{B}$ una misura del grado di verosimiglianza dell’evento in termini di un intervallo compreso tra 0 e 1 come avviene per la frequenza (relativa), associata a ogni possibile risultato, di una successione di osservazioni replicate.

A completamento dei concetti precedenti si può considerare quello di “evento casuale” ovvero quell’evento che si manifesta con una frequenza (relativa) a seguito di una lunga sequenza di osservazioni (replicazioni), con una modalità che si approssima a un valore stabile, che viene inteso come “valore limite” della frequenza (relativa), appunto definito come “probabilità” dell’evento casuale. Il concetto di regolarità statistica o di convergenza in senso statistico/probabilistico viene evidenziato dall’esempio di figura.



Regolarità e convergenza in senso statistico

Nel grafico vengono riportati i risultati delle frequenze relative del verificarsi di uno specifico evento casuale, per un numero di replicazioni N dell’esperimento fino a 150. Si può notare che, per ciascuna delle due serie di replicazioni considerate, la frequenza relativa f risulta differente, con oscillazioni del valore che tendono ad attenuarsi al crescere di N. Inoltre, si verifica una forma di convergenza attorno a un valore (nel grafico circa 0,46) che può essere intesa come probabilità dell'evento considerato. Si tratta di una convergenza anche se non di tipo matematico pur sempre indicativa di regolarità statistica del fenomeno osservato.

4. Impostazione assiomatica della Teoria della Probabilità

L'interpretazione del concetto e della misura dell'aleatorietà data dalla probabilità di un fenomeno avente regolarità statistica, è stata delineata e indicata come "frequentistica", che per la sua impostazione osservazionale e quindi statistica secondo la concezione ordinaria del termine, risulta la più immediata.

Va, tuttavia, sottolineato che la Teoria della Probabilità è una "branca della Matematica" che come altre, ad es. la *Geometria*, ha avuto uno sviluppo che partendo dalle definizioni degli elementi oggetto di trattazione si avvale di alcune (poche) proposizioni – detti *assiomi* o *postulati* – che vengono assunti per procedere, mediante il metodo deduttivo, a dimostrare la validità di teoremi di carattere particolare, attinenti all'oggetto considerato.

Pertanto tale teoria si basa sulla "deduzione logica", che permette da conoscenze generali di ottenere proposizioni valide nel particolare.

Mediante le regole della Teoria della Probabilità è possibile ottenere un "modello" della componente aleatoria presente nel fenomeno reale che permette di completare la modellizzazione della realtà fenomenica allo studio.

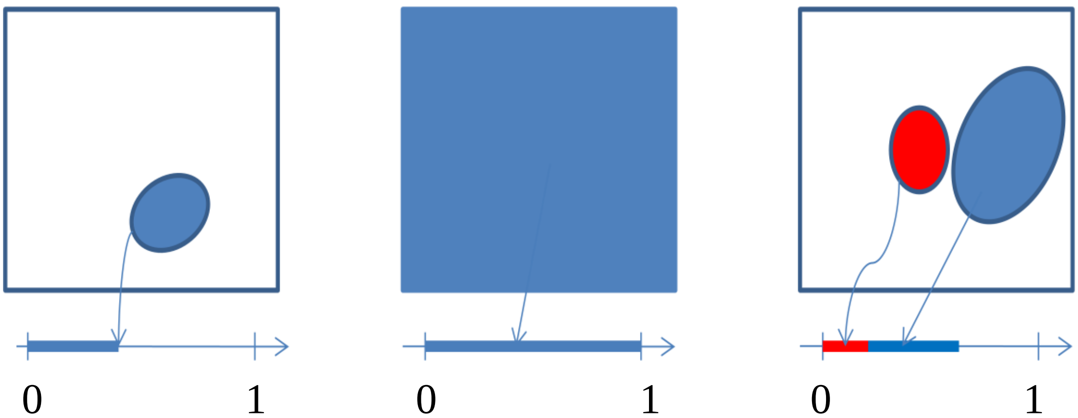
Gli indubbi vantaggi di questa procedura non vanno esenti da *critiche* di carattere sia *filosofico* sia *matematico* che si diversificano secondo le differenti scuole di pensiero, ma sollevano anche problemi interni alla disciplina statistica riguardanti ad esempio le *assunzioni di regolarità* e la *costruzione di in modello* complessivo da ritenersi immagine della realtà.

Non potendoci addentrare nel sistema logico della Teoria assiomatica della Probabilità, presentiamo gli assiomi che costituiscono la base del Calcolo delle Probabilità. Nella figura sottostante, avvalendoci della rappresentazione grafica degli insiemi, possiamo indicare il generico evento casuale con una *area* (ovale in figura), l'evento elementare con ogni *punto* del rettangolo che è lo spazio degli eventi, unione di tutti gli eventi elementari.

- *Assiomi del calcolo delle probabilità*

1. La probabilità di ogni evento casuale è un numero non negativo: $0 \leq \mathrm{P(E)} \leq 1$ per $\mathrm{E} \subset \Omega, \mathrm{E} \in \mathcal{B}$.
2. La probabilità dell'evento "certo" è pari a *uno*:
$$P(\Omega) = 1 \text{ per } \Omega \in \mathcal{B}.$$
3. La probabilità dell'unione di due eventi *disgiunti* e pari alla *somma* delle probabilità dei due eventi:
$$P(E_1 \cup E_2) = P(E_1) + P(E_2)$$
$$\text{per } E_1, E_2 \subset \Omega, E_1, E_2 \in \mathcal{B} \text{ e } E_1 \cap E_2 = \emptyset$$

Si può osservare che, mentre i primi due assiomi sono delle definizioni che permettono di assegnare a ogni evento una probabilità come misura dello stesso, il terzo assioma costituisce la base del vero e proprio Calcolo della Probabilità, precisando le caratteristiche di *additività*, perfettamente analoga alla relazione di somma delle aree per figure geometriche non sovrapposte.



Assiomi del Calcolo della Probabilità

5. Teorie probabilistiche

Oltre alle due teorie già presentate, "frequentistica" e "assiomatica", nell'evoluzione storica del pensiero sui problemi della casualità sono state proposte altre teorie, tra le quali si ricordano quella "classica" che si rifà oltre che a J. Bernoulli, a Newton, Pascal e Laplace, che studiarono il caso nei cosiddetti giochi: lancio della moneta, estrazione di carte da un mazzo, gioco dei dadi, ecc., in cui ogni evento elementare o risultato è inteso come equiprobabile, permettendo così di assegnare la probabilità degli eventi mediante conteggio. Tale proposta secondo alcuni non è che una tautologia. Vi è poi la teoria "soggettivistica"che supera alcuni limiti concettuali delle teorie precedenti, assegnando la probabilità di un evento alla valutazione personale di chi è interessato alle decisioni, nel rispetto di coerenza e additività quali vengono stabilite dalla teoria "assiomatica".

E' possibile indicare in modo sintetico come sono state introdotte nel tempo tali approcci e a quali studiosi si fa riferimento:

1. *Teoria "Classica"* (J. Bernoulli; 1654-1705);
2. *Teoria "Frequentistica"* (R. Von Mises; 1883-1953);
3. *Teoria "Assiomatica"* (A.N. Kolmogorov ; 1903-1987);
4. *Teoria "Soggettivistica"* (B. De Finetti; 1906-1985 – L.J. Savage; 1917-1971).

6. Definizione di spazio probabilistico

Esiste una condizione sufficiente affinché $\mathcal{B}$ sia una classe di Borel che consiste nell'essere una classe "chiusa" rispetto:

a) all'operazione di unione, di una infinità numerabile

$$E_1, E_2, \cdots, E_j, \cdots \in \mathcal{B} \rightarrow \bigcup_{j=1}^{\infty} E_j \in \mathcal{B};$$

b) all'operazione di intersezione, di una infinità numerabile

$$E_1, E_2, \cdots, E_j, \cdots \in \mathcal{B} \to \bigcap_{j=1}^{\infty} E_j \in \mathcal{B};$$

c) all'operazione di complemento

$$E \in \mathcal{B} \to \Omega - E \in \mathcal{B}$$

Si introduca una funzione a valori reali per ogni elemento della classe di Borel $\mathcal{B}$ che denominiamo "distribuzione di probabilità" $\boldsymbol{\mathcal{P}} = \{\{P(E)\ \forall\ E \in \mathcal{B}\}$ che soddisfa gli assiomi del calcolo delle probabilità, $0 \leq P(E) \leq 1\ \ \forall E \in \mathcal{B}$, $P(\Omega) = 1$.

Sia $E_1, E_2, \cdots, E_j, \cdots \in \mathcal{B}$ una successione di eventi e inoltre $E_j \cap E_l = \emptyset\ \ \forall\ j \neq l$ allora:

$$P\left(\bigcup_{j=1}^{\infty} E_j\right) = \sum_{j=1}^{\infty} P(E_j)$$

condizione che assicura la "completa additività" nei confronti della successione di eventi.

La proprietà dell'additività numerabile è garantita aggiungendo, a quella per successioni finite, la seguente condizione limite

$$\forall\ E_1 \supset E_2 \supset \cdots \supset E_j \supset \cdots \quad \to \quad \lim_{j\to\infty} E_j = \cap_{j=1}^{\infty} E_j = \emptyset$$
$$\to \quad \lim_{j\to\infty} P(E_j) = 0$$

La terna di spazi: Ω spazio degli eventi elementari, $\mathcal{B}$ classe di Borel, $\boldsymbol{\mathcal{P}}$ distribuzione di probabilità, viene a costituire $\{\Omega, \mathcal{B}, \boldsymbol{\mathcal{P}}\}$ il cosiddetto "spazio probabilistico" che permette di definire in modo completo quanto riguarda il fenomeno oggetto dell'esperimento. Se $\Omega \equiv \Re^d$, come si verifica nello studio di *v.c.* a *d* dimensioni, la classe di Borel è generata dai "rettangoli a *d* dimensioni", allora $\mathcal{B} \equiv \mathcal{B}_d$.

Si ricorda che con $P(E)$ si misura il grado di aspettativa o di verosimiglianza del verificarsi del generico evento E.

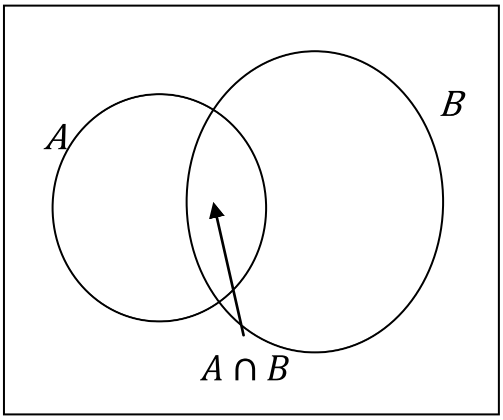
7. Alcuni teoremi di calcolo delle probabilità

- **Teorema 1**. Siano $A, B \in \Omega$ due eventi "non disgiunti", la probabilità dell'evento unione $A \cup B$ vale

$$P(A \cup B) = P(A) + P(\cup B) - P(A \cap B)$$

Dimostrazione. Impiegando l'algebra degli insiemi abbiamo (v. il corrispondente diagramma di Eulero-Venn)

$$A \cup B = A \cup (B - A \cap B) \qquad B = (B - A \cap B) \cup (A \cap B)$$



Gli insiemi al secondo membro delle relazioni precedenti sono disgiunti, potendosi così applicare il 3° assioma, quindi:

$$P(A \cup B) = P(A) + P(B - A \cap B)$$

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

Sottraendo membro a membro si ha

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

da cui si conclude

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

- **Teorema 2**. Siano $A, B \in \Omega$ due eventi "non disgiunti", la probabilità dell'evento differenza $A - B$ vale

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

Dimostrazione. Impiegando l'algebra degli insiemi abbiamo (v. il corrispondente diagramma di Eulero-Venn): gli eventi $A - B$ e B sono disgiunti e, quindi, anche $A - B$ e $A \cap B$. L'evento A può considerarsi unione tra $A - B$ e $A \cap B$

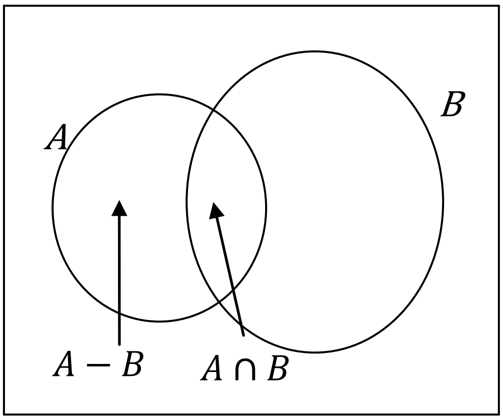
$$A = (A - B) \cup (A \cap B) \quad \text{con} \quad (A - B) \cap (A \cap B) = \emptyset$$

Quindi, per il 3° assioma:

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

da cui

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



- **Teorema 3**. Siano $A, B \in \Omega$ due eventi e l'evento A contiene B

$$A \supset B$$

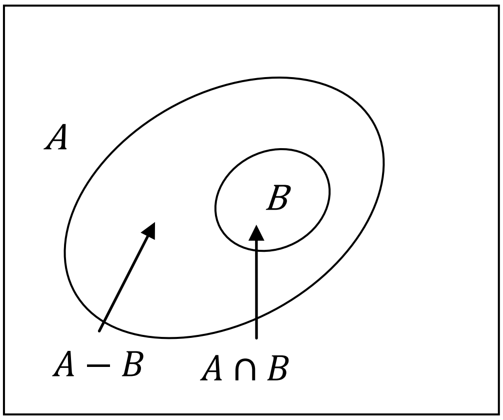
allora si avrà

$$P(A) \geq P(B)$$

Dimostrazione. Se $A = B$ allora $P(A) = P(B)$. Se B è contenuto propriamente in A, si ha la seguente relazione tra gli eventi

$A = (A - B) \cup (A \cap B)$ con $(A - B) \neq \emptyset$ e $(A \cap B) \neq \emptyset$ ed essendo poi $A \cap B = B$ ne consegue che $A = (A - B) \cup B$.

Poiché gli eventi al secondo membro sono disgiunti, per il 3° assioma abbiamo $P(A) = P(A - B) + P(B)$ e per il 1° assioma $P(A - B) \geq 0$ abbiamo $P(A) \geq P(B)$, che completa la dimostrazione del teorema.



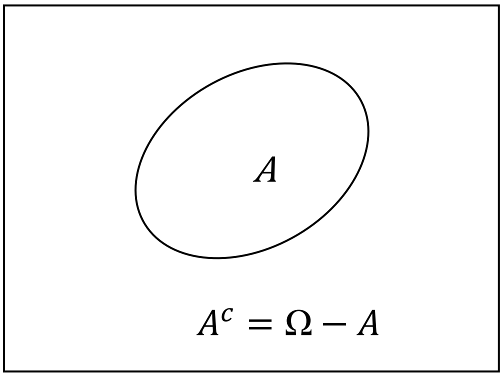
- **Teorema 4**. La probabilità dell'evento "complementare" $A^c = \Omega - A$ è pari a:

$$P(A^c) = 1 - P(A)$$

Dimostrazione. Ricordando che $A^c = \Omega - A$ e quindi $A^c \cup A = \Omega$, $A^c \cap A = \emptyset$, per i assiomi 2° e 3° si ottiene

$$P(A^c \cup A) = P(A^c) + P(A) = P(\Omega) = 1$$

e quindi $P(A^c) = 1 - P(A)$.



- **Teorema 5**. La probabilità dell'evento "impossibile" $\emptyset$ è pari a zero.

$$P(\emptyset) = 0$$

Dimostrazione. Consideriamo l'evento unione tra Ω e $\emptyset$: $\Omega \cup \emptyset$. Tale evento appartiene alla classe di Borel $\mathcal{B}$ ma non è contenuto in Ω (Ω è costituito solo da elementi sperimentabili) quindi per il 3° assioma, si ottiene: $P(\Omega \cup \emptyset) = P(\Omega) + P(\emptyset)$ in quanto eventi disgiunti. Per il 1° assioma $0 \leq P(\Omega) + P(\emptyset) \leq 1$, ma per il 2° assioma $P(\Omega) = 1$ e quindi ne consegue

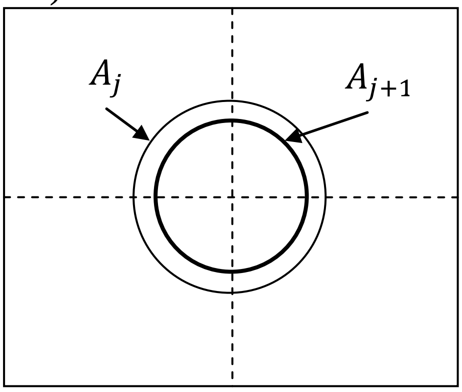
$$0 \leq P(\emptyset) \leq 1 - P(\Omega) = 1 - 1 = 0 \Longrightarrow P(\emptyset) = 0.$$

<u>*Osservazioni*</u>

1. *Probabilità del limite di una successione "non crescente" di eventi*
   Sia $\{A_1, A_2, \cdots, A_j, \cdots\}$ una successione, con $A_j \supset A_{j+1}$, si ha allora: $\lim_{j\to\infty} A_j = \bigcap_{l=1}^{\infty} A_l$. Si dimostra che la probabilità del limite di A_j è pari al limite della successione di probabilità $P(A_j)$: $P(\lim_{j\to\infty} A_j) = \lim\limits_{j\to\infty} P(A_j)$.

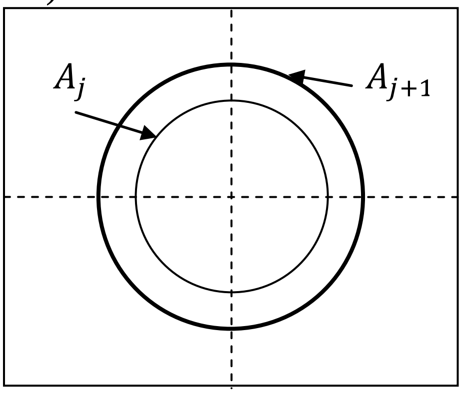
2. *Probabilità del limite di una successione "non decrescente" di eventi*
   Sia $\{A_1, A_2, \cdots, A_j, \cdots\}$ una successione, con $A_j \subset A_{j+1}$, si ha allora: $\lim_{j\to\infty} A_j = \bigcup_{l=1}^{\infty} A_l$. Si dimostra che la probabilità del limite di A_j è pari al limite della successione di probabilità $P(A_j)$: $P(\lim_{j\to\infty} A_j) = \lim\limits_{j\to\infty} P(A_j)$.

a)



Successione di cerchi concentrici
che via via si restringono

b)



Successione di cerchi concentrici
che via via si allargano

3. *Probabilità dell'intersezione tra due o più eventi: la probabilità congiunta*
   Siano A e B due eventi contenuti in Ω, l'evento intersezione $A \cap B$, costituito da tutti gli eventi elementari contenuti in Ω che sono presenti sia in A sia in B, viene detto evento "congiunto" e indicato semplicemente come: $A \cap B \equiv A \cdot B \equiv AB$. Tale rappresentazione può essere generalizzata anche alla situazione di più di due eventi, estendendola a quella di una successione di un'infinità di eventi numerabili, appartenenti alla classe di Borel.

4. *Esigenza di assegnare una distribuzione di probabilità: la situazione dello spazio degli eventi elementari finito – classe di Boole*
   In numerose situazioni sperimentali è possibile modellizzare il fenomeno, in particolare la "casualità", ricorrendo alla classica estrazione da una o più urne di una pallina con re-immissione. Tale situazione generalizza molte varianti riguardanti i così detti "giochi classici", quali: a) lancio di una o più monete; b) estrazione di carte da un mazzo; c) lancio di uno o più dadi; ecc..

   Per assegnare la distribuzione di probabilità $\boldsymbol{\mathcal{P}}$ si rende necessario fare una "assunzione" sulla probabilità di ogni evento elementare (che spesso viene indicata impropriamente come 4° Assioma del calcolo delle probabilità) vale a dire che ogni evento elementare sia "equiprobabile".

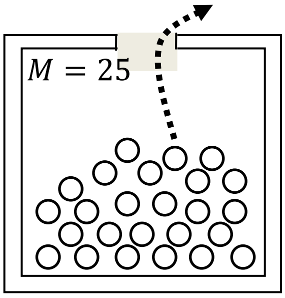
Ad esempio consideriamo un urna contenente M palline indicate come: $\{e_j; j = 1,2,\cdots,M\}$, l'esperimento consiste nell'estrazione di una pallina e successiva re-immissione. La probabilità che ogni evento elementare e_j si verifichi è "costante" e pari a p, per l'assunzione di equiprobabilità, abbiamo $P(e_j) = p \ \forall \ j = 1,2,\cdots,M$. Per determinare il valore p si ricordi che per il 2° assioma $P(\Omega) = P\left(\bigcup_{j=1}^{M} e_j\right) = 1$ e per il 3° assioma, essendo gli eventi elementari disgiunti, $P\left(\bigcup_{j=1}^{M} e_j\right) = \sum_{j=1}^{M} P(e_j) = \sum_{j=1}^{M} p = Mp$ ne consegue che $Mp = 1$, da cui si ottiene $p = 1/M$, condizione che soddisfa il 1° assioma.

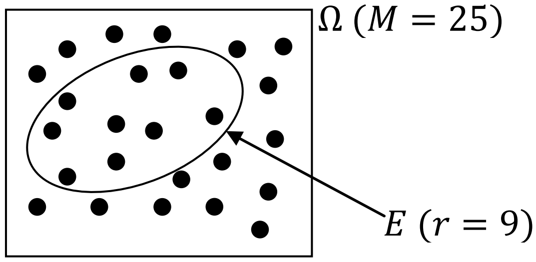
Ad ogni elemento della classe di Boole $E = \{e_1, e_2, \cdots, e_r\}$, costituito da r eventi elementari di Ω, è assegnabile una probabilità:

$$P(E) = P\left(\bigcup_{j=1}^{r} e_j\right) = \sum_{j=1}^{r} P(e_j) = \sum_{j=1}^{r} \frac{1}{M} = \frac{r}{M}$$
$$= \frac{numero\ dei\ casi\ \boldsymbol{favorevoli}: "r"}{numero\ dei\ casi\ \boldsymbol{possibili}: "M"}$$

La definizione data della probabilità di un evento come rapporto tra casi favorevoli e possibili corrisponde a quella propria della teoria "classica" alla cui base sta un conteggio di elementi equiprobabili. Si comprende il ruolo preminente, in questo ambito del "calcolo combinatorio" e quindi delle seguenti grandezze.

- Permutazioni di *n* elementi $P_n = n! = n \cdot (n-1) \cdots 2 \cdot 1$;
- Disposizioni di *n* elementi a *r* a *r*
$$D_{n,r} = n \cdot (n-1) \cdots (n-r+1) = \frac{n!}{(n-r)!};$$
- Combinazioni *n* elementi a *r* a *r*
$$C_{n,r} = \frac{D_{n,r}}{P_r} = \binom{n}{r} = \frac{n!}{r!(n-r)!}.$$





- Urna contenente M palline.
- Fenomeno da osservare: colore della pallina.
- Esperimento: estrazione di una pallina e sua reimmissione.
- Evento elementare: pallina estratta.

5. *Esigenza di assegnare una distribuzione di probabilità: la situazione generale dello spazio degli eventi elementari costituito da una infinità numerabile o avente la potenza del continuo – classe di Borel*

   Per insiemi costituiti da infinità numerabili si deve ricorre a una definizione dello spazio probabilistico e quindi della probabilità di ciascun evento elementare mediante una funzione che si basi sulla teoria "frequentistica" oppure su quella "soggettivistica" per assegnare, nel rispetto degli assiomi del calcolo delle probabilità, le probabilità a ogni evento appartenente alla classe di Borel. Ancor più questo avviene nel caso in cui Ω si presenti con le caratteristiche della potenza del continuo potendosi assegnare le probabilità facendo ricorso alla legge "empirica del caso".

8. Concetto di probabilità condizionata

Dato lo spazio probabilistico $\{\Omega, \mathcal{B}, \boldsymbol{\mathcal{P}}\}$ si considerino due eventi A e B, contenuti in $\mathcal{B}$.

La probabilità non gode della proprietà distributiva rispetto all'operazione "prodotto d'insiemi" ossia dell'intersezione, come si è già visto per l'operazione "somma d'insiemi" ossia dell'unione, in cui la proprietà distributiva è valida solo se gli eventi A e B sono disgiunti, quindi in generale si ha

$$P(A \cdot B) \neq P(A) \cdot P(B)$$

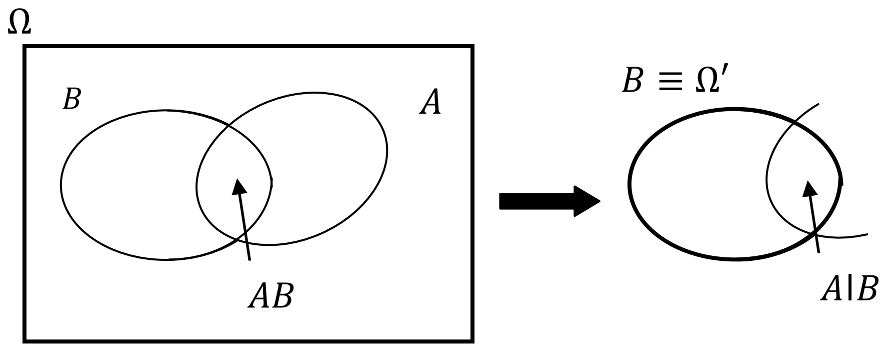
Da cui consegue, se $P(B) > 0$, che

$$\frac{P(AB)}{P(B)} \neq P(A)$$

Il rapporto al primo membro è indicato come $P(A|B)$ e definisce la "probabilità condizionata" dell'evento A al verificarsi dell'evento B.

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \quad \text{con } \ P(B) > 0$$

dove con $A \cap B \equiv A \cdot B \equiv AB$ si indica l'evento congiunto.



Esempio 1.

Consideriamo l'esperimento casuale del lancio di un dado. Come possibili risultati abbiamo le sei facce, identificate dai numeri segnati. Lo spazio degli eventi elementari è $\Omega = \{1,2,3,4,5,6\}$; assunta la condizione di equiprobabilità si può probabilizzare qualsiasi evento contenuto in Ω. Consideriamo $A = \{1,2\}$, cioè che il risultato dell'esperimento sia dato dalla faccia "1" o "2", la probabilità di A è $P(A) = \frac{numero\ dei\ casi\ \boldsymbol{favorevoli}}{numero\ dei\ casi\ \boldsymbol{possibili}} = \frac{2}{6} = \frac{1}{3}$. Ci si limiti ora a considerare come spazio di interesse, al posto di Ω, quello ottenuto scartando gli eventi elementari "5" e "6" ossia solo $\Omega^{'} \equiv B = \{1,2,3,4\}$. In questo nuovo spazio istituiamo un nuovo sistema di probabilità con relativa classe di Borel. La probabilità dell'evento $A = \{1,2\}$ in questo spazio vale $P'(A) = \frac{numero\ dei\ casi\ \boldsymbol{favorevoli}}{numero\ dei\ casi\ \boldsymbol{possibili}} = \frac{2}{4} = \frac{1}{2}$. La probabilità $P'(A)$ è pari alla probabilità "condizionata di A dato B", infatti: $P(A|B) = \frac{P(AB)}{P(B)} = \frac{2/6}{4/6} = \frac{1}{2}$. In questo caso si ha che $P(A|B) \neq A$ e possiamo dire che l'evento condizionante B influisce sul valore della probabilità dell'evento A.

E' possibile verificare sia mediante i teoremi del calcolo delle probabilità sia mediante le proprietà della teoria frequentistica (legge empirica del caso) che la trasformazione: $A \to A' = AB$ per $\forall A \in \mathcal{B}$ stabilisce una nuova classe di Borel $\mathcal{B}'$ la cui distribuzione di probabilità $P(A|B)$ gode anch'essa degli assiomi del calcolo delle probabilità.

Infatti, $\mathcal{B}'$ è una classe "chiusa" rispetto alle operazioni di "unione", intersezione" e "differenza".

Data una successione di eventi $\{A_jB; j = 1,2 \cdots\}$, ottenuta come intersezione del generico A_j, con B prefissato e con elementi contenuti in $\mathcal{B}'$, abbiamo

- l'<u>unione</u> $\bigcup_{j=1}(A_jB) \in \mathcal{B}'$, infatti essendo $\bigcup_{j=1} A_j \in \mathcal{B}$ e inoltre $\bigcup_{j=1}(A_jB) = (\bigcup_{j=1} A_j)B$ abbiamo un insieme del tipo $AB \in \mathcal{B}'$;
- l'<u>intersezione</u> $\bigcap_{j=1}(A_jB) \in \mathcal{B}'$, infatti per la proprietà commutativa si ha $\bigcap_{j=1}(A_jB) = (\bigcap_{j=1} A_j)(B \cdot B \cdot \cdots) = (\bigcap_{j=1} A_j)B$ e quindi abbiamo un insieme del tipo $AB \in \mathcal{B}'$;
- la <u>differenza</u> $A_1B - A_2B = (A_1 - A_2)B$, essendo $A = A_1 - A_2 \in \mathcal{B}$ e quindi $A_1B - A_2B = AB \in \mathcal{B}'$.

La classe $\mathcal{B}'$ è "chiusa" rispetto alle operazioni di: unione, intersezione, differenza. Inoltre si possono fare alcune osservazioni.

- Se si considera $A \in \mathcal{B}$ disgiunto da B, allora $AB = \emptyset$ e quindi anche $\emptyset \in \mathcal{B}'$.
- Se si considera $A \supset B \in \mathcal{B}$, allora $AB = B = \Omega'$ e quindi anche $\Omega' \in \mathcal{B}'$.

Ritornando alla diseguaglianza $P(A \cdot B) \neq P(A) \cdot P(B)$ che può scriversi in generale $P(A|B) = \frac{P(AB)}{P(B)} \neq P(A)$, possiamo dire che:

- $P(A)$ è la probabilità relativa all'evento A nell'intero spazio Ω;
- $P(A|B)$ è la probabilità relativa all'evento A nel sottospazio $B = \Omega'$.

9. Concetto di eventi indipendenti

Qualora valga l'uguaglianza tra $P(A|B)$ e $P(A)$, allora l'evento condizionante B non modifica la probabilità di verificarsi di A, si dirà, quindi, che A è "indipendente stocasticamente" da B:

$$P(A|B) = P(A) \rightarrow P(AB) = P(A) \cdot P(B)$$

Se $P(A) < 0$, si può definire, analogamente, la probabilità condizionata $P(B|A)$ che vale

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

Qualora valga l'uguaglianza tra $P(B|A)$ e $P(B)$ allora l'evento condizionante A non modifica la probabilità di verificarsi di B, si dirà, peraltro, che B è "indipendente stocasticamente" da A e questo comporta

$$P(B|A) = P(B) \rightarrow P(AB) = P(A) \cdot P(B)$$

Ne consegue che la condizione di indipendenza (stocastica) di B dato A è equivalente a quella di A dato B, potendosi parlare di "indipendenza tra gli eventi A e B". Tale situazione è verificata dall'uguaglianza della probabilità degli eventi congiunti rispetto al prodotto delle probabilità dei componenti:

$$P(AB) = P(A) \cdot P(B)$$

L'indipendenza equivale ad affermare che:

- La probabilità condizionata di A non dipende da B, che è l'evento condizionante;
- La probabilità condizionata di B non dipende da A, che è l'evento condizionante.

$$\frac{P(AB)}{P(B)} = P(A) \qquad \frac{P(AB)}{P(A)} = P(B)$$

$$\Updownarrow \qquad\qquad\qquad \Updownarrow$$

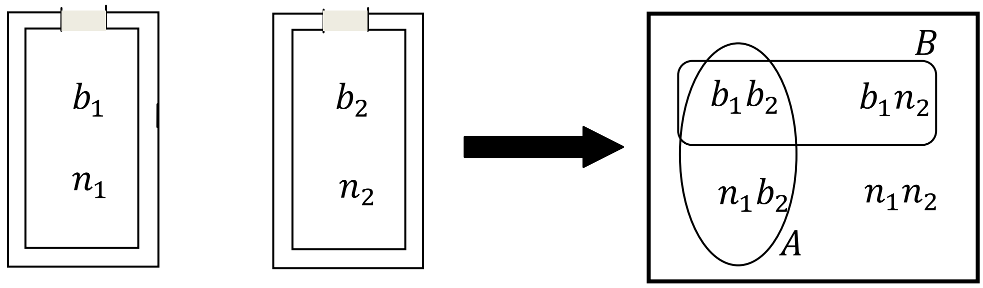
$$= P(A|B) \qquad\qquad = P(B|A)$$

La situazione di "indipendenza" di deve ipotizzare *a priori* in relazione alla struttura probabilistica considerata.

<u>Esempio 2.</u>

Si considerino due urne con palline "bianche" e "nere" in numero pari a: b_1, n_1 per la 1° urna e b_2, n_2 per la 2° urna. L'esperimento consiste nell'estrazione (con rimessa) di una pallina prima dalla 1° urna e poi una pallina dalla 2ª. L'evento elementare è dato da una coppia di palline, rispettando l'ordine di estrazione dalle urne, lo spazio degli eventi elementari Ω è costituito da un numero di punti pari a: $m_T = b_1 \cdot b_2 + b_1 \cdot n_2 + n_1 \cdot b_2 + n_1 \cdot n_2 = (b_1 + n_1) \cdot (b_2 + n_2) = m_1 \cdot m_2$ (ogni pallina delle b_1 si associa con ogni pallina delle b_2, e così via).

Si chiede quale sia la probabilità che la seconda pallina estratta (2ª urna) sia "bianca" supposto che la prima pallina estratta (1ª urna) sia bianca?



Si tratta di ottenere la probabilità condizionata $P(A|B)$, avendo definito gli eventi:

$$A \equiv\ 2°\, pallina\ "bianca";\ B \equiv\ 1°\, pallina\ "bianca".$$

Ci interessa la probabilità di AB nel sottospazio di B. La probabilità congiunta $P(AB)$ risulta

$$P(AB) = \frac{b_1 \cdot b_2}{m_T}$$

e quindi

$$P(AB) = \frac{b_1 \cdot b_2}{m_1 \cdot m_2} = \frac{b_1}{m_1}\frac{b_2}{m_2}$$

Ma essendo

$$P(A) = \frac{b_1 \cdot b_2 + b_1 \cdot b_2}{m_T} = \frac{b_2(b_1 + n_1)}{m_1 \cdot m_2} = \frac{b_2}{m_2}$$

$$P(B) = \frac{b_1 \cdot b_2 + b_1 \cdot n_2}{m_T} = \frac{b_1(b_2 + n_2)}{m_1 \cdot m_2} = \frac{b_1}{m_1}$$

otteniamo la condizione di indipendenza tra i due eventi A e B

$$P(AB) = P(A)P(B)$$

avendosi

$$P(A|B) = \frac{P(AB)}{P(B)} = P(A) = \frac{b_2}{m_2}$$

La condizione di indipendenza riscontrata è dovuta al fatto che ogni caso favorevole per una data modalità si può associare a un caso dell'altra modalità (p.es. *ogni* pallina bianca della 2° urna si può associare con *tutte* le palline bianche della 2° urna).

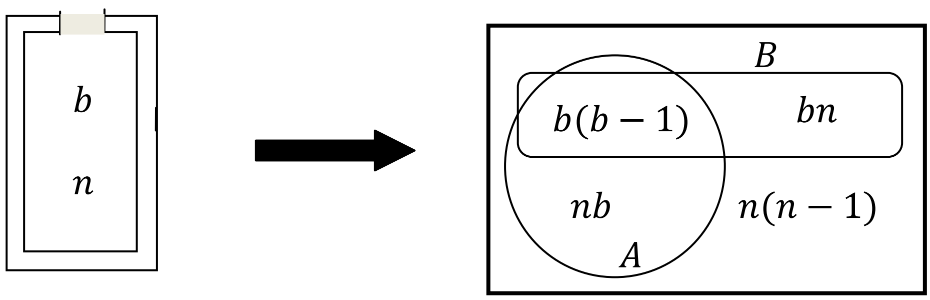
<u>Esempio 3.</u>

Si consideri l'analoga estrazione di due palline da una stessa urna senza "rimessa". Alla prima estrazione se si ha una pallina bianca nell'urna rimangono $(b-1)$ palline bianche e n nere, su cui eseguire la seconda estrazione; qualora alla prima estrazione si abbia una pallina nera nell'urna rimangono b palline bianche e $(n-1)$ nere.

Il numero complessivo di eventi elementari risulta

$$m_T = b \cdot (b-1) + b \cdot n + n \cdot b + n \cdot (n-1) = b(b+n-1) + n(b+n-1) = (b+n)(b+n-1) = m(m-1)$$

essendo $m = b + n$



Si tratta di ottenere ancora la probabilità condizionata $P(A|B)$, avendo definito gli eventi:

$$A \equiv 2° \, pallina \text{ "}bianca\text{"}; \; B \equiv 1° \, pallina \text{ "}bianca\text{"}.$$

$$P(A) = \frac{b \cdot (b-1) + b \cdot n}{m_T} = \frac{b(b+n-1)}{m(m-1)} = \frac{b}{m}$$

$$P(B) = \frac{b \cdot (b-1) + b \cdot n}{m_T} = \frac{b(b+n-1)}{m(m-1)} = \frac{b}{m}$$

$$P(AB) = \frac{b \cdot (b-1)}{m_T} = \frac{b(b-1)}{m(m-1)} = \frac{b}{m}\frac{(b-1)}{(m-1)} \neq P(A)P(B)$$

$$P(A|B) = \frac{b(b-1)}{b(b-1)+bn} = \frac{b(b-1)}{b(m-1)} = \frac{b-1}{m-1} \neq P(A)$$

In questa situazione gli eventi A e B non sono indipendenti e conseguentemente la probabilità $P(A|B) \neq P(A)$.

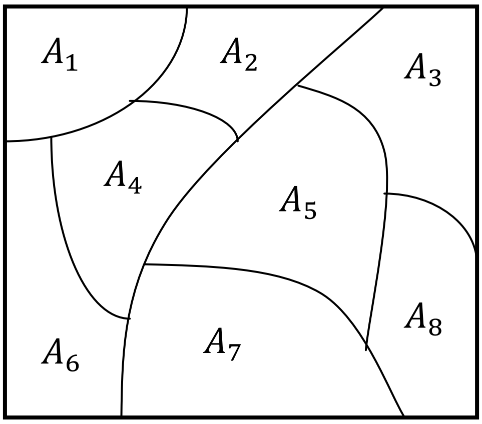
10. La partizione dello spazio degli eventi elementari e formula di Bayes

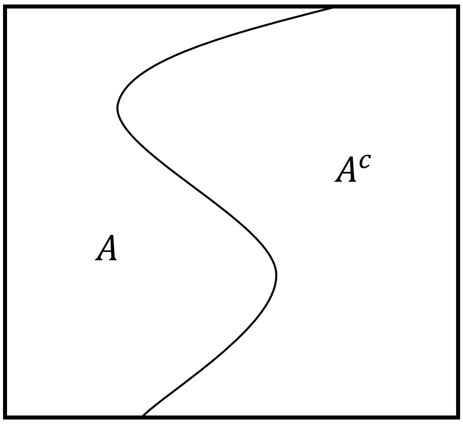
Consideriamo g eventi $\{A_j \subset \Omega; 1, 2, \cdots, g\}$ tali che:

$$\bigcup_{j=1}^{g} A_j = \Omega \quad \text{con} \quad A_j \cap A_l = \emptyset \ \text{ per } j \neq l$$

Una tale successione di eventi costituisce una “partizione” di Ω.

Un esempio è dato dalla coppia $\{A, A^c\}$, situazione dicotomica di generale interesse.





Se si considera l’intersezione di $B \subset \Omega$ con Ω si ha

$$B = B\Omega = \mathrm{B}\left(\bigcup_{j=1}^{g} A_j\right) = \bigcup_{j=1}^{g} (BA_j)$$

dove $\{A_j \subset \Omega; 1, 2, \cdots, g\}$ è una partizione di Ω e $\{BA_j \subset \Omega; 1, 2, \cdots, g\}$ risulta anch'essa una "partizione", ma di B, che è detta "partizione condizionata".

Relativamente alla coppia di eventi A_i e B possiamo definire le "probabilità condizionate":

$$P(A_i|B) = \frac{P(A_iB)}{P(B)} \quad \text{e} \quad P(B|A_i) = \frac{P(BA_i)}{P(A_i)}$$

supposto $P(B) > 0$ e $P(A_i) > 0$, allora $P(BA_i) = P(B|A_i)P(A_i) = P(A_iB)$ da cui

$$P(A_i|B) = \frac{P(A_iB)}{P(B)} = \frac{P(B|A_i)P(A_i)}{P(B)} \qquad (*)$$

Da quanto è stato detto sulle partizioni, si ha

$$P(B) = P\left(\bigcup_{j=1}^{g} (BA_j)\right) = \sum_{j=1}^{g} P(BA_j)$$

essendo $(BA_j)(BA_l) = \emptyset$ per $j \neq l$ e per il 3° assioma si ottiene

$$P(B) = \sum_{j=1}^{g} P(BA_j) = \sum_{j=1}^{g} P(B|A_j)P(A_j)$$

Sostituendo tale espressione nella eq. $(*)$ abbiamo la relazione che è nota come "teorema o formula di Bayes (1707-1761)"

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{\sum_{j=1}^{g} P(B|A_j)P(A_j)}$$

in cui, al 1° membro compare la probabilità di A_i condizionata da B, al 2° membro sono presenti le probabilità di B condizionate dai singoli eventi $\{A_j \subset \Omega; 1, 2, \cdots, g\}$ della partizione di Ω.

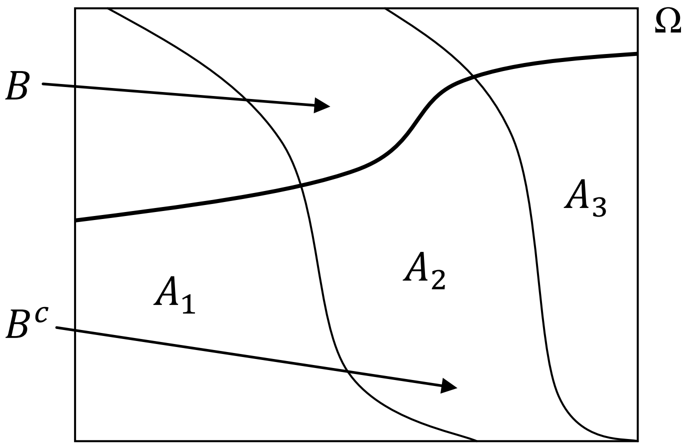
L'importanza della formula di Bayes, in ambito probabilistico e statistico, è legata a una sua duplice interpretazione.

- Gli eventi A_j son intesi come le possibili "cause" che originano un certo "effetto" rappresentato dall'evento B. Si desidera determinare la probabilità $P(A_i|B)$ che sia imputabile al manifestarsi di una particolare "causa" A_i il verificarsi dell'evento B, disponendo dei valori della probabilità $P(A_j)$ associata a ogni "causa" ed essendo note le generiche probabilità $P(B|A_j)$ che le "cause" $\{A_j; 1, 2, \cdots, g\}$ producano l'effetto B.

  Esempio 4.

  In un'azienda vi sono tre catene di montaggio dedicate alla fabbricazione di un prodotto digitale. Dalle esperienze pregresse è noto che le probabilità che il prodotto provenga da ciascuna delle tre catene (cause) sono pari a: $P(A_1) = 1/6$, $P(A_2) = 2/6$, $P(A_2) = 3/6$. Inoltre si conoscono, in relazione alle caratteristiche delle singole catene di montaggio, le probabilità di superamento dei test di verifica (effetto) sono $P(B|A_1) = 0{,}99$, $P(B|A_2) = 0{,}80$, $P(B|A_3) = 0{,}90$. Si richiede la probabilità $P(A_2|B)$ che il prodotto che ha superato il test provenga dalla catena di montaggio A_2.

$$
\begin{aligned}
P(A_2|B) &= \frac{P(B|A_2)P(A_2)}{P(B|A_1)P(A_1) + P(B|A_2)P(A_2) + P(B|A_3)P(A_3)} \\
&= \frac{0{,}80 \cdot 0{,}333}{0{,}99 \cdot 0{,}167 + 0{,}80 \cdot 0{,}333 + 0{,}90 \cdot 0{,}5} = \frac{0{,}2667}{0{,}8817} \\
&= 0{,}3025 = 30{,}25\%
\end{aligned}
$$



- Le probabilità $\{P(A_j); 1, 2, \cdots, g\}$ sono intese come probabilità "*a priori*" degli eventi costituenti la partizione di Ω. Le probabilità $\{P(A_j|B); 1, 2, \cdots, g\}$ sono intese come probabilità "*a posteriori*" degli stessi eventi costituenti la partizione di Ω in relazione al realizzarsi dell'evento B.

<u>Esempio 5.</u>

Si consideri un processo produttivo che esegua la lavorazione di pezzi provenienti da tre fornitori di cui è noto l'impiego di materiali aventi prestazioni differenti, che si traduce in una diversa probabilità di presentare pezzi difettosi. Si richiede, dato che il sistema produttivo ha riscontrato la comparsa di un pezzo difettoso, di determinare la probabilità che tale pezzo provenga da un certo fornitore.

In relazione all'acquisizione dei materiali dai diversi fornitori si considerino le seguenti probabilità "*a priori*"

- $P(A_1) = 0{,}30$, con A_1 l'evento che il pezzo sia proveniente dal 1° fornitore;
- $P(A_2) = 0{,}50$, con A_2 l'evento che il pezzo sia proveniente dal 1° fornitore;
- $P(A_3) = 0{,}20$, con A_3 l'evento che il pezzo sia proveniente dal 1° fornitore.

Sulla base di una serie di controlli di vari lotti presentati dai fornitori si dispone di una stima (legge empirica del caso) della probabilità di pezzi difettosi relativi ad ogni fornitore. Indicato con B l'evento "pezzo difettoso" abbiamo

$$P(B|A_1) = 0{,}05$$

$$P(B|A_2) = 0{,}02$$

$$P(B|A_3) = 0{,}01$$

La soluzione del quesito è data dalle determinazioni delle probabilità condizionate $P(A_i|B)$ per $i = 1{,}2{,}3$ ottenute mediante la formula di Bayes:

$$P(A_1|B) = \frac{0{,}05 \cdot 0{,}30}{0{,}05 \cdot 0{,}30 + 0{,}02 \cdot 0{,}50 + 0{,}01 \cdot 0{,}20} = \frac{0{,}015}{0{,}027} = 0{,}556$$

$$P(A_2|B) = \frac{0{,}02 \cdot 0{,}50}{0{,}027} = \frac{0{,}010}{0{,}027} = 0{,}370$$

$$P(A_3|B) = \frac{0{,}01 \cdot 0{,}20}{0{,}027} = \frac{0{,}002}{0{,}027} = 0{,}074$$

Le singole probabilità condizionate $P(A_i|B)$ "*a posteriori*" risultano tutte diverse dalle probabilità "*a priori*" $P(A_i)$:

$$P(A_1) = 0{,}30 \rightarrow P(A_1|B) = 0{,}556$$

$$P(A_2) = 0{,}50 \rightarrow P(A_2|B) = 0{,}370$$

$$P(A_3) = 0{,}20 \rightarrow P(A_3|B) = 0{,}074$$

Si osserva, da $P(A_i|B) \neq P(A_i)\ \forall i$, la non indipendenza stocastica tra gli eventi A_i e l'evento B. Risulta verificata, inoltre, la relazione $\sum_{i=1}^{g} P(A_i|B) = 1 \rightarrow 0{,}556 + 0{,}370 + 1{,}074 = 1$.

Osservazioni

- Lo studio delle probabilità degli eventi aleatori richiede l'individuazione, a seconda del fenomeno analizzato di uno spazio algebrico costituito dalla terna $\{\Omega, \mathcal{B}, \mathcal{P}\}$ in cui Ω è lo spazio degli eventi elementari cioè dei risultati possibili dell'esperimento aleatorio, $\mathcal{B}$ è la classe di Borel o σ-algebra ossia l'insieme di tutti gli eventi sottoinsiemi di Ω, compreso $\emptyset$ e Ω, classe chiusa rispetto alle operazioni di unione, intersezione e differenza, dove $\mathcal{P}$ è la distribuzione di probabilità che assegna una misura a ogni evento contenuto in $\mathcal{B}$.
- La formula di Bayes è stata presentata considerando partizioni di Ω costituite da un numero finito g di eventi mutualmente esclusivi ed esaustivi, ma è possibile generalizzare la formula anche nel caso di g infinito.
- Si può generalizzare le relazioni date per le condizioni di indipendenza al caso di più di due eventi, per es. di $g > 2$, avendosi

$$P\left(A_1 A_2 \cdots A_g\right) = P(A_1) \cdot P(A_2) \cdot \cdots \cdot P(A_g)$$

- Analogamente si può generalizzare la probabilità congiunta di $g > 2$ eventi mediante una sequenza di probabilità condizionate

$$\begin{aligned} P\left(A_1 A_2 \cdots A_g\right) \\ &= P(A_1) \cdot P(A_2 | A_1) \cdot P(A_3 | A_1 A_2) \cdot \cdots \\ &\cdot P(A_g | A_1 A_2 \cdots A_{g-1}) \end{aligned}$$

espressione nota come "legge delle probabilità composte".

11. Variabili casuali unidimensionali

Ricordando le tipiche fasi con cui è condotta una "ricerca statistica": a) esperimento; b) fenomeno ripetitivo; c) carattere di interesse; d) unità osservabile; e) rilevazione; f) raccolta dei risultati, Si rende, inoltre, necessario ricorrere a un modello che associ i vari eventi alla loro misura di verosimiglianza. Allo spazio probabilistico $\{\Omega, \mathcal{B}, \mathcal{P}\}$ che assomma in sé tali aspetti si affianca una variabile numerica X, che consideriamo in questa prima fase unidimensionale, ottenuta mediante un'opportuna applicazione che trasforma Ω, spazio degli eventi elementari, in un insieme di numeri reali in cui i risultati dell'esperimento si estrinsecano mediante la funzione $x(\cdot): \Omega \rightarrow \Re$. Si dispone così di un nuovo spazio detto "probabilizzato" definito dalla terna $\{\Re, \mathcal{B}_1, \mathcal{P}_X\}$

$$\{\Omega, \mathcal{B}, \mathcal{P}\} \Longrightarrow \{\Re, \mathcal{B}_1, \mathcal{P}_X\}$$

In cui $x(e) \in \Re$, con $e \in \Omega$, è una determinazione di un risultato casuale dell'esperimento ossia della "variabile casuale" (*v.c.*) unidimensionale *X*, ed è detta "immagine diretta" di e.

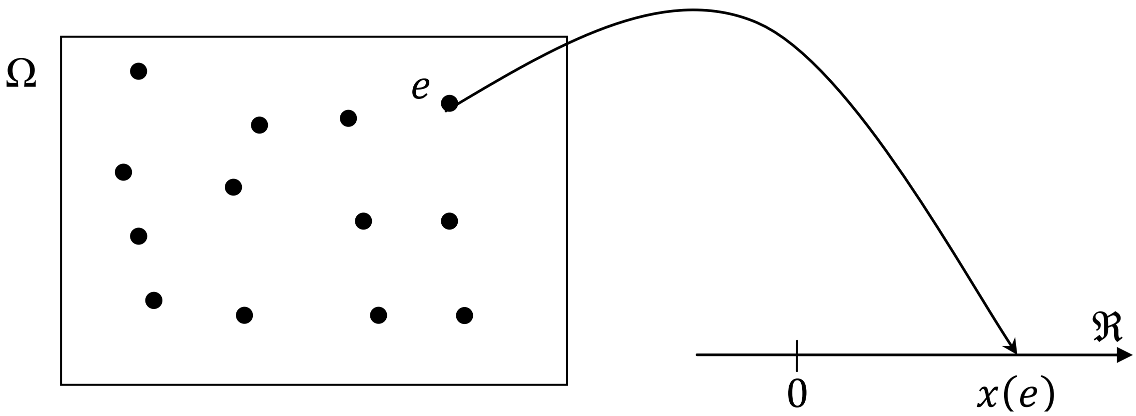
$\mathcal{B}_1$ è la classe di Borel (σ-algebra) generata sulla base degli intervalli $I = \{x: a < x \leq b, a < b \in \Re\}$ relativi allo spazio euclideo $\Re$, classe "chiusa" nei confronti delle operazioni di complemento (differenza), unione, intersezione, operazioni queste ultime al più numerabili, e inoltre, comprende anche $\Re$ e $\emptyset$.

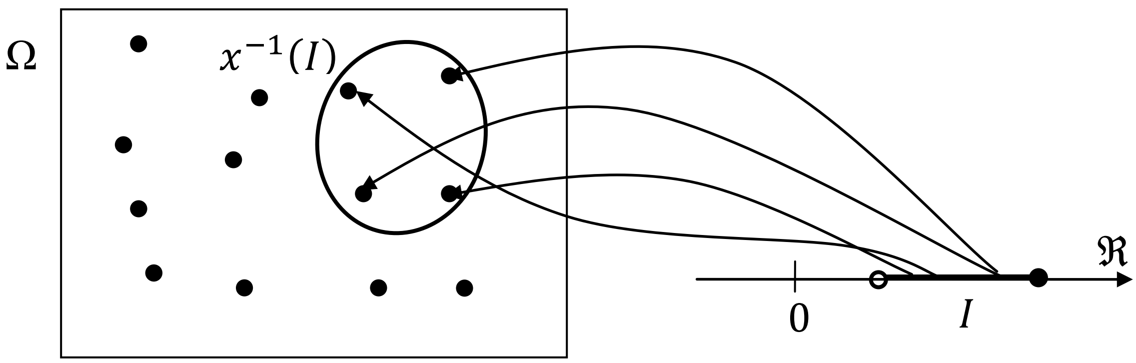
In corrispondenza a ogni sottoinsieme (evento) $A \supset \Omega$ si ha una trasformazione in un sottoinsieme $x(A) \supset \Re$ che è di A la sua "immagine diretta".

$\mathcal{P}_X$ è la "distribuzione di probabilità" associabile a ogni elemento di $\mathcal{B}_1$ definito, in particolare per ogni intervallo *I*, mediante la probabilità in $\mathcal{P}$ dell'immagine "inversa" di *I*.

$$P_X(I) \equiv P(x^{-1}(I) \supset \Omega)$$

con $x^{-1}(I) = \{e \in \Omega: \forall x(e) \in I\}$.



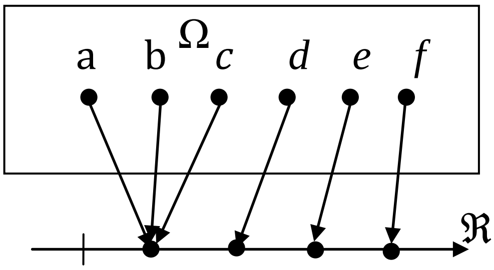


Esempio 6.

Consideriamo come esperimento casuale il lancio di un dado in cui le facce (a, b, c) portano il numero 1 e le altre facce (d, e, f) i numeri 2, 3, 4 rispettivamente.

- Unità sperimentale: faccia;
- Rilevazione: numero segnato sulla faccia;
- Corrispondenze:

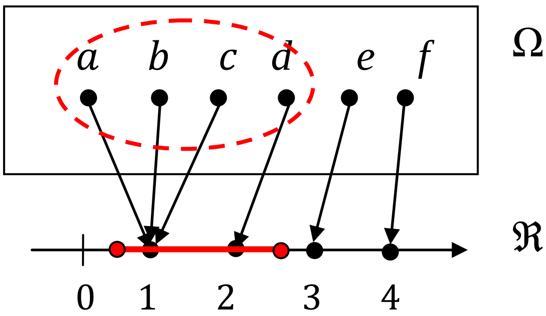
$$\text{faccia} \rightarrow \text{numero}\begin{cases} a, b, c \rightarrow 1 \\ \quad d \rightarrow 2 \\ \quad e \rightarrow 3 \\ \quad f \rightarrow 4 \end{cases}$$



0 1 2 3 4

Osservazioni

- Mentre la "rilevazione" $x(e)$ è una funzione degli elementi di Ω che vengono trasformati in un altro spazio S (nel nostro caso, in numeri appartenenti all'asse reale $\Re$), la "probabilità" è una funzione d'insieme, in termini di un valore reale $P(A) \in [0,1] \subset \Re$ per ogni evento A contenuto in Ω.
- L'immagine inversa si può definire, per qualunque sottoinsieme del codominio S, non solo quindi per i sottoinsiemi dell'immagine diretta di Ω: $x(\Omega)$. Riprendendo l'esempio precedente in cui $x(\Omega) = \{1,2,3,4\}$, si ha: $x^{-1}(\{1\}) = \{a, b, c\}$; $x^{-1}(\{2,3\}) = \{d, e\}$; $x^{-1}(\{x: 0{,}5 \leq x \leq 2{,}5\}) = \{a, b, c, d\}$.



- Lo spazio S può essere diverso dall'insieme dei numeri reali $\Re$, può essere anche di tipo "qualitativo", come avviene nel caso di estrazione di una carta da un mazzo in cui si è interessati al tipo di "seme".

12. Funzione di ripartizione di una variabile casuale unidimensionale

La misura di probabilità $P_X(\cdot)$ relativa agli elementi della classe di Borel $\mathcal{B}_1$ che interessano la *v.c.* X è data, in modo completo e univoco, dalla "Funzione di Ripartizione" (*F.d.R.*) $F: \mathfrak{R} \to [0,1]$ definita come:

$$F(x) = P_X(-\infty < X \leq x) \quad \text{per} \quad \forall x \in \mathfrak{R}$$

che dà la probabilità che la *v.c.* X presenti valori "non maggiori" (minori o uguali) del generico valore *x*.

Affinché una $F(\cdot)$ sia una *F.d.R.* deve soddisfare alle seguenti proprietà.

- $\lim_{x\to-\infty} F(x) = 0 \quad \lim_{x\to\infty} F(x) = 1$ (asintoticità).
- Se $a < b$ allora $F(a) \leq F(b)$ (monotonicità non decrescente).
- $F(\cdot)$ può presentare "salti" (discontinuità del primo tipo) ma è continua "a destra".
  Abbiamo infatti

$$\lim_{x\to x_0^-} F(x) = F(x_0) - P_X(x_0) \begin{matrix} \nearrow = F(x_0) \\ \searrow \neq F(x_0) \end{matrix}$$

$$\lim_{x\to x_0^+} F(x) = F(x_0)$$

  I punti di discontinuità costituiscono un insieme al più numerabile.

La *F.d.R.* di una *v.c.* X presenta "salti" nei punti di discontinuità che di alternano con "tratti" di continuità.

Qualora la *F.d.R.* si presenti continua la probabilità relativa ad ogni valore x_0 risulta infinitesima e quindi è pari a $P_X(X = x_0) = 0$ pur essendo x_0 un valore "possibile" come risultato dell'esperimento, mentre l'elemento della classe di Borel $\emptyset$ ha anch'esso probabilità nulla ma dal punto di vista osservativo è "impossibile".

Negli esempi dati si è già evidenziata la tipica struttura della *F.d.R.* di una *v.c.* X e, generalizzando, possiamo dire: ogni *F.d.R.* può scomporsi, in modo univoco nella somma di tre funzioni:

$$F(x) = F_d(x) + F_c(x) + F_s(x)$$

dove

$$F_d(x) = \sum\nolimits_{i:x_i \leq x} P_X(\{x_i\})$$

essendo $\{x_i\}$, l'insieme di punti di discontinuità di F, è la componente "discreta";

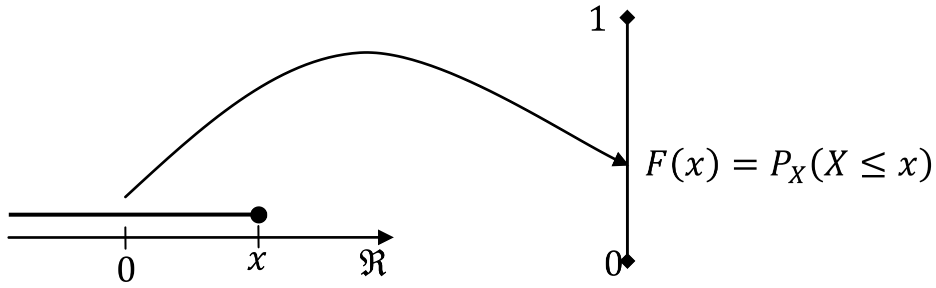
$$F_c(x) = \int_{-\infty}^{x} f(t)dt$$

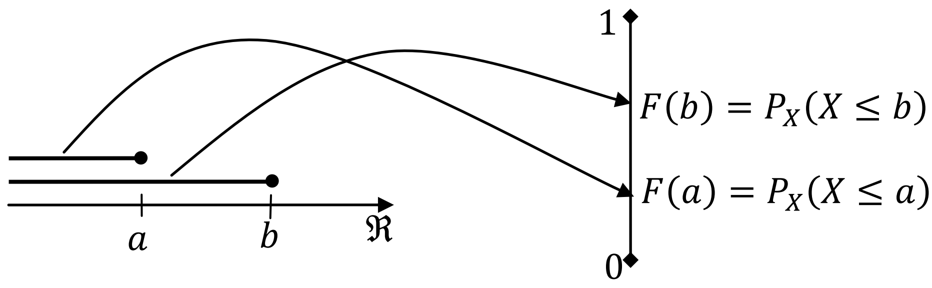
in cui $f(t) \geq 0 \ \forall x \in \Re$, è la componente "assolutamente continua";

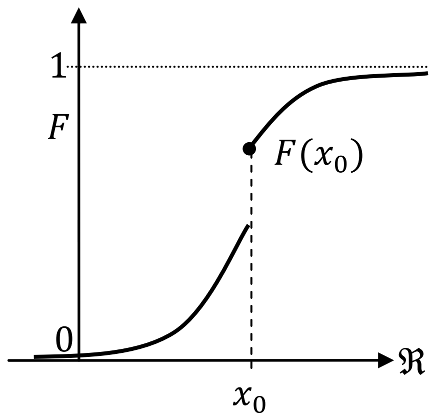
$F_s(x)$ è la componente "singolare" che pur essendo continua non può esprimersi come integrale di una funzione tipo $f(t)$.

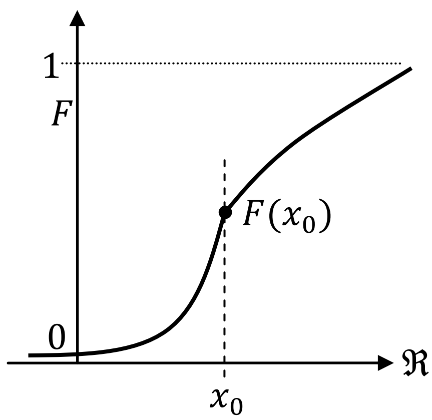
Le tre funzioni $F_d(\cdot), F_c(\cdot), F_s(\cdot)$ hanno un andamento simile a quanto detto per $F(\cdot)$ ad esclusione del fatto che il loro limite per $x \to \infty$ è minore di uno, essendo pari a uno solo il limite della loro somma.

Nel seguito considereremo, per semplicità e per la rilevante importanza in ambito applicativo solo *v.c.* ad una sola componente: a) "discreta"; b) "assolutamente continua"









Con la *F.d.R.* $F(x)$ siamo in grado di assegnare a tutti gli elementi della classe di Borel $\mathcal{B}_1$, generati a partire dagli intervalli, una probabilità nel rispetto degli assiomi del calcolo delle probabilità.

In particolare sia $I = (a, b]$ un intervallo generico, abbiamo

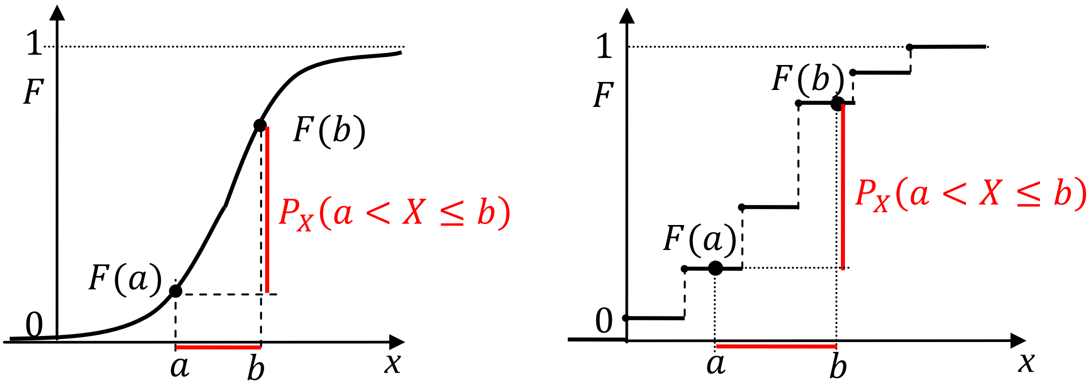
$$P_X(X \in I) = P_X(a < X \leq b) = F(b) - F(a) \text{ per } \forall a \leq b$$

L'intervallo $(a, b]$ si può interpretare come differenza di due insiemi $S_a = (X \leq a)$ e $S_b = (X \leq b)$ che corrispondono in $\Re$ alle due semirette $(-\infty, a]$ e $(-\infty, b]$, rispettivamente. Dai teoremi delle probabilità abbiamo: $P_X(S_b - S_a) = P_X(S_b) - P_X(S_a \cdot S_b)$, ma $S_a \cdot S_b = S_a$ in quanto $S_a \subset S_b$, allora $P_X(S_b - S_a) = P_X(S_b) - P_X(S_b)$. Ricordando che:

$$\begin{cases} P_X(S_a) = P_X(X \le a) = F(a) \\ P_X(S_b) = P_X(X \le b) = F(b) \end{cases}$$

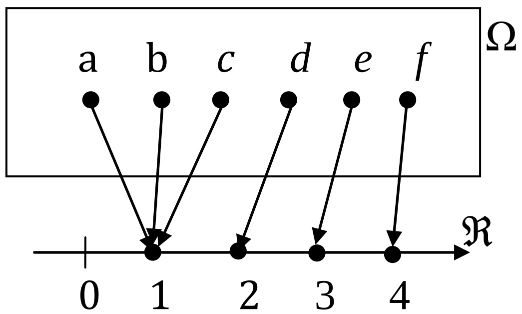
otteniamo quindi

$$P_X(a < X \le b) = P_X(S_b) - P_X(S_a) = F(b) - F(a)$$



<u>Esempio 7.</u>

Si riprenda l'esempio 6. in cui la *v.c.* X proveniva da un esperimento casuale dato dal lancio di un dado con eventi elementari equiprobabili.



La *F.d.R.* risulta

$$\begin{array}{lll} \text{Per } -\infty < x < 1 & \rightarrow & F(x) = 0 \\ \text{Per } 1 \le x < 2 & \rightarrow & F(x) = 1/2 \\ \text{Per } 2 \le x < 3 & \rightarrow & F(x) = 2/3 \end{array}$$

$$\text{Per } 3 \leq x < 4 \qquad \rightarrow \quad F(x) = 5/6$$
$$\text{Per } 4 \leq x < \infty \qquad \rightarrow \quad F(x) = 1$$

<u>Esempio 8.</u>

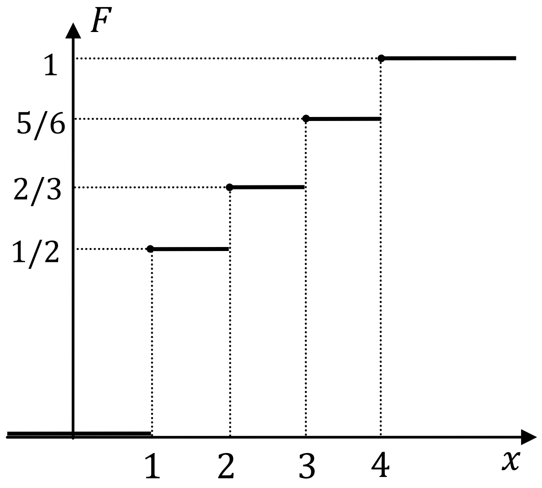
Sia data la seguente funzione:

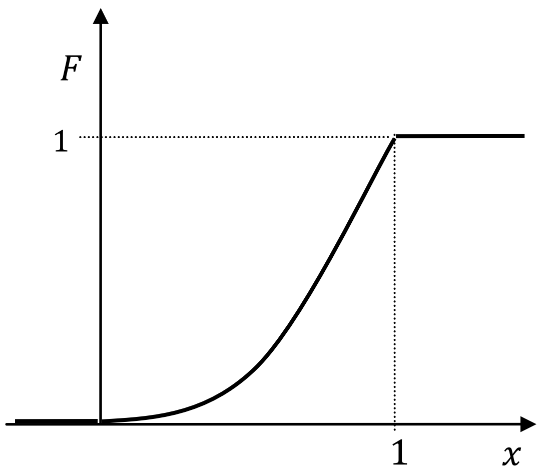
$$F(x) = \begin{cases} 0 & per \ -\infty < x < 0 \\ x^2 & per \ 0 \leq x < 1 \\ 1 & per \ 1 \leq x < \infty \end{cases}$$

Essa può rappresentare la *F.d.R.* di una *v.c.* X, infatti: $\lim_{x \to -\infty} F(x) = 0$, $\lim_{x \to \infty} F(x) = 1$, è "non decrescente" ed è anche continua.

E' possibile determinare la probabilità di un generico intervallo, come ad esempio:

$$P_X(0{,}5 < x \leq 0{,}7) = F(0{,}7) - F(0{,}5) = 0{,}49 - 0{,}25 = 0{,}24$$





13. Variabili casuali unidimensionali discrete

Una *v.c.* unidimensionale discreta X è definita in un insieme $S = \{x_1, x_2, \cdots, x_i, \cdots\}$ costituito da un insieme finito o una infinità numerabile di valori con probabilità diverse da zero.

Si dovrà verificare

$$P_X(S) = \sum_i P_X(X = x_i) = \sum_i p_i = 1$$

I valori $\{x_1, x_2, \cdots, x_i, \cdots\}$ sono tra loro disgiunti, ne consegue che per ogni insieme S' disgiunto da S avremo $S \cap S' = \emptyset$, quindi $P_X(S \cup S') = P_X(S) + P_X(S') = 1$ da cui si ha $P_X(S') = 0$. Quindi la *v.c.* X presenta probabilità nulla di manifestarsi in S'.

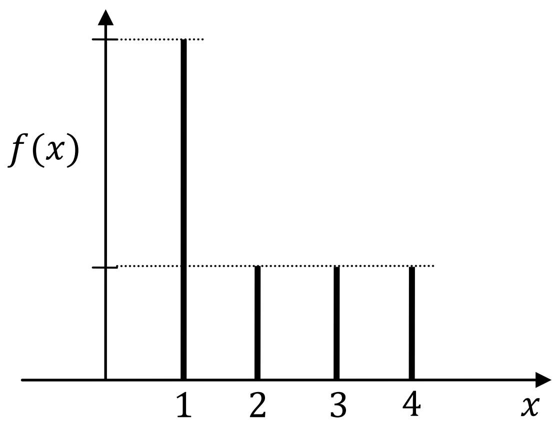
Oltre alla *F.d.R.*, data da $F(x) = \sum_{x_i \leq x} P_X(X = x_i) = \sum_{x_i \leq x} p_i$, si definisce anche una funzione $f(x)$ (o $p(x)$) detta "funzione di probabilità" (*f.d.p.*) data da

$$f(x) = \begin{cases} p_i & \text{per } x = x_i \\ 0 & \text{per } x \neq x_i \end{cases} \quad \text{per } x \in \Re$$

Quindi la *F.d.R.* di una *v.c.* discreta può esprimersi come:

$$F(x) = \sum_{x_i \leq x} f(x_i)$$

Un esempio di *v.c.* discreta è stato dato negli esempi 6. e 7 e la funzione di probabilità ha l'andamento del grafico sottostante.



14. Variabili casuali unidimensionali continue

Consideriamo specificatamente una *v.c.* X "assolutamente continua, la cui *F.d.R.* si deve presentare le seguenti proprietà:

- $F(x)$ è una funzione continua;
- Esiste una funzione $f(x) \geq 0$ per $x \in \Re$ tale che

$$F(x) = \int_{-\infty}^{x} f(t)dt$$

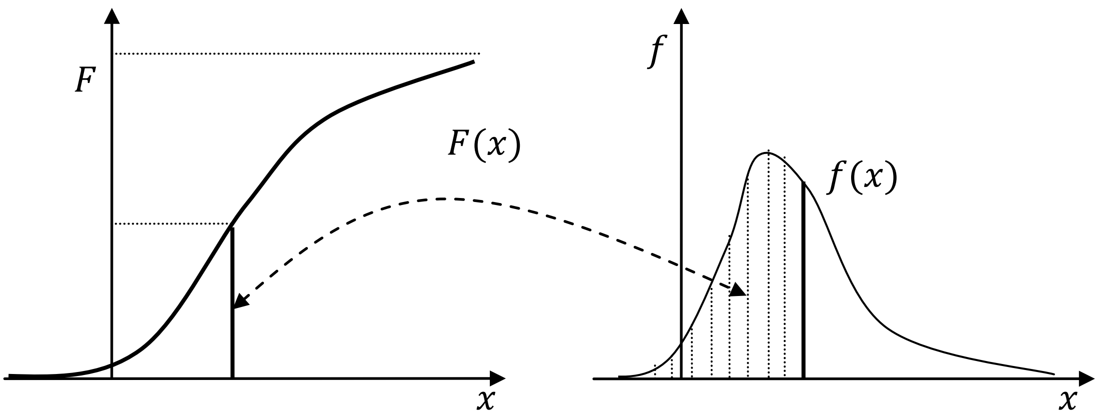
Nei punti in cui $f(x)$ è continua si ha

$$\frac{dF}{dx} = f(x)$$

La funzione $f(x)$ può presentare punti di discontinuità in un numero finito o al più un'infinità numerabile.

La funzione $f(x)$ è detta "funzione di densità di probabilità" o semplicemente "funzione di densità" (*f.d.d.*) della *v.c.* X e deve soddisfare le seguenti condizioni:

$$f(x) \geq 0 \text{ per } x \in \Re \quad \text{e} \quad \int_{-\infty}^{\infty} f(t)dt = 1$$



<u>Esempio 9.</u>

Riprendiamo l'esempio 8.. E si voglia verificare che la *v.c.* X è assolutamente continua.

$$F(x) = \begin{cases} 0 & per \ -\infty < x < 0 \\ x^2 & per \ 0 \leq x < 1 \\ 1 & per \ 1 \leq x < \infty \end{cases}$$

Assumiamo $f(x) = dF/dx$ nei diversi campi di definizione:

$$f(x) = \begin{cases} 0 & per \ -\infty < x < 0 \\ 2x & per \ 0 \le x < 1 \\ 0 & per \ 1 \le x < \infty \end{cases}$$

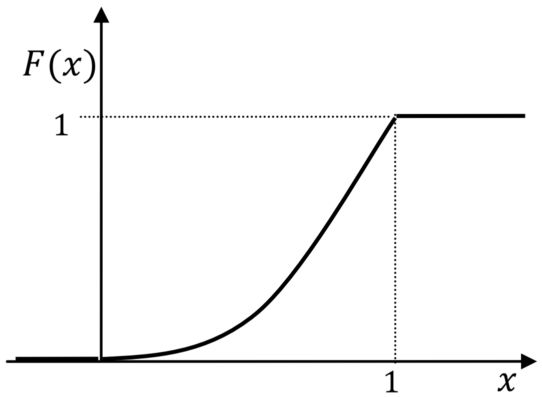
1. La $F(x)$ è ovunque continua.
2. Verifichiamo che $F(x) = \int_{-\infty}^{x} f(t)dt$
   - Se $x < 0$, allora $f(x) = 0 \Longrightarrow F(x) = 0$;
   - Se $x \in [0,1)$, allora

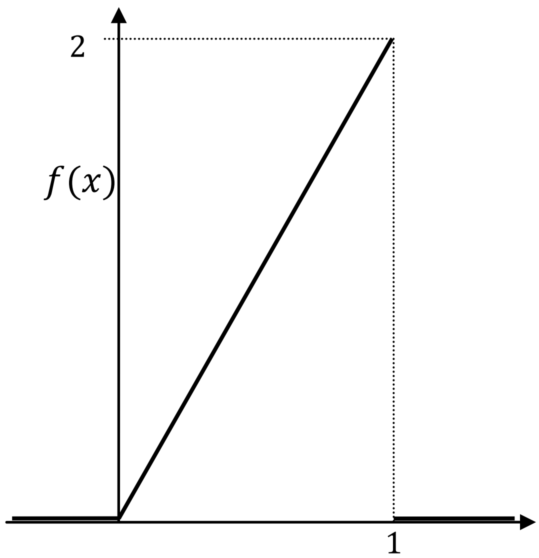
$$F(x) = \int_{-\infty}^{x} f(t)dt = \int_{-\infty}^{0} 0 \cdot dt + \int_{0}^{x} 2t \cdot dt = |t^2|_0^x = x^2$$

   - Se $x \ge 1$, allora

$$F(x) = \int_{-\infty}^{x} f(t)dt = \int_{-\infty}^{0} 0 \cdot dt + \int_{0}^{1} 2t \cdot dt + \int_{1}^{x} 0 \cdot dt = |t^2|_0^1 = 1$$

Quindi la *v.c.* X è assolutamente continua con *f.d.d.* pari a $f(x)$. La *F.d.R.* $F(x)$ non ha derivata in $x = 1$ e non si può determinare $f(x) = dF/dx$ in $x = 1$. La $f(x)$ presenta un punto di discontinuità in $x = 1$.





Osservazioni

- Una *v.c.* X assolutamente continua ha la *F.d.R.* $F(x)$ derivabile ovunque a meno di un numero di punti numerabile.
- Tali punti si presentano con associata una probabilità pari a "zero".
- In generale la *f.d.d.* $f(x) = dF/dx$ permette di definire un "differenziale di probabilità" $dF = f(x)dx = F(x + dx) - F(x) = P_X(x < X \leq x + dx)$.
- Se Δx è un intervallo "piccolo": $\Delta x \cong dx$ allora $\Delta F \cong f(x)\Delta x$.
- Per una *v.c.* X assolutamente continua, si ha

  $$P_X(a < X \leq b) = P_X(a < X < b) = P_X(a \leq X < b) = P_X(a \leq X \leq b)$$

  in quanto $P_X(a) = 0$ e $P_X(b) = 0$. Infatti, ad esempio, si dimostra che $P_X(b) = 0$, in quanto $\lim_{x \to b} P_X(x < X \leq b) = \lim_{x \to b}[F(b) - F(x)] = \lim_{x \to b} \int_x^b f(t)dt = P_X(b)$, ma per una *v.c.* X assolutamente continua $\lim_{x \to b} F(x) = F(b)$ e quindi $P_X(b) = \lim_{x \to b} P_X(x < X \leq b) = F(b) - F(b) = 0$.

- A ogni evento $X = a$ corrisponde una probabilità "nulla" anche se non si tratta di un evento "impossibile": $P_X(X = a) = 0$.
- Analogamente abbiamo per $P_X(\Re - \{a\}) = 1$ che corrisponde a un evento con probabilità "uno" senza che si tratti di un evento "certo"

  $$P_X(\Re - \{a\}) = P_X(\Re) - P_X\{a\} = 1 - 0 = 1$$

15. Variabile casuale come funzione di una variabile casuale

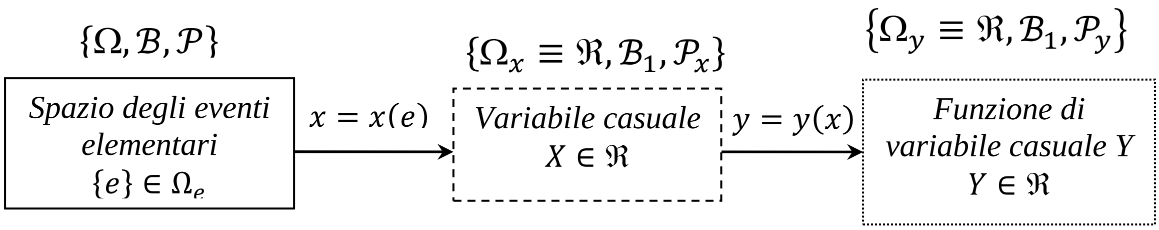
Considerato lo spazio Ω degli eventi elementari e abbiamo ottenuto una *v.c.* X mediante la trasformazione $x(e): \Omega \to \Re \; \forall e \in \Omega$. Questo ci ha permesso di disporre un nuovo spazio probabilistico $\{\Omega_x \equiv \Re, \mathcal{B}_1, \mathcal{P}_x\}$ derivato dall'originario spazio $\{\Omega, \mathcal{B}, \mathcal{P}\}$, in cui la distribuzione di probabilità $\mathcal{P}_x$ è ottenuta mediante l'immagine inversa $x^{-1}(S_x \in \Re)$ di ogni insieme S_x appartenente alla σ-algebra $\mathcal{B}_1$ come:

$$P_X(S_x) = P(x^{-1}(S_x \subset \Re))$$

Abbiamo anche visto come la distribuzione di probabilità $P_X(\cdot)$ possa sintetizzarsi mediante la *F.d.R.* di *X* data da

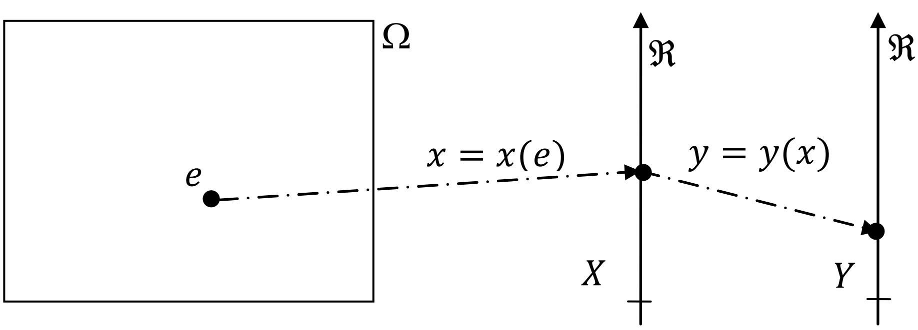
$$F_x(x) = P_X(X \in (-\infty, x])$$

Si introduca ora la trasformazione da $\Re$ in $\Re$, mediante la funzione $y(x)$ in modo da disporre di una nuova *v.c.* *Y* a cui si associa uno spazio probabilistico completo $\{\Omega_y \equiv \Re, \mathcal{B}_1, \mathcal{P}_y\}$ legato a quello della *v.c.* *X* in termini di distribuzione di probabilità mediante l'immagine inversa $y^{-1}(S_y \in \Re)$ di ogni insieme S_y appartenente alla σ-algebra $\mathcal{B}_1$ come $P_Y(S_y) = P_X(y^{-1}(S_y \subset \Re))$.



La *F.d.R.* di *Y* si ottiene dalla distribuzione di probabilità $P_X(\cdot)$ mediante l'immagine inversa $F_Y(y) = P_X(y^{-1}((-\infty, y]))$.

Si osservi che essendo la trasformazione $X \to Y$ fatta su "dominio" e "codominio" identici $\Re$ non si sono distinte le corrispondenti σ-algebre, indicate entrambe con $\mathcal{B}_1$.



Esempio 10.

Consideriamo la trasformazione lineare $Y = y(X) = aX + b$ e indichiamo con $F_X(x)$ e $F_Y(y)$ le *F.d.R.* di *X* e *Y*, abbiamo

$$F_Y(y) = P_Y(Y \leq y) = P_X(aX + b \leq y) = \begin{cases} P_X\left(X \leq \dfrac{y-b}{a}\right) \text{se } a > 0 \\ P_X\left(X \geq \dfrac{y-b}{a}\right) \text{se } a < 0 \end{cases}$$

Per $a > 0$ abbiamo

$$P_X\left(X \leq \frac{y-b}{a}\right) = F_X\left(\frac{y-b}{a}\right)$$

Per $a < 0$ abbiamo

$$\begin{aligned} P_X\left(X \geq \frac{y-b}{a}\right) &= 1 - P_X\left(X < \frac{y-b}{a}\right) \\ &= 1 - \left[F_X\left(\frac{y-b}{a}\right) - P_X\left(X = \frac{y-b}{a}\right)\right] \end{aligned}$$

Quindi in generale, abbiamo

$$F_Y(y) = \begin{cases} F_X\left(\dfrac{y-b}{a}\right) & \text{se } a > 0 \\ 1 - F_X\left(\dfrac{y-b}{a}\right) + P_X\left(X = \dfrac{y-b}{a}\right) & \text{se } a < 0 \end{cases}$$

Qualora la *v.c. X*, di conseguenza anche la *Y*, è assolutamente continua abbiamo $P_X\left(X = \frac{y-b}{a}\right) \equiv P_Y(Y = y) \equiv 0$ per ogni valore di *y*.

$$F_Y(y) = \begin{cases} F_X\left(\dfrac{y-b}{a}\right) = \displaystyle\int_{-\infty}^{\frac{y-b}{a}} f_X(t)dt & \text{se } a > 0 \\ 1 - F_X\left(\dfrac{y-b}{a}\right) = 1 - \displaystyle\int_{-\infty}^{\frac{y-b}{a}} f_X(t)dt & \text{se } a < 0 \end{cases}$$

essendo $f_X(t)$ la *f.d.d.* della *v.c.* X. La $F_Y(y)$ è derivabile e quindi si ha la *f.d.d.* della *v.c.* Y data da

$$\frac{dF_Y}{dy} = f_Y(y) = \begin{cases} \frac{1}{a} f_X\left(\frac{y-b}{a}\right) & \text{se } a > 0 \\ -\frac{1}{a} f_X\left(\frac{y-b}{a}\right) & \text{se } a < 0 \end{cases}$$

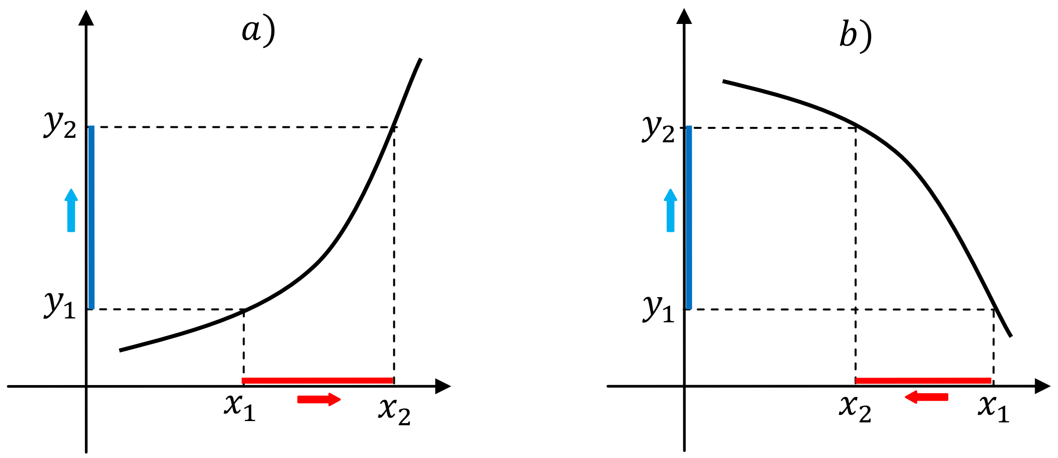
Risultando, in entrambe le situazioni, $f_Y(y) \geq 0$, potendosi scrivere, in forma compatta, come

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

confermando, che la Y è una *v.c.* “assolutamente continua.

I risultati ottenuti nel caso di *v.c.* continue e di una trasformazione lineare è generalizzabile qualora si consideri una trasformazione $X \rightarrow Y$ “monotona in senso stretto” e quindi “biettiva”.

Sia $y = y(x)$ una funzione reale in $\Re \rightarrow \Re$, derivabile con derivata continua, possiamo considerare simultaneamente le due situazioni: a) con $y(x)$ “crescente”; b) con $y(x)$ “crescente”.



Esistendo biunivocità tra x e y e univocità di percorso degli assi, abbiamo

$$P_Y(y_1 \leq Y \leq y_2) = \begin{cases} P_X(x_1 \leq X \leq x_2) = \int_{x_1}^{x_2} f_X(x)dx & a) \\ P_X(x_2 \leq X \leq x_1) = \int_{x_2}^{x_1} f_X(x)dx & b) \end{cases}$$

Esprimendo le "*x*" mediante le "*y*", tramite $y = y(x)$ e la sua funzione inversa $x = y^{-1}(y) = x(y)$

$$x_1 \Rightarrow y_1, x_2 \Rightarrow y_2, f_X(x) = f_X(x(y))$$

con differenziale $dx = x'(y)dy$, abbiamo

$$P_Y(y_1 \leq Y \leq y_2) = \begin{cases} \int_{y_1}^{y_2} f_X[x(y)][x'(y)]dy & a) \\ \int_{y_2}^{y_1} f_X[x(y)][x'(y)]dy = -\int_{y_2}^{y_1} f_X[x(y)][x'(y)]dy & b) \end{cases}$$

con $x'(y) > 0$ e $x'(y) < 0$ rispettivamente nella situazione *a*) e *b*), da cui possiamo scrivere sinteticamente

$$P_Y(y_1 \leq Y \leq y_2) = \int_{y_1}^{y_2} f_X[x(y)]|x'(y)|dy$$

Se si pone $y_1 \to -\infty$, $y_2 \to y$ e $y \to t$ si ottiene la *F.d.R.* della *v.c. Y*

$$F_Y(y) = \int_{-\infty}^{y} \{f_X[x(t)]|x'(t)|\}dt \qquad (*)$$

Dove la funzione integranda è $f_X[x(t)]|x'(t)| \geq 0$ rappresenta la *f.d.d.* della *v.c.* di *Y*, di conseguenza *Y* è una *v.c.* "assolutamente continua"

$$f_Y(y) = f_X[x(y)]|x'(y)| \qquad (**)$$

Esempio 11.

Sia X una *v.c.* assolutamente continua con *f.d.d.*

$$f_X(x) = \begin{cases} 0{,}5 & \text{per} -1 < x \leq 1 \\ 0 & \text{altrove} \end{cases}$$

Consideriamo la *v.c.* Y, funzione di X pari a $Y = y(X) = X^2$. Essendo $y(x)$ funzione monotona a tratti nel dominio che interessa ($-1 < x \leq 1$) conviene procedere mediante le immagini inverse degli eventi.

La *F.d.R.* di *Y* è data da:

$$F_Y(y) = \begin{cases} 0 & \text{per } y < 0 \\ P_Y(0 < Y \leq y) & \text{per } 0 \leq y \leq 1 \\ 1 & \text{per } y > 1 \end{cases}$$

In particolare per $0 \leq y \leq 1$ abbiamo $Y \leq y \Rightarrow -\sqrt{y} \leq X \leq \sqrt{y}$ e quindi

$$P_Y(0 < Y \leq y) = P_X\left(-\sqrt{y} \leq X \leq \sqrt{y}\right) = F_X\left(\sqrt{y}\right) - F_X\left(-\sqrt{y}\right)$$

Essendo, per $-1 \leq x \leq 1$, $F_X(x) = \int_{-1}^{x} 0{,}5\, dx = 0{,}5(x+1)$, abbiamo

$$\begin{aligned} F_Y(y) &= F_X\left(\sqrt{y}\right) - F_X\left(-\sqrt{y}\right) = 0{,}5\left(\sqrt{y}+1\right) - 0{,}5\left(-\sqrt{y}+1\right) \\ &= \sqrt{y} \quad \text{per } 0 \leq y \leq 1 \end{aligned}$$

Con *f.d.d.* pari a

$$f_Y(y) = \begin{cases} 0{,}5/\sqrt{y} & \text{per } 0 < y \leq 1 \\ 0 & \text{altrove} \end{cases}$$

Si lascia al lettore la rappresentazione grafica delle funzioni: $y(x)$, $f_X(x)$, $F_X(x)$, $f_Y(y)$, $F_Y(y)$.

Esempio 12.

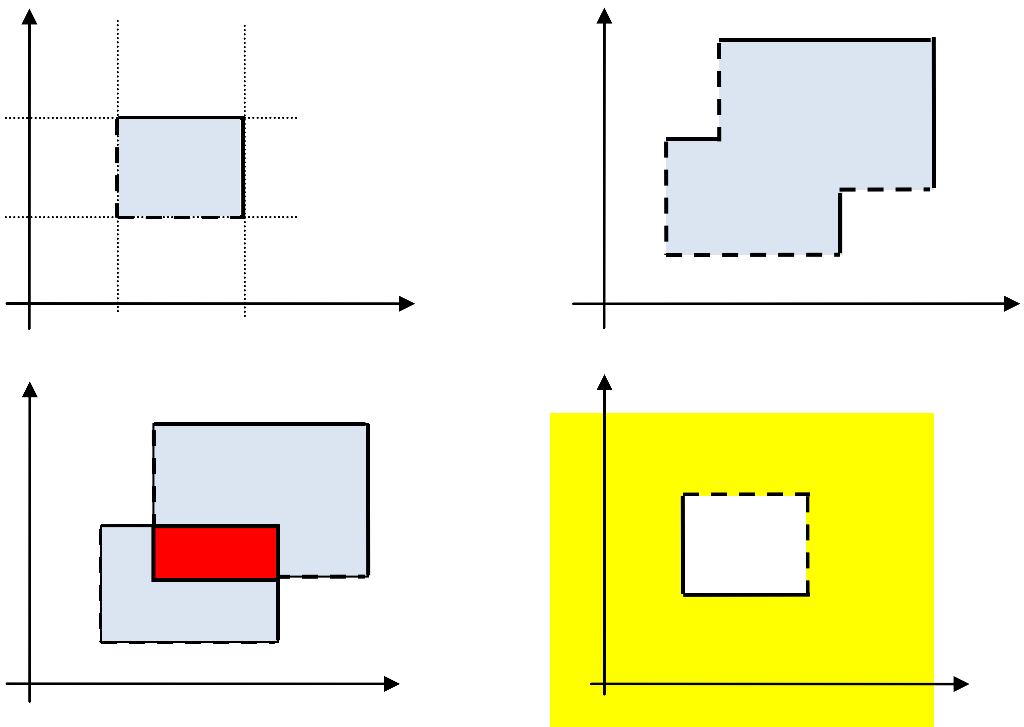
Si consideri la *v.c.* X dell'esercizio 11. e si determinino le *F.d.R. e f.d.d.* della *v.c.* $Y = y(X) = X^3$ mediante le relazioni (*) e (**), tracciandone i corrispondenti grafici.

16. Variabili casuali bidimensionali

Come *v.c.* "bidimensionale" $\boldsymbol{X} = (X_1, X_2)$, oppure semplicemente (X, Y), intendiamo lo "spazio probabilistico" definito dalla terna $\{\Re^2, \mathcal{B}_2, \mathcal{P}_{\boldsymbol{X}}\}$, dove la σ-algebra $\mathcal{B}_2$ è relativa allo spazio euclideo $\Re^2$ ed è generato dalla classe di intervalli tipo $\boldsymbol{I}$ (rettangoli ≡ prodotto cartesiano degli intervalli delle due componenti)

$$\boldsymbol{I} = \{(x_1, x_2): a' < x_1 \leq b'; a'' < x_2 \leq b''\} = (a', b'] \times (a'', b'']$$

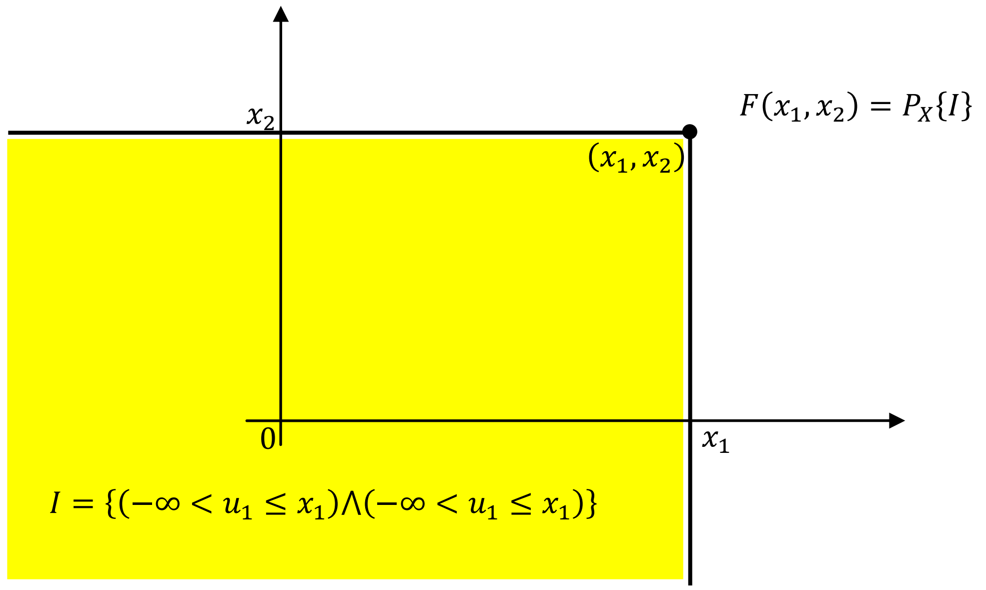
dove $a' \leq b', a'' \leq b'' \in \Re$. Lo spazio è completato mediante le operazioni differenza (complemento), unione, intersezione tra gli elementi di $\mathcal{B}_2$, al più numerabili.



Analogamente a quanto fatto nel caso di *v.c.* unidimensionale si utilizza, per probabilizzare $\boldsymbol{X} = (X_1, X_2)$, la *F.d.R.* $F(x_1, x_2): \Re^2 \to [0,1]$, data da

$$F(x_1, x_2) = P_X\{(u_1, u_2) \in \Re^2: (-\infty < u_1 \leq x_1) \land (-\infty < u_2 \leq x_1)\}$$

che costituisce una "regione angolare" in $\Re^2$.



Dalla *F.d.R.* della *v.c.* bidimensionale $\boldsymbol{X} = (X_1, X_2)$ è possibile ottenere le *F.d.R.* delle *v.c.* componenti unidimensionali X_1 e X_2, $F_1(x_1)$ e $F_2(x_2)$, rispettivamente.

$$F_1(x_1) = P_X\{(u_1, u_2) \in \Re^2: (u_1 \leq x_1) \land (-\infty < u_2 < \infty)\}$$
$$= F(x_1, \infty)$$

$$F_2(x_2) = P_X\{(u_1, u_2) \in \Re^2: (-\infty < u_1 \leq x_1) \land (u_2 \leq x_2)\}$$
$$= F(\infty, x_2)$$

Similmente a quanto è stato indicato per la *F.d.R.* di una *v.c.* unidimensionale è possibile caratterizzare la *F.d.R.* nel caso bidimensionale.

1. *Comportamento asintotico*

$$\lim_{\substack{x_1 \to \infty \\ x_2 \to \infty}} F(x_1, x_2) = 1$$

$$\lim_{x_1 \to -\infty} F(x_1, x_2) = 0 \;\text{ per } \forall\; x_2; \lim_{x_2 \to -\infty} F(x_1, x_2) = 0 \;\text{ per } \forall\; x_1$$

da cui

$$\lim_{\substack{x_1 \to -\infty \\ x_2 \to -\infty}} F(x_1, x_2) = 0$$

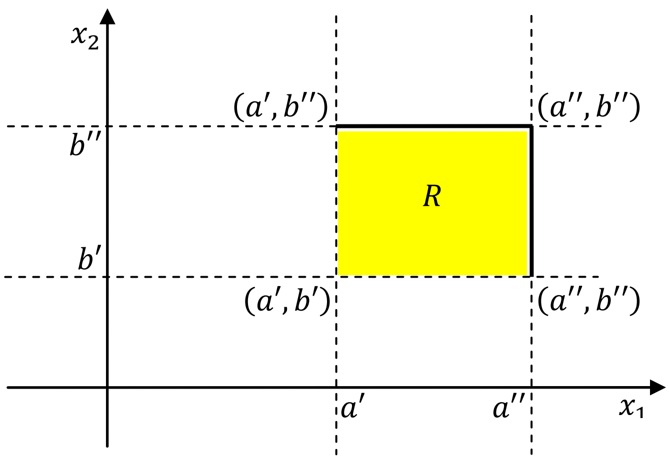
2. *Monotonicità*
   Se $a' \leq b'$ e $a'' \leq b''$ allora $F(a', a'') \leq F(b', b'')$. Infatti essendo

$$F(a', a'') = P_{\boldsymbol{X}}(S_a) = P_{\boldsymbol{X}}\{(-\infty, a'] \times (-\infty, a'']\}$$

$$F(b', b'') = P_{\boldsymbol{X}}(S_b) = P_{\boldsymbol{X}}\{(-\infty, b'] \times (-\infty, b'']\}$$

   si ha $S_a \subset S_b$, di conseguenza $P_{\boldsymbol{X}}(S_a) \leq P_{\boldsymbol{X}}(S_b)$ e quindi $F(a', a'') \leq F(b', b'')$.

3. *Condizione di Positività*
   Consideriamo il generico "rettangolo" definito come prodotto cartesiano $R = (a', a''] \times (b', b'']$



Oltre a R, consideriamo le seguenti regioni del piano:

$A = (-\infty, a''] \times (-\infty, b'']$, $B = (-\infty, a'] \times (-\infty, b']$,

$C = (-\infty, a'] \times (-\infty, b''], D = (-\infty, a''] \times (-\infty, b'],$

$E = (a', a''] \times (-\infty, b'],\ \ F = (-\infty, a'] \times (b', b'']$

Tra tali regioni del piano esistono le seguenti relazioni di unione e intersezione:

$$A = R \cup B \cup E \cup F$$

con i componenti a destra tra loro disgiunti e quindi

$$P_X\{A\} = P_X\{R\} + P_X\{B\} + P_X\{E\} + P_X\{F\} \quad \text{da cui } P_X\{R\} = P_X\{A\} - P_X\{B\} - P_X\{E\} - P_X\{F\} \quad (*)$$

Inoltre

$$C = B \cup F \ \text{ ed essendo } B \cap F = \emptyset$$

abbiamo

$$P_X\{C\} = P_X\{B\} + P_X\{F\} \ \text{ da cui } P_X\{F\} = P_X\{C\} - P_X\{B\} \quad (**)$$

Similmente

$$D = B \cup E \ \text{ ed essendo } B \cap E = \emptyset$$

abbiamo

$$P_X\{D\} = P_X\{B\} + P_X\{E\} \ \text{ da cui } = P_X\{D\} - P_X\{B\} \quad (***)$$

Sostituendo nella eq. (*) le espressioni di $P_X\{E\}$ e $P_X\{F\}$ date , rispettivamente dalle (**) e (***) si ottiene

$$\begin{aligned} P_X\{R\} &= P_X\{A\} - P_X\{B\} - P_X\{E\} - P_X\{F\} \\ &= P_X\{A\} - P_X\{B\} - [P_X\{C\} - P_X\{B\}] \\ &\quad - [P_X\{D\} - P_X\{B\}] \\ &= P_X\{A\} + P_X\{B\} - P_X\{C\} - P_X\{D\} \end{aligned}$$

Essendo le regioni *A*, *B*, *C*, *D* “angolari” e di conseguenza le relative probabilità sono date mediante la *F.d.R.* della *v.c.* $\boldsymbol{X} = (X_1, X_2)$, abbiamo

$$P_{\boldsymbol{X}}\{R\} = F(a'', b'') + F(a', b') - F(a', b'') - F(a'', b')$$

Tale relazione permette di probabilizzare ogni “rettangolo” in $\Re^2$ e quindi ogni elemento della σ-algebra $\mathcal{B}_2$, risultando la probabilità di *R* pari a

$$F(a'', b'') + F(a', b') - F(a', b'') - F(a'', b') \geq 0$$

per ogni scelta di $(a' \leq a'', b' \leq b'' \in \Re)$.

4. *Condizione di Continuità*

La *F.d.R* $F(x_1, x_2)$ è sempre “continua a destra” rispetto alle due componenti, può presentare punti di “discontinuità” del 1° tipo, i punti di discontinuità sono al più una infinità numerabile.

In generale, dato un punto (x_{10}, x_{20}), abbiamo

$$\lim_{\substack{x_1 \to x_{10}^+ \\ x_2 \to x_{20}^+}} F(x_1, x_2) = F(x_{10}, x_{20})$$

5. *Decomposizione della funzione* $F(x_1, x_2)$

Anche per le *v.c.* bidimensionali (o bivariate) è valido un teorema di decomposizione ossia $F(x_1, x_2)$ può definirsi, in modo univoco, mediante la somma di tre funzioni

$$F(x_1, x_2) = F_d(x_1, x_2) + F_c(x_1, x_2) + F_s(x_1, x_2)$$

- dove $F_d(x_1, x_2)$ è una funzione “a gradini”

$$F_d(x_1, x_2) = \sum_{k: x_{1k} \leq x_1, x_{2k} \leq x_2} P_{\boldsymbol{X}}(x_{1k}, x_{2k})$$

essendo $(x_{1k}, x_{2k}) \in S$ un generico punto di discontinuità della $F(x_1, x_2)$ e S è l'insieme di tali punti che viene a costituire una "nube" di punti in $\Re^2$;

- dove $F_c(x_1, x_2)$ è una funzione "assolutamente continua" che si può esprimere mediante il seguente l'integrale

$$F_c(x_1, x_2) = \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} f(u_1, u_2) du_2 du_1$$

essendo $f(\cdot,\cdot) \geq 0$ è la *f.d.d.* congiunta;

- dove $F_s(x_1, x_2)$ è una funzione "continua" ma non esprimibile mediante l'integrale di una funzione e rappresenta la componente "singolare" (non assolutamente continua).

Le tre funzioni F_d, F_c, F_s hanno andamenti simili alla *F.d.R.* $F(x_1, x_2)$ solo che il limite per x_1 e x_2 tenenti a ∞, di tali funzioni non è pari a 1.

Nel seguito considereremo solo *v.c.* aventi una sola delle componenti: o "discrete" oppure "assolutamente continua".

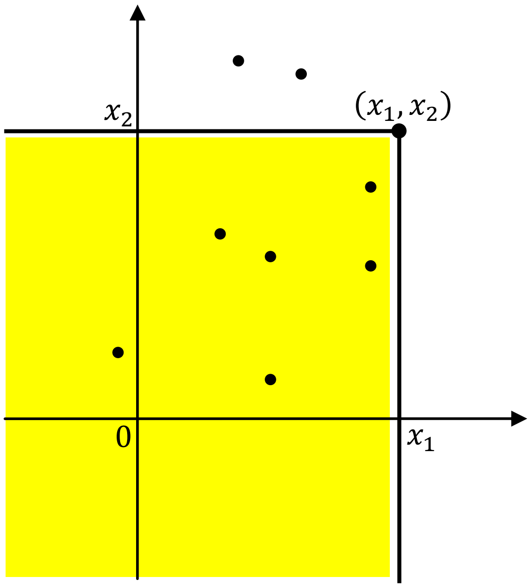
17. Variabili casuali bidimensionali discrete

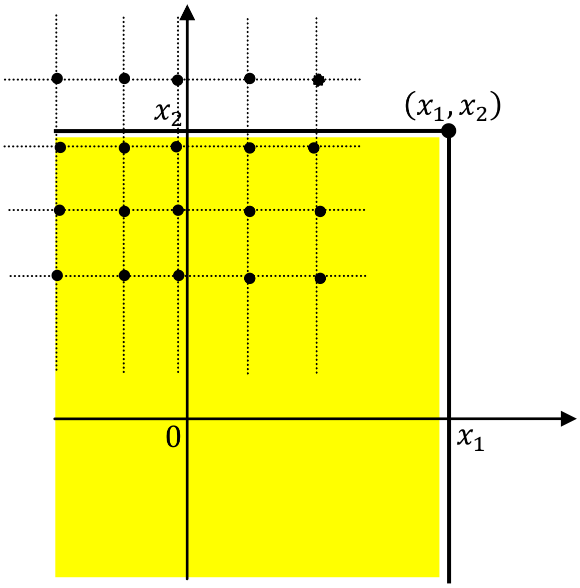
In una *v.c.* discreta nella *F.d.R.* $F(x_1, x_2)$ è presente il solo termine $F_d(x_1, x_2)$

$$F(x_1, x_2) = \sum_{k: x_{1k} \leq x_1, x_{2k} \leq x_2} P_{\boldsymbol{X}}(x_{1k}, x_{2k}) = \sum_{k: x_{1k} \leq x_1, x_{2k} \leq x_2} p_k$$

dove $(x_{1k}, x_{2k}) \in S$ essendo l'insieme di punti con probabilità diversa da zero $\{p_k = 1,2, \cdots\}$ costituito da una numerosità finita o numerabile. Qualora i punti siano disposti in forma ordinata come una matrice di dimensione $(l \times m)$ e quindi $S = \{(x_{1i}, x_{2j}): i = 1,2, \cdots, l; j = 1,2, \cdots, m\}$, posto $p_{ij} = P_{\boldsymbol{X}}(x_{1i}, x_{2j}) \geq 0$, abbiamo

$$F(x_1,x_2) = \sum_{x_{1i}\le x_1} \sum_{x_{2j}\le x_2} P_{\boldsymbol{X}}(x_{1i},x_{2j}) = \sum_{x_{1i}\le x_1} \sum_{x_{2j}\le x_2} p_{ij}$$





Nota la *F.d.R.* della *v.c.* bidimensionale $\boldsymbol{X} = (X_1, X_2)$ è possibile ottenere la *F.d.R.* delle sue due componenti X_1 o X_2 (*v.c.* “marginali”) che sono “discrete” se lo è la *v.c.* bidimensionale. Consideriamo la *v.c.* X_1 i cui punti in $\Re^2$, con probabilità diversa da zero, sono disposti in forma matriciale: $\{(x_{1i},x_{2j})\text{: per } i = 1,2,\cdots; j = 1,2,\cdots\}$

$$P\{X_1 = x_{1i}\} = \sum_j P_{\boldsymbol{X}}(x_{1i},x_{2j}) = \sum_j p_{ij} = p_{i\cdot}$$

da cui le funzioni *f.d.p.* e *F.d.R.* risultano

$$f_1(x_1) = \begin{cases} p_{i\cdot} & \text{se } x_1 = x_{1i} \\ 0 & \text{altrimenti} \end{cases}$$

$$F_1(x_1) = F(x_1,\infty) = \sum_{i:x_{1i}\le x_1} p_{i\cdot}$$

E’ possibile definire, oltre alle *v.c.* marginali unidimensionale X_1 e X_2, anche delle *v.c.* “condizionate” unidimensionali: $X_1|x_{2j}$ per $j =$

$1{,}2,\cdots$, applicando la formula delle probabilità condizionate: $P(A|B) = \frac{P(AB)}{P(B)}$ per $P(B) > 0$.

$$P\left(X_1 = x_{1i}\middle|x_{2j}\right) = \frac{p_{ij}}{p_{\cdot j}} \text{ per } i = 1, 2, \cdots; \text{fissato } j \text{ pari a } 1, 2, \cdots$$

da cui le funzioni *f.d.p.* e *F.d.R.* risultano

$$f_{1|x_{2j}}(x_1) = \begin{cases} p_{ij}/p_{\cdot j} & \text{se } x_1 = x_{1i} \\ 0 & \text{altrimenti} \end{cases}$$

$$F_{1|x_{2j}}(x_1) = \sum_{i:x_{1i} \le x_1} p_{ij}/p_{\cdot j} = \frac{1}{p_{\cdot j}} \sum_{i:x_{1i} \le x_1} p_{ij} \text{ fissato } j \text{ pari a } 1, 2, \cdots$$

Analogamente si ottengono le funzioni *f.d.p.* e *F.d.R.* delle *v.c.* marginale X_2 e condizionate $X_2|x_{1i}$ per $i = 1{,}2,\cdots$, le cui espressioni sono lasciate al lettore come pure la dimostrazione delle seguenti relazioni

$$\sum_i \sum_j p_{ij} = 1, \sum_i p_{i\cdot} = 1, \sum_j p_{\cdot j} = 1$$

dove le sommatorie sono estese a tutti i valori degli indici i e *j*.

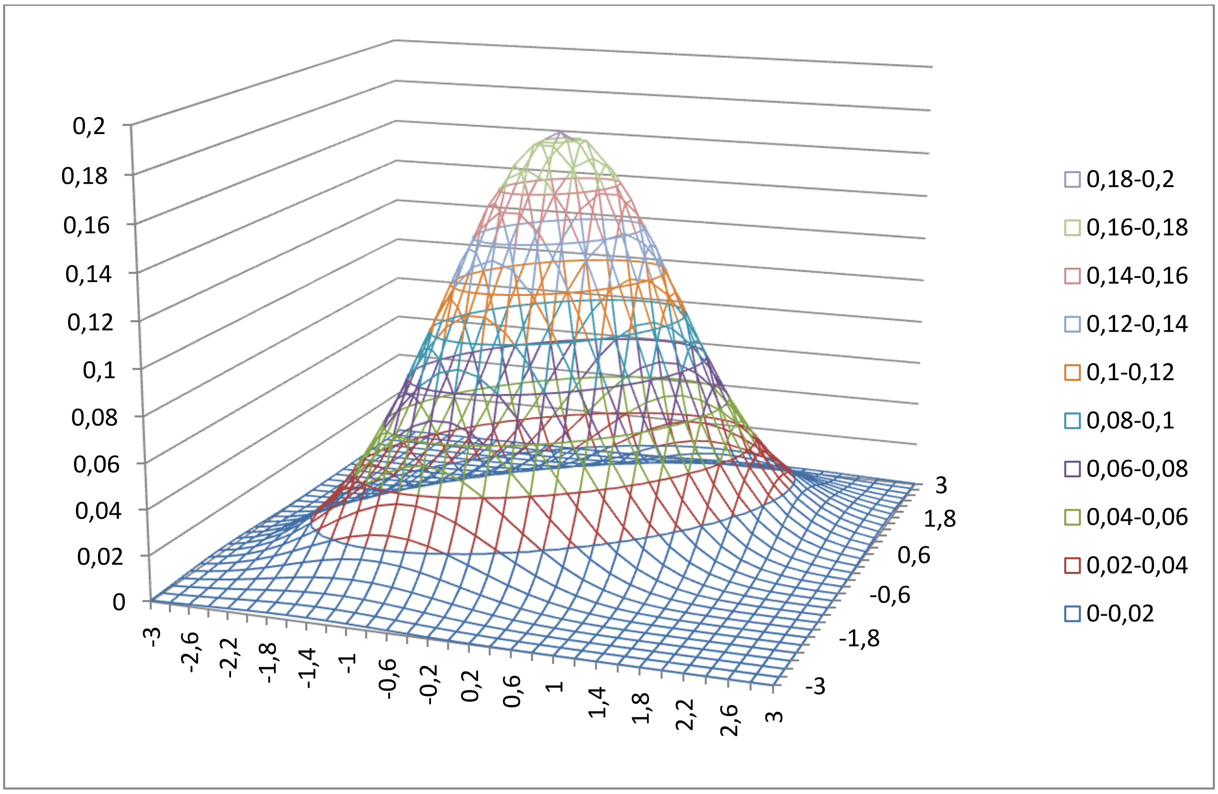
18. Variabili casuali bidimensionali continue

Qualora la *v.c.* $\boldsymbol{X} = (X_1, X_2)$ sia "assolutamente continua" la *F.d.R.* è data da

$$F(x_1, x_2) = \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} f(u_1, u_2) du_2 du_1$$

Essendo $f(x_1, x_2) \ge 0$ la *f.d.d.* "congiunta" che soddisfa la condizione

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x_1, x_2)dx_2 dx_1 = 1$$



Nota la *F.d.R.* della *v.c.* bidimensionale $\boldsymbol{X} = (X_1, X_2)$ è possibile ottenere la *F.d.R.* delle sue due componenti X_1 o X_2 (*v.c.* "marginali") che sono "continue", se lo è la *v.c.* bidimensionale. Consideriamo, per semplicità la sola *v.c.* X_1, allora la sua *F.d.R.* è data da

$$F_1(x_1) = F(x_1, \infty) = \int_{-\infty}^{x_1}\left[\int_{-\infty}^{\infty} f(u_1, u_2)du_2\right]du_1 = \int_{-\infty}^{x_1} f_1(u_1)du_1$$

dove $f_1(x_1)$ è la *f.d.d.* di X_1 che risulta ottenuta mediante l'integrale

$$f_1(x_1) = \int_{x_2=-\infty}^{\infty} f(x_1, x_2)dx_2$$

Analogamente, per X_2, si ottengono

$$F_2(x_2) = F(\infty, x_2) = \int_{-\infty}^{x_2}\left[\int_{-\infty}^{\infty} f(u_1, u_2)du_1\right]du_2 = \int_{-\infty}^{x_2} f_2(u_2)du_2$$

$$f_2(x_2) = \int_{x_1=-\infty}^{\infty} f(x_1, x_2)dx_1$$

Anche per la *v.c.* “continua” è possibile definire, oltre alla *v.c.* marginali unidimensionali X_1 e X_2, anche delle *v.c.* “condizionate” unidimensionali: $X_1|x_2$ per $\forall\ x_2$ tale che $f_2(x_2) > 0$, applicando la formula delle probabilità condizionate.

Le funzioni *f.d.p.* e *F.d.R.* di $X_1|x_2$, per ogni $x_2 \in \Re$, tale che $f_2(x_2) > 0$, risultano

$$f_{1|x_2}(x_1) = \frac{f(x_1, x_2)}{f_2(x_2)}$$

$$F_{1|x_2}(x_1) = \int_{-\infty}^{x_1} f_{1|x_2}(u_1)du_1 = \frac{1}{f_2(x_2)} \int_{-\infty}^{x_1} f(u_1, x_2)du_1$$

Analogamente si ottengono le funzioni *f.d.d.* e *F.d.R.* delle *v.c.* condizionate $X_2|x_i$, le cui espressioni sono lasciate al lettore come pure la dimostrazione delle seguenti relazioni

$$\int_{-\infty}^{\infty} f_{1|x_2}(x_1)dx_1 = \int_{-\infty}^{\infty} f_{2|x_2}(x_2)dx_2 = 1$$

E’ lasciata al lettore, l’estensione per analogia formale, dal caso bidimensionale (d = 2) al caso di *v.c.* multidimensionali (d > 2). Occorre fare attenzione alla simbologia che può essere appesantita dato il numero di *v.c.* in gioco, sia marginali sia condizionate.

19. Variabili casuali con componenti stocasticamente indipendenti

Una *v.c.* a d dimensioni $\boldsymbol{X} = (X_1, X_2, \cdots, X_d)$ presenta le componenti mutualmente “indipendenti” se per ogni regione $S = (s_1, s_2, \cdots s_d) \in \Re^d \subset \mathcal{B}_d$ si verifica la condizione di “fattorizzazione” propria degli eventi indipendenti. Per questo è sufficiente la “fattorizzazione” della *F.d.R.*

$$F(x_1, x_2, \cdots, x_d) = F_1(x_1) \cdot F_2(x_2) \cdot \cdots \cdot F_d(x_d)$$

La fattorizzazione si ripercuote, sia per *v.c.* discrete sia per *v.c.* continue, sulle di probabilità che di densità

$$f(x_1, x_2, \cdots, x_d) = f_1(x_1) \cdot f_2(x_2) \cdot \cdots \cdot f_d(x_d)$$

Consideriamo, come esempio, una *v.c.* bidimensionale $\boldsymbol{X} = (X_1, X_2)$ assolutamente continua, abbiamo infatti

$$F(x_1, x_2) = F_1(x_1) \cdot F_2(x_2) = \left[\int_{-\infty}^{x_1} f_1(u_1) du_1\right]\left[\int_{-\infty}^{x_2} f_2(u_2) du_2\right]$$
$$= \int_{-\infty}^{x_1}\int_{-\infty}^{x_2} [f_1(u_1) \cdot f_2(u_2)] du_2 du_1 \int_{-\infty}^{x_1}\int_{-\infty}^{x_2} f(u_1, u_2) du_2 du_1$$

Risultando così

$$f(x_1, x_2) = f_1(x_1) \cdot f_2(x_2)$$

In condizione di indipendenza le *v.c.* condizionate $X_1|x_2$ e $X_2|x_1$ risultano equivalenti alle componenti marginali corrispondenti X_1 e X_2, come è immediato verificare

$$f_{1|x_2}(x_1) = f_1(x_1), \;\; F_{1|x_2}(x_1) = F_1(x_1)$$

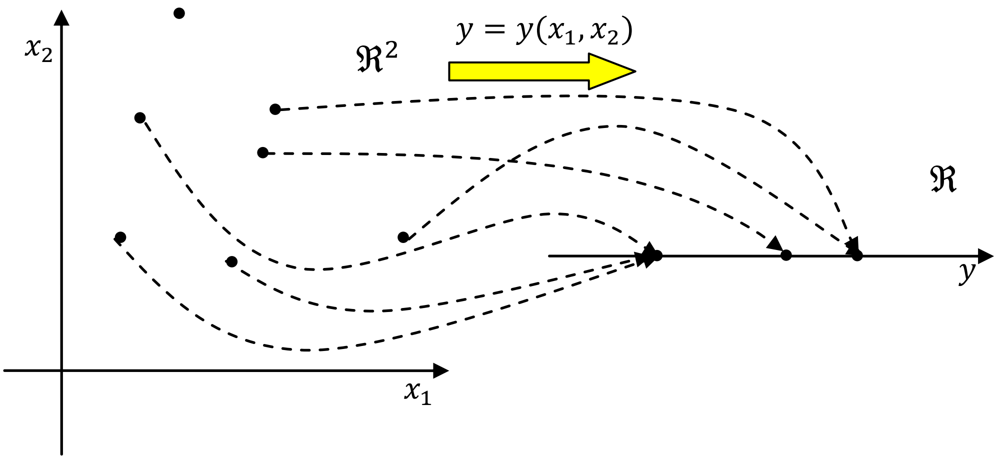
$$f_{2|x_2}(x_2) = f_2(x_2), \;\; F_{2|x_1}(x_2) = F_2(x_2)$$

Questa condizione di equivalenza delle *v.c.* in termini distributivi è indicata comunemente con: $X_1 \sim X_1|x_2$ e $X_2 \sim X_2|x_1$.

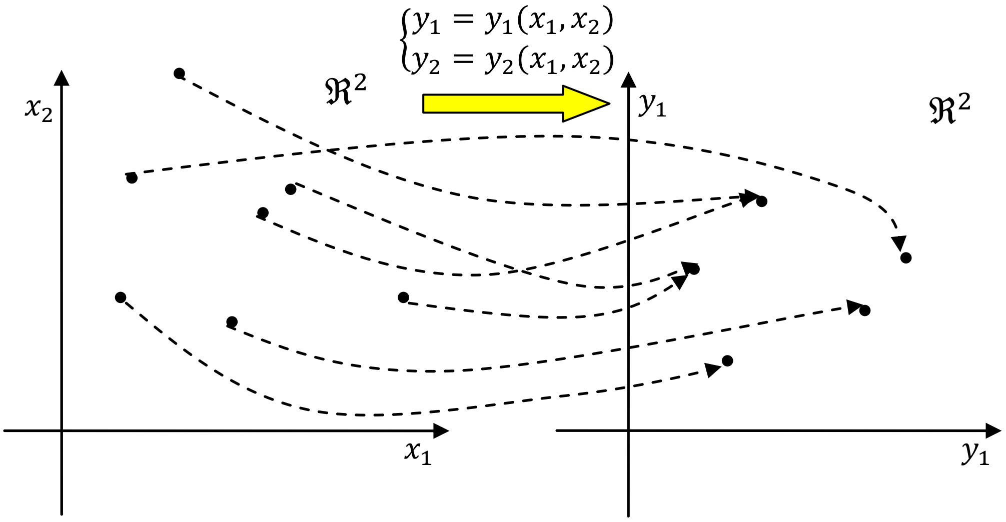
20. Variabili casuali funzioni di una variabile casuale bidimensionale

Consideriamo una *v.c.* $\boldsymbol{X} = (X_1, X_2)$ e il relativo spazio probabilistico $\{\Re^2, \mathcal{B}_2, \mathcal{P}_{\boldsymbol{X}}\}$, di cui sia nota la *F.d.R.* e la *f.d.p.* o *f.d.d.*, a seconda che si tratti di *v.c.* discreta o assolutamente continua.

Consideriamo una nuova *v.c.* *Y*, unidimensionale ottenuta mediante la trasformazione $Y = y(X_1, X_2)$, che fa corrispondere ogni punto di $\Re^2$ in uno in $\Re$. La *v.c.* *Y* è una *v.c.* funzione della *v.c.* $\boldsymbol{X} = (X_1, X_2)$.



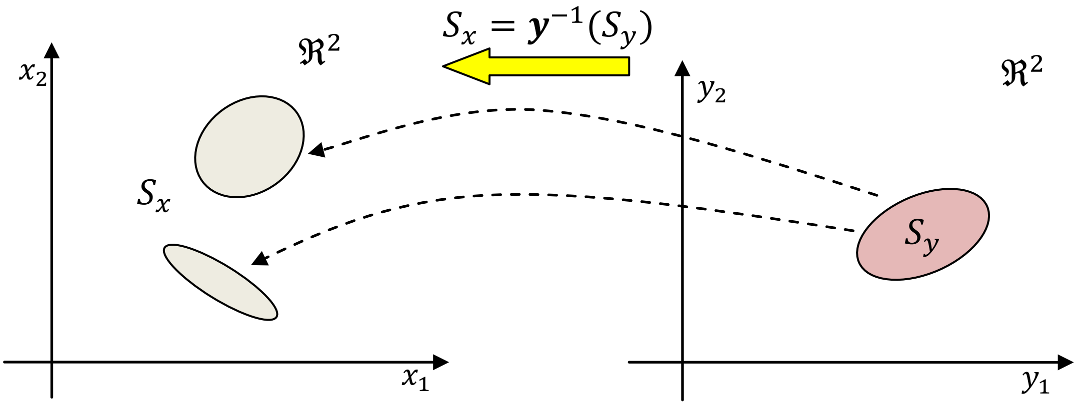
Similmente possiamo definire una nuova *v.c.* bidimensionale $\boldsymbol{Y} = (Y_1, Y_2)$ ottenuta mediante la trasformazione $\boldsymbol{Y} = \boldsymbol{y}(\boldsymbol{X}) = \boldsymbol{y}(X_1, X_2) = \big(Y_1 = y_1(X_1, X_2), Y_2 = y_2(X_1, X_2)\big)$, che fa corrispondere ogni punto di $\Re^2$ in uno in $\Re^2$. La *v.c.* $\boldsymbol{Y}$ è una *v.c.* funzione della *v.c.* $\boldsymbol{X} = (X_1, X_2)$.



La trasformazione deve essere tale che, per ogni elemento della classe di Borel $S_y \in \mathcal{B}_2$ corrisponda una immagine inversa $S_x = \boldsymbol{y}^{-1}(S_y)$ che si può probabilizzarsi nella regione della *v.c.* $\boldsymbol{X} = (X_1, X_2)$.

$$P_{\boldsymbol{Y}}\{S_y\} = P_{\boldsymbol{X}}\{S_x = \boldsymbol{y}^{-1}(S_y)\}$$

essendo $S_x = \{(x_1, x_2) \in \Re^2 : \boldsymbol{y}(x_1, x_2) \in S_y\}$



Lo studio delle *v.c.* funzioni di una *v.c.* bidimensionale comporta principalmente la determinazione della *F.d.R.* della *v.c.* $\boldsymbol{Y} = \boldsymbol{y}(\boldsymbol{X})$ partendo dalla *F.d.R.* della *v.c.* $\boldsymbol{X}$, come viene illustrato negli esempi seguenti.

<u>Esempio 13.</u>

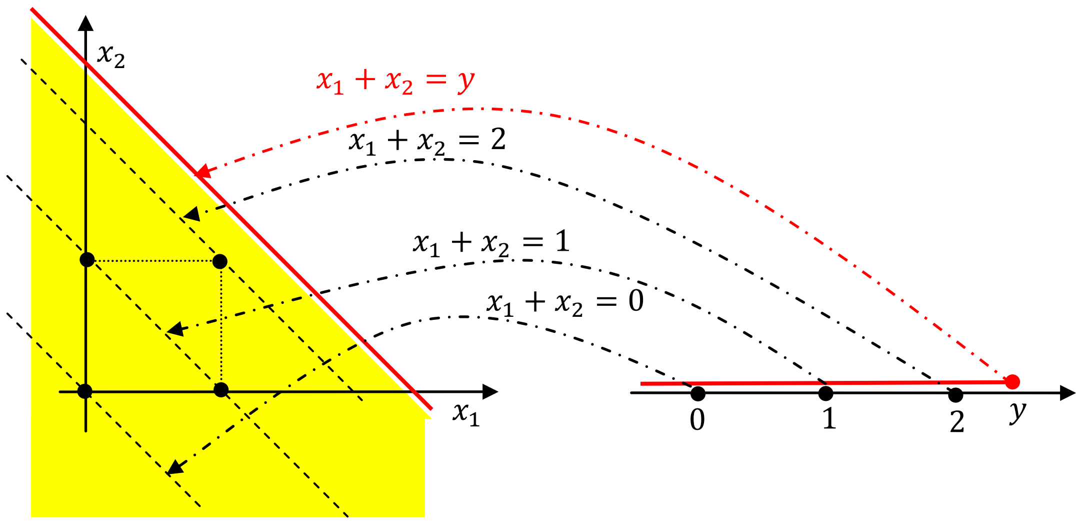
Si consideri la *v.c.* discreta $\boldsymbol{X} = (X_1, X_2)$ definita dalla *f.d.p.* riportata in tabella e si richiede di determinare la distribuzione della *v.c.* "somma" $Y = y(X_1, X_2) = X_1 + X_2$

$\{p_{ij}\}$		x_2		
		0	1	$p_{i\cdot}$
x_1	0	2/6	2/6	4/6
	1	1/6	1/6	2/6
	$p_{\cdot j}$	3/6	3/6	1

La Y è discreta e la sua *F.d.R.* è data da

$$F_Y(y) = P_Y(Y \leq y) = P_{\boldsymbol{X}}(S_x = X_1 + X_2 \leq y)$$

dove S_x è definito come immagine inversa di $S_y = Y \leq y$ ossia da tutti i punti di $\Re^2$ del semipiano limitato dalla linea di frontiera $(x_1 + x_2 = y)$.



La v.c. Y presenta valori con probabilità diverse da zero solo in corrispondenza di 0, 1 e 2. La *f.d.p.* e la *F.d.R.* di Y sono date nella tabella seguente.

y_k	p_k
0	2/6
1	3/6
2	1/6
	1

y	$F_y(y)$
$-\infty < y < 0$	0
$0 \leq y < 1$	2/6
$1 \leq y < 2$	5/6
$y \geq 2$	1

Generalizzando, nel caso di somma delle componenti della *v.c.* $\boldsymbol{X} = (X_1, X_2)$, possiamo definire la *F.d.R.* di $Y = X_1 + X_2$ come:

$$F_Y(y) = \sum_{\substack{i,j \\ x_{1i}+x_{2j}\leq y}} p_{ij}$$

oppure

$$F_Y(y) = \sum_{y_k \leq y} p_k = \sum_{y_k \leq y} \sum_{\substack{i,j \\ x_{1i}+x_{2j}=y_k}} p_{ij}$$

essendo $p_k = \sum_{\substack{i,j \\ x_{1i}+x_{2j}=y_k}} p_{ij}$ la *f.d.p.* della *v.c.* Y $p_k = P_y(Y = y_k)$.

21. Variabile casuale somma delle componenti di una variabile casuale bidimensionale continua

Dato l'interesse per le v.c. "assolutamente continue" ci soffermiamo in modo più dettagliato ad analizzare il caso della v.c. somma $Y = y(X_1, X_2) = X_1 + X_2$.

Dalla definizione della F.d.R. di Y, come abbiamo già visto, si ha

$$F_Y(y) = P_Y(Y \leq y) = P_{\boldsymbol{X}}(X_1 + X_2 \leq y)$$

Se scomponiamo l'insieme di punti $(x_1 + x_2 \leq y)$ in $\Re^2$in "strisce" (infinitesimali) definite dalle rette di equazione

$$\begin{cases} x_1 + x_2 = t \\ x_1 + x_2 = t + dt \end{cases}$$

La probabilità associata a ciascuna di tali strisce è

$$\int_t^{t+dt} \left[\int_{-\infty}^{\infty} f(x_1, t - x_1) dx_1 \right] dt \cong \left[\int_{-\infty}^{\infty} f(x_1, t - x_1) dx_1 \right] dt$$

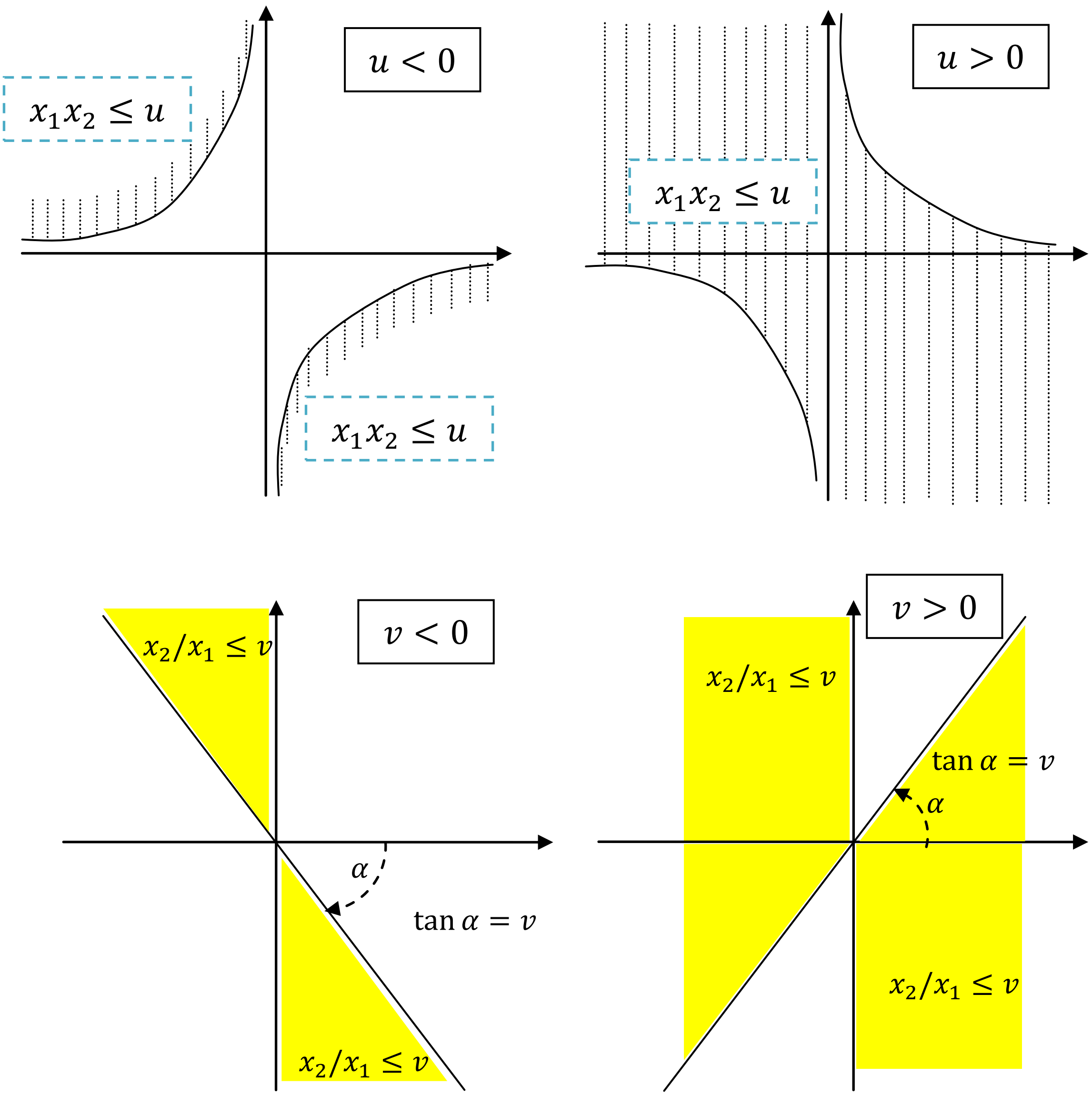
essendo la $f(x_1, x_2)$ vincolata da rispettare la condizione $x_1 + x_2 = t$ per cui $x_2 = t - x_1$.

Per il 3° assioma del calcolo delle probabilità si ha

$$F_Y(y) = \int_{t=-\infty}^{y} \left[\int_{-\infty}^{\infty} f(x_1, t - x_1) dx_1 \right] dt = \int_{t=-\infty}^{y} f_y(t)\, dt$$

Dove $f_y(y) = \int_{-\infty}^{\infty} f(x_1, y - x_1) dx_1$ è la *f.d.d.* della *v.c.* *Y*, che è anche essa di tipo "continuo", essendo $f_y(y) \geq 0$.

Analogamente a quanto fatto nel caso della "somma" delle componenti di $\boldsymbol{X} = (X_1, X_2)$, che comprende banalmente anche il caso della "differenza", di può procedere anche nel caso di "prodotto" $U = u(X_1, X_2) = X_1 \cdot X_2$ e di "quoziente" $V = v(X_1, X_2) = X_2 / X_1$, definendo le regioni in (x_1, x_2) che soddisfano la condizione $x_1 \cdot x_2 \leq u$ e $x_1 / x_2 \leq v$, rispettivamente. Si lascia al lettore l'individuazione delle *F.d.R.* nelle due situazioni descritte in precedenza (si osservino le figure seguenti).



22. La regolarità nelle trasformazioni di variabili casuali continue bidimensionali

Spesso si è interessati a trasformazioni di $\boldsymbol{X} = (X_1, X_2)$, *v.c.* "assolutamente continua", in $\boldsymbol{Y} = (Y_1, Y_2)$ in cui le funzioni $y_1(x_1, x_2)$ e $y_2(x_1, x_2)$ sono espresse mediante delle combinazioni lineari. In situazioni similari, che indicheremo con il termine "regolari", la *f.d.d.* congiunta $f_{\boldsymbol{Y}}(y_1, y_2)$ si può ottenere con il metodo di sostituzione di variabili proprio dell'integrazione multipla. Diremo che una trasformazione è "regolare" se:

- Sia dato uno spazio probabilistico $\{\Re^{d=2}, \mathcal{B}_2, \mathcal{P}_{\boldsymbol{X}}\}$ relativo alla v.c. $\boldsymbol{X} = (X_1, X_2)$ e uno spazio probabilizzato indotto dalla trasformazione $\{\boldsymbol{X} \to \boldsymbol{Y} = (Y_1, Y_2): \boldsymbol{y}(\boldsymbol{X})\}$, indicato con $\{\Re^2, \mathcal{B}_2, \mathcal{P}_{\boldsymbol{Y}}\}$;

- $\boldsymbol{y} = (y_1, y_2) = \boldsymbol{y}(x_1, x_2)$ è una funzione continua e derivabile;
- $(y_1, y_2) = \boldsymbol{y}(x_1, x_2)$ è una funzione "biettiva", tale quindi da determinare in modo univoco sia le coordinate (y_1, y_2) mediante le (x_1, x_2) e, in modo inverso le coordinate (x_1, x_2) mediante le (y_1, y_2) tramite $\boldsymbol{x}(y_1, y_2) \equiv \boldsymbol{y}^{-1}(y_1, y_2)$

$$\begin{cases} y_1 = y_1(x_1, x_2) \\ y_2 = y_2(x_1, x_2) \end{cases} \Leftrightarrow \begin{cases} x_1 = x_1(y_1, y_2) \\ x_2 = x_2(y_1, y_2) \end{cases}$$

La f.d.d. di $\boldsymbol{Y} = (Y_1, Y_2)$ può ottenersi considerando equivalenza, in termini di differenziale di probabilità tra i due spazi di $\boldsymbol{X}$ e di $\boldsymbol{Y}$

$$\begin{aligned} f_{\boldsymbol{Y}}(y_1, y_2) dy_1 dy_2 &\equiv f_{\boldsymbol{X}}(x_1, x_2) dx_1 dx_2 \\ &= f_{\boldsymbol{X}}(x_1(y_1, y_2), x_2(y_1, y_2)) \cdot \left| det\{\boldsymbol{J}_{\boldsymbol{y}}\} \right| dy_1 dy_2 \end{aligned}$$

dove $\boldsymbol{J}_{\boldsymbol{y}}$ è la matrice "jacobiana" data dalle derivate parziali della trasformazione inversa $\boldsymbol{x}(y_1, y_2)$ e $\left| det\{\boldsymbol{J}_{\boldsymbol{y}}\} \right|$ ne è il valore assoluto del determinante

$$\boldsymbol{J}_{\boldsymbol{y}} = \left[\frac{\partial x_i}{\partial y_j}\right]_{i,j=1,2} = \begin{bmatrix} \dfrac{\partial x_1}{\partial y_1} & \dfrac{\partial x_1}{\partial y_2} \\ \dfrac{\partial x_2}{\partial y_1} & \dfrac{\partial x_2}{\partial y_2} \end{bmatrix}$$

<u>Esempio 14.</u>

Consideriamo la v.c. continua $\boldsymbol{X} = (X_1, X_2)$ in cui le due componenti sono indipendenti e aventi la stessa distribuzione $X_1 \sim X_2 \sim X$, definita dalla seguente *f.d.d.*

$$f_x(x) = \frac{1}{k} \exp\left\{-\frac{x^2}{2}\right\} \; per - \infty < x < \infty$$

la costante k, detta costante di normalizzazione, risulta pari

$$k = \sqrt{2\pi} = \int_{-\infty}^{\infty} f_x(x)dx$$

La f.d.d. di $\boldsymbol{X} = (X_1, X_2)$ è quindi

$$f_{\boldsymbol{X}}(x_1, x_2) = f_1(x_1)f_2(x_2) = \frac{1}{2\pi}\exp\left\{-\frac{x_1^2 + x_2^2}{2}\right\}$$

Consideriamo ora la v.c. $\boldsymbol{Y} = (Y_1, Y_2)$ in cui le componenti sono date rispettivamente dalla "somma" e dalla "differenza" delle componenti di $\boldsymbol{X} = (X_1, X_2)$, mediante quindi la trasformazione regolare

$$\begin{cases} y_1 = x_1 + x_2 \\ y_2 = x_1 - x_2 \end{cases} \Leftrightarrow \begin{cases} x_1 = (y_1 + y_2)/2 \\ x_2 = (y_1 - y_2)/2 \end{cases}$$

I valori di y_1, y_2 sono estesi all'intero asse reale, come quelli di x_1, x_2. La matrice jacobiana risulta

$$\boldsymbol{J_y} = \begin{bmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} \end{bmatrix} = \begin{bmatrix} 1/2 & 1/2 \\ 1/2 & -1/2 \end{bmatrix}$$

Da cui $det\{\boldsymbol{J_y}\} = -1/2$ e quindi $\left|det\{\boldsymbol{J_y}\}\right| = \frac{1}{2} > 0$

La *f.d.d.* congiunta della v.c. $\boldsymbol{Y} = (Y_1, Y_2)$ risulta

$$\begin{aligned} f_{\boldsymbol{Y}}(y_1, y_2) &= f_{\boldsymbol{X}}(x_1(y_1, y_2), x_2(y_1, y_2)) \cdot \left|det\{\boldsymbol{J_y}\}\right| \\ &= \frac{1}{2\pi}\exp\left\{-\frac{[(y_1 + y_2)/2]^2 + [(y_1 - y_2)/2]^2}{2}\right\}\frac{1}{2} \end{aligned}$$

essendo

$$[(y_1+y_2)/2]^2+[(y_1-y_2)/2]^2=\frac{[(y_1+y_2)^2+(y_1-y_2)^2]}{4}$$
$$=\frac{y_1^2+y_2^2}{2}$$

otteniamo

$$f_{\boldsymbol{Y}}(y_1,y_2)=\left(\frac{1}{2\pi}\right)\left(\frac{1}{2}\right)\exp\left\{-\frac{(y_1^2+y_2^2)/2}{2}\right\}$$

Con alcuni passaggi algebrici la *f.d.d.* di (Y_1,Y_2) può porsi nella forma di fattorizzazione

$$f_{\boldsymbol{Y}}(y_1,y_2)=\left(\frac{1}{2\pi}\right)\left(\frac{1}{2}\right)\exp\left\{-\frac{(y_1^2+y_2^2)/2}{2}\right\}$$
$$=\left[\frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{2}}\exp\left\{-\frac{\left(y_1/\sqrt{2}\right)^2}{2}\right\}\right]\left[\frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{2}}\exp\left\{-\frac{\left(y_2/\sqrt{2}\right)^2}{2}\right\}\right]$$
$$=f_1(y_1)f_2(y_2)$$

Possiamo quindi osservare che $\boldsymbol{Y}=(Y_1,Y_2)$ presenta componenti indipendenti, la legge di distribuzione delle singole componenti è la stessa $Y_1\sim Y_2\sim Y$ con f.d.d.

$$f_y(y)=\frac{1}{k}\left(\frac{1}{\sqrt{2}}\right)\exp\left\{-\frac{\left(y/\sqrt{2}\right)^2}{2}\right\}$$

Avente quindi lo stesso andamento di $X\sim X_1\sim X_2$ e quindi delle componenti di $\boldsymbol{X}=(X_1,X_2)$ a cui è legata da una relazione lineare

$$Y=\sqrt{2}X$$

23. Valor medio e momenti di variabili casuali uni- e bi-dimensionali

Come per le variabili statistiche che sono state introdotte nella Parte I, riguardante argomenti di “statistica descrittiva”, si richiede di fornire alcune caratteristiche delle v.c. in forma sintetica così da effettuare rapidi confronti sul loro comportamento distributivo.

L'indicatore più importante è costituito ancora dalla "media aritmetica" detta anche "valore atteso" ("mean", "average", "expected value"; "moyenne", "valeur attendue", "espérance mathématique".

Sia X una v.c. e $Y = y(X)$ una v.c. funzione di Y, definiamo come media aritmetica di X e di Y, rispettivamente

$$\mu_x = M(X) = \begin{cases} \nearrow \sum_k x_k f(x_k) = \sum_k x_k p_k \quad per\ v.c.discreta \\ \searrow \int_{-\infty}^{\infty} x\, f(x)dx \qquad\qquad per\ v.c.continua \end{cases}$$

$$\mu_y = M(Y) = \begin{cases} \nearrow \sum_k y(x_k) f(x_k) = \sum_k y(x_k) p_k \quad per\ v.c.discreta \\ \searrow \int_{-\infty}^{\infty} y(x)\, f(x)dx \qquad\qquad per\ v.c.continua \end{cases}$$

Si osservi come anche la media di $Y = y(X)$ possa ottenersi impiegando direttamente la *f.d.p.* o la *f.d.d.* della v.c. di partenza X.

Rispetto a quanto avviene nel caso descrittivo per la definizione di media si richiede la condizione di "convergenza assoluta" per la serie o per l'integrale, cioè un valore "finito" per

$$M(|X|) = \begin{cases} \nearrow \sum_k |x_k| p_k < \infty \qquad \text{per } v.c.\text{discreta} \\ \searrow \int_{-\infty}^{\infty} |x|\, f(x)dx < \infty \ \text{ per } v.c.\text{continua} \end{cases}$$

$$M(|Y|) = \begin{cases} \nearrow \sum_k |y(x_k)| p_k < \infty \qquad \text{per } v.c.\text{discreta} \\ \searrow \int_{-\infty}^{\infty} |y(x)|\, f(x)dx < \infty \ \text{ per } v.c.\text{continua} \end{cases}$$

La "convergenza assoluta" è condizione sufficiente per la definizione di media.

Esempio 15

Riprendiamo l'esempio 13, di v.c. discrete $\{(X_1, X_2)$ e $Y = X_1 + X_2\}$

$\{p_{ij}\}$		x_2		
		0	1	$p_{i\cdot}$
x_1	0	2/6	2/6	4/6
	1	1/6	1/6	2/6
	$p_{\cdot j}$	3/6	3/6	1

y_k	p_k
0	2/6
1	3/6
2	1/6
	1

Per la media di X_1 e X_2, abbiamo

$$\mu_{x_1} = M(X_1) = 0 \cdot (4/6) + 1 \cdot (2/6) = 2/6 = 0{,}3333$$

$$\mu_{x_2} = M(X_2) = 0 \cdot (3/6) + 1 \cdot (3/6) = 3/6 = 0{,}5$$

$$\mu_y = M(Y) = 0 \cdot (2/6) + 1 \cdot (3/6) + 2 \cdot (1/6) = 5/6 = 0{,}8333$$

oppure

$$\mu_y = (0 + 0) \cdot (2/6) + (0 + 1) \cdot (2/6) + (1 + 0) \cdot (1/6) + (1 + 1) \cdot (1/6) = \\ = 5/6 = 0{,}8333$$

e, ricordando la proprietà di operatore lineare della media aritmetica, abbiamo

$$\mu_y = M(Y) = M(X_1 + X_2) = M(X_1) + M(X_2) = 0{,}3333 + 0{,}5 \\ = 0{,}8333$$

Esempio 16

Riprendiamo l'esempio 14. considerando la *v.c.* continua X avente *f.d.d.*

$$f_x(x) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{x^2}{2}\right\} \text{ per } -\infty < x < \infty$$

di cui si vuole ottenere la media $\mu_x = M(X)$ e la media di $Y = y(X) = X^2$.

Si verifichi la convergenza in valore assoluto per $M(X)$

$$\int_{-\infty}^{\infty} |x|\, f_x(x)dx = 2\int_0^{\infty} x \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{x^2}{2}\right\} dx$$
$$= \frac{2}{\sqrt{2\pi}} \int_0^{\infty} x \exp\left\{-\frac{x^2}{2}\right\} dx$$

data la simmetria rispetto all'origine di $f_x(x)$: $f_x(a) = f_x(-a)\ \forall\, a$. L'integrale al secondo membro può scriversi

$$I = \int_0^{\infty} x \exp\left\{-\frac{x^2}{2}\right\} dx$$

Posto $u = -x^2/2$, abbiamo $du = -x \cdot dx$, sostituendo otteniamo

$$I = -\int_0^{-\infty} \exp\{u\}\, du = \int_{-\infty}^{0} \exp\{u\}\, du = |\exp\{u\}|_{-\infty}^{0} = 1 < \infty$$

E quindi la convergenza assoluta di μ_x è assicurata.

La media $\mu_x = M(X)$ risulta

$$\mu_x = \int_{-\infty}^{\infty} x\, f_x(x)dx = \int_{-\infty}^{0} x\, f_x(x)dx + \int_0^{\infty} x\, f_x(x)dx$$
$$= -\int_0^{\infty} x\, f_x(x)dx + \int_0^{\infty} x\, f_x(x)dx = 0$$

data la simmetria rispetto all'origine di $f_x(x)$.

Per μ_y, media di $Y = y(X) = X^2$, si ha

$$\mu_y = \int_{-\infty}^{\infty} x^2 f_x(x)dx = 2\int_0^{\infty} x^2 f_x(x)dx$$
$$= \frac{2}{\sqrt{2\pi}}\int_0^{\infty} x^2 \exp\left\{-\frac{x^2}{2}\right\}dx$$

Posto $u = x^2/2$, si ha $du = x\,dx$ e $x = +\sqrt{2}u^{1/2}$, da cui sostiyiendo nell'integrale precedente abbiamo

$$\mu_y = \frac{2}{\sqrt{2\pi}}\int_0^{\infty} x^2 \exp\left\{-\frac{x^2}{2}\right\}dx = \frac{2}{\sqrt{2\pi}}\int_0^{\infty} \sqrt{2}u^{1/2}\exp\{-u\}du$$
$$= \frac{2}{\sqrt{\pi}}\int_0^{\infty} u^{\frac{3}{2}-1}\exp\{-u\}du = \frac{2}{\sqrt{\pi}}\Gamma\left(\frac{3}{2}\right)$$

Essendo $\Gamma(a) = \int_0^{\infty} u^{a-1}\exp\{-u\}du$ l'integrale euleriano di 2[a] specie (funzione Gamma), che per $a = 3/2$ è pari a $\Gamma\left(\frac{3}{2}\right) = \frac{1}{2}\Gamma\left(\frac{1}{2}\right) = \frac{\sqrt{\pi}}{2}$, da cui consegue che

$$\mu_y = M(X^2) = 1 < \infty$$

E quindi è rispettata anche per $M(X^2) = M(|X|^2)$ la convergenza assoluta.

Analogamente per una v.c. *X* si possono definire i "momenti dall'origine" μ_re i "momenti centrali" di ordine *r* sempre sotto la condizione di convergenza assoluta delle serie o degli integrali

$$\mu_r = M(X^r) = \begin{cases} \nearrow \sum_k x_k^r f(x_k) = \sum_k x_k^r p_k & \text{per } v.c.\text{ discreta} \\ \searrow \int_{-\infty}^{\infty} x^r f(x)dx & \text{per } v.c.\text{ continua} \end{cases}$$

$$\bar{\mu}_r = M((X-\mu)^r) = \begin{cases} \nearrow \sum_k (x_k-\mu)^r f(x_k) = \sum_k (x_k-\mu)^r p_k & \text{per } v.c. \text{ discreta} \\ \searrow \int_{-\infty}^{\infty} (x-\mu)^r\, f(x)dx & \text{per } v.c. \text{ continua} \end{cases}$$

Per $l = 0, 1, 2, \cdots$

Come si è visto in ambito descrittivo esiste una relazione tra i momenti dall'origine e quelli centrali, che vale anche nel caso di *v.c.*, in particolare ricordiamo

1. Indice di variabilità: varianza di *X*
$$\sigma^2 = Var(X) = \bar{\mu}_2 = M((X-\mu)^2) = \mu_2 - \mu^2$$

2. Indice di simmetria: ottenuto dal momento 3° centrale
$$\gamma_1 = \bar{\mu}_3/\sigma^3 \gtreqless 0$$

3. Indice di forma: curtosi di *X*
$$\gamma_2 = \bar{\mu}_4/\sigma^4 \geq 0$$

Analogamente a quanto fatto in statistica descrittiva si possono definire i momenti anche per v.c. bidimensionali (X_1, X_2), sempre con l'attenzione della esistenza in convergenza assoluta

Limitiamoci a considerare il momento misto di ordine $(r = 1, s = 1)$

$$\mu_{11} = M(X_1 \cdot X_2) \begin{cases} \nearrow \sum_k x_k y_k f(x_k, y_k) & \text{per } v.c. \text{ discreta} \\ \searrow \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x\, y\, f(x,y)dxdy & \text{per } v.c. \text{ continua} \end{cases}$$

$$\begin{cases} \bar{\mu}_{11} = \mu_{11} - \mu_{10}\mu_{01} \\ Cov(X_1, X_2) = M(X_1X_2) - M(X_1)M(X_2) \end{cases}$$

La covarianza esiste se le varianze di X_1 e X_2 sono finite, si ricordi infatti che $|Cov(X_1, X_2)| \leq [Var(X_1) \cdot Var(Y_1)]^{1/2}$.

Osservazioni

Riprendiamo alcuni teoremi sul valor medio di *v.c.* e sui momenti di *v.c.*.

1. Sia X una v.c. con media $M(X)$ e $Y = a + bX$ allora $M(Y) = a + bM(X)$;
2. Sia $\{X_l; l = 1,2, \cdots\}$ una successione di v.c. aventi media, allora posto $Y = a_0 + \sum_l a_l X_l$ una v.c. combinazione lineare delle $\{X_l\}$, allora $M(Y) = a_0 + \sum_l a_l M(X_l)$ (combinazione lineare delle medie). Per la verifica della "convergenza assoluta" si ricordi che $|a + b| \leq |a| + |b|$.

Per quanto riguarda l'operatore "varianza"

3. Se una *v.c.* presenta un valore finito della varianza anche la media esiste e quindi converge in valore assoluto, il contrario può essere non vero;
4. Sia *X* una *v.c.* con media $\mu = M(X)$ e varianza $\sigma^2 = Var(X)$, allora per $Y = a + bX$ abbiamo: $Var(Y) = b^2 Var(X)$;
5. Sia $\{X_l; l = 1,2, \cdots\}$ una successione di *v.c.* aventi media e varianza e "mutualmente indipendenti", posto $Y = a_0 + \sum_l a_l X_l$ una v.c. combinazione lineare delle $\{X_l\}$, allora abbiamo: $Var(Y) = \sum_l a_l^2 Var(X_l)$;
6. Consideriamo una *v.c.* bidimensionale (X, Y) e sia $Z = aX + bY$, la varianza di *Z* è data da:

   $$Var(Z) = a^2 Var(X) + 2abCov(X, Y) + b^2 Var(Y)$$

   in cui è coinvolta la covarianza tra *X* e *Y*;

7. Se si considera una v.c. X avente media e varianza, μ_x e σ_x^2 rispettivamente, consideriamo la v.c. $Z = (X - \mu_x)/\sigma_x$ la media e la varianza di Z risultano rispettivamente $\mu_z = 0$ e $\sigma_x^2 = 1$, come si dimostra considerando le precedenti proprietà degli operatori. Una v.c. come Z è detta "v.c. standardizzata" rispetto a

X. Un esempio è stato dato negli esempi 14. e 16., tale v.c. è detta "gaussiana o normale standardizzata" e indicata come $Z \sim N(0,1)$.

24. La funzione generatrice dei momenti

La Funzione Generatrice dei Momenti (*F.G.M.*) sintetizza e permette di ottenere tutti i momenti di una *v.c.* *X* e, se esiste nel senso di convergenza assoluta, è in corrispondenza biunivoca con la *F.d.R.* costituendo un modo alternativo di individuazione della distribuzione di *X*.

La *F.G.M.* è definita come media di una funzione di *X*

$$G_X(t) = M(\mathrm{e}^{tX})$$

media che deve esistere almeno valori del parametro *t* in un intorno dello zero: $|t| < t_0$. In tal caso esistono tutti i momenti di *X* (dall'origine, di conseguenza anche quelli centrali) che sono dati

$$\mu_r = \left(\frac{d^r G_x}{dt^r}\right)_{t=0} \qquad (*)$$

Si richiede quindi che $G_X(t)$ sia derivabile, nell'intorno $|t| < t_0$, fino all'*r* termine.

Per dimostrare tale caratteristica di $G_X(t)$, consideriamo il caso di una *v.c.* "continua", ricordando espressine dello sviluppo di $e^u = \sum_{k=0}^{\infty} \frac{u^k}{k!}$, ottenendo

$$G_X(t) = \int_{-\infty}^{\infty} \mathrm{e}^{xt} f(x)dx = \int_{-\infty}^{\infty} \mathrm{e}^{xt} f(x)dx$$
$$= \int_{-\infty}^{\infty} \sum_{k=0}^{\infty} \left(\frac{(xt)^k}{k!}\right) f(x)dx = \sum_{k=0}^{\infty} \frac{t^k}{k!} \int_{-\infty}^{\infty} x^k f(x)dx$$

Se i momenti μ_k esistono, abbiamo

$$G_X(t) = \sum_{k=0}^{\infty} \frac{\mu_k}{k!} t^k$$

e ricordando la sviluppo di Mac-Laurin, otteniamo la $(*)$.

L'importanza della *F.G.M.* $G_X(t)$ non sta solo nella possibilità di ottenere i momenti di una *v.c.* *X* ma tanto del "teorema di unicità", che possiamo sintetizzare: se due *v.c.* hanno uguale *F.G.M.* hanno anche uguale distribuzione *F.d.R.* e viceversa. Questo permette di definire più facilmente l'appartenenza di una *v.c.* ad una particolare classe di distribuzioni.

Nelle seguenti osservazioni diamo alcuni teoremi, basati sulle proprietà dell'operatore media aritmetica $M(\cdot)$, riguardanti la *F.G.M.*

Osservazioni

1. Se $Y = a + bX$, con $b \neq 0$, e X è una *v.c.* con *F.G.M.* data da $G_X(t)$, allora la *F.G.M.* di *Y* risulta

$$G_Y(t) = M(\mathrm{e}^{tY}) = M\left(\mathrm{e}^{t(a+bX)}\right) = M(\mathrm{e}^{at} \cdot \mathrm{e}^{btX}) = \mathrm{e}^{at} M\left(\mathrm{e}^{(bt)X}\right)$$
$$= \mathrm{e}^{at} G_X(bt)$$

2. Se X è una *v.c.* con media μ e varianza σ^2 avente *F.G.M.* pari a $G_X(t)$, la *v.c.* $Z = (X - \mu)/\sigma$, detta *v.c.* standardizzata, avente $M(Z) = 0$ e $Var(Z) = 1$, presenta la seguente *F.G.M.*

$$G_Z(t) = M(\mathrm{e}^{tZ}) = M\left(\mathrm{e}^{t((X-\mu)/\sigma)}\right) = M\left(\mathrm{e}^{\left(-\frac{\mu}{\sigma}+\frac{X}{\sigma}\right)t}\right)$$
$$= M\left(\mathrm{e}^{\mu t/\sigma} \mathrm{e}^{X(\frac{t}{\sigma})}\right) = \mathrm{e}^{\mu t/\sigma} G_X\left(\frac{t}{\sigma}\right)$$

3. Se X è una *v.c.* con media μ e varianza σ^2 e si conosca la *F.G.M.* della corrispondente *v.c.* $Z = (X - \mu)/\sigma$ standardizzata pari a $G_Z(t)$, allora la *F.G.M.* di X si ottiene, ricordando che $X = \mu + \sigma Z$ e applicando l'osservazione 1.

$$G_X(t) = \mathrm{e}^{\mu t} G_Z(\sigma t)$$

4. Consideriamo due *v.c.* X e Y indipendenti, per semplicità entrambe continue, con relative *F.G.M.* $G_X(t)$ e $G_Y(t)$. La *v.c.* somma $W = X + Y$ presenta la *F.G.M.* $G_W(t)$, che si ottiene

$$G_W(t) = M(\mathrm{e}^{tW}) = M\left(\mathrm{e}^{t(X+Y)}\right) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mathrm{e}^{t(x+y)} f(x,y)dxdy$$

Data la condizione di indipendenza tra X e Y la *f.d.d.* congiunta $f(x,y)$ è uguale al prodotto delle *f.d.d.*: $f(x,y) = f_x(x) \cdot f_y(y)$, quindi sostituendo nella relazione precedente abbiamo

$$\begin{aligned} G_W(t) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mathrm{e}^{xt}\mathrm{e}^{yt} f_x(x) f_y(y)dxdy \\ &= \left[\int_{-\infty}^{\infty} \mathrm{e}^{xt} f_x(x)dx\right] \cdot \left[\int_{-\infty}^{\infty} \mathrm{e}^{yt} f_y(y)dy\right] \\ &= G_X(t) \cdot G_Y(t) \end{aligned}$$

Relazione che in caso di indipendenza delle componenti dà la *F.G.M.* della "somma" come "prodotto" delle singole *F.G.M.*. Il risultato ottenuto può estendersi alla somma di n componenti $W = \sum_{i=1}^{n} X_i$ con $G_{X_i}(t)$ per $i = 1, 2, \cdots, n$

$$G_W(t) = \prod_{i=1}^{n} G_{X_i}(t)$$

Qualora le componenti $X_i \sim X: i = 1, 2, \cdots, n$ siano oltre che indipendenti anche "identicamente distribuite" (*i.i.d.*) la *F.G.M.* della somma risulta

$$G_W(t) = [G_X(t)]^n$$

essendo $G_X(t)$ la *F.G.M.* di ogni componente.

Esempio 17.

Consideriamo la *v.c.* discreta “indicatore” X che assume il valore 1 se si verifica l’evento favorevole A, con probabilità $0 < p < 1$, e il valore 0 se si ha A^c cioè l’evento contrario, con probabilità $q = 1 - p$. La *F.G.M.* di X risulta

$$G_X(t) = M(\mathrm{e}^{tX}) = \mathrm{e}^{0t}q + \mathrm{e}^{1t}p = q + \mathrm{e}^t p$$

La media di X si ottiene

$$\mu = M(X) = \left[\frac{d}{dt}G_X(t)\right]_{t=0} = \left[\frac{d}{dt}(q + \mathrm{e}^t p)\right]_{t=0} = p\left[\frac{d}{dt}\mathrm{e}^t\right]_{t=0} = p$$

E il generico momento dall’origine

$$\mu_r = \left[\frac{d^r}{dt^r}G_X(t)\right]_{t=0} = p \text{ per } r = 1, 2, \cdots$$

La varianza di X si ottiene immediatamente come

$$Var(X) = \mu_2 - \mu^2 = p - p^2 = p(1-p) = pq$$

Esempio 18.

Consideriamo ora n *v.c.* indipendenti X_i distribuite come la *v.c.* X dell’esempio precedente $X_i \sim X: i = 1, 2, \cdots, n$ con *F.G.M.* pari a

$$G_X(t) = M(\mathrm{e}^{tX}) = q + \mathrm{e}^t p$$

La *v.c.* “somma” $W = \sum_{i=1}^n X_i$ presenta la seguente *F.G.M.*

$$G_W(t) = [G_X(t)]^n = (q + \mathrm{e}^t p)^n$$

La media di W si ottiene dalla relativa *F.G.M.* come

$$\mu = M(W) = \left[\frac{d}{dt}G_W(t)\right]_{t=0} = \left[\frac{d}{dt}(q + \mathrm{e}^t p)^n\right]_{t=0}$$
$$= [n(q + \mathrm{e}^t p)^{n-1}\mathrm{e}^t p]_{t=0} = np$$

Il momento secondo $\mu_2 = M(W^2)$ si ottiene dalla $G_W(t)$ come

$$\mu_2 = M(W^2) = \left[\frac{d^2}{dt^2} G_W(t)\right]_{t=0} = \left[\frac{d}{dt} n(q + \mathrm{e}^t p)^{n-1} \mathrm{e}^t p\right]_{t=0}$$
$$= \left[np \frac{d}{dt} (q + \mathrm{e}^t p)^{n-1} \mathrm{e}^t\right]_{t=0}$$
$$= \{np[(n-1)(q + \mathrm{e}^t p)^{n-2} \mathrm{e}^t p + (q + \mathrm{e}^t p)^{n-1} \mathrm{e}^t]\}_{t=0}$$
$$= n[(n-1)p + 1]p = n^2p^2 - np^2 + np$$

Si può ottenere la varianza di W come

$$Var(W) = \mu_2 - \mu^2 = (n^2p^2 - np^2 + np) - (np)^2 = np - np^2$$
$$= np(1-p) = npq$$

Si può notare come tale calcolo sia piuttosto laborioso, infatti la media e la varianza di *W* si ottengono più immediatamente come

$$\mu = M(W) = M\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} M(X_i) = nM(X) = np$$

$$Var(W) = Var\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} Var(X_i) = nVar(X) = npq$$

Esempio 19.

Consideriamo la *v.c.* continua *Z* corrispondente alla legge gaussiana standardizzata $Z \sim N(0,1)$, si veda gli esempi 14. e 16., se ne voglia costruire la *F.G.M.*, ricordando che la *f.d.d.* di *Z* è

$$f_z(z) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{z^2}{2}\right\} \text{ per } -\infty < z < \infty$$

Abbiamo così

$$G_Z(t) = M(\mathrm{e}^{Zt}) = \int_{-\infty}^{\infty} \mathrm{e}^{zt}\, f_z(z)dz$$
$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp(zt) \exp\left\{-\frac{z^2}{2}\right\} dz$$
$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left\{zt - \frac{z^2}{2}\right\} dz \quad (**)$$

L'espressione entro parentesi graffa può scriversi

$$zt - \frac{z^2}{2} = -\frac{1}{2}(z^2 - 2zt) = -\frac{1}{2}(z^2 - 2zt \pm t^2) = -\frac{(z-t)^2}{2} + \frac{t^2}{2}$$

Introducendo tale espressione effettuando inoltre il cambiamento di variabili all'interno dell'integrale $u = z - t$, abbiamo

$$G_z(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left\{zt - \frac{z^2}{2}\right\} dz$$
$$= \frac{1}{\sqrt{2\pi}} \exp\left(\frac{t^2}{2}\right) \int_{-\infty}^{\infty} \exp\left\{-\frac{u^2}{2}\right\} du$$

Ricordando che $\int_{-\infty}^{\infty} \exp\left\{-\frac{u^2}{2}\right\} du = \sqrt{2\pi}$, otteniamo per $G_W(t)$ l'espressione

$$G_Z(t) = \exp(t^2/2)$$

Le derivate di $G_Z(t)$, fino al quarto ordine risultano:

$$G'_Z(t) = t \cdot \exp(t^2/2)$$

$$G''_Z(t) = \frac{d}{dt}[t \cdot \exp(t^2/2)] = \exp(t^2/2) + t\exp(t^2/2)$$
$$= ([1 + t^2])\exp(t^2/2)$$

$$G'''_Z(t) = \frac{d}{dt}[(1+t^2)\cdot \exp(t^2/2)]$$
$$= 2t\exp(t^2/2) + (1+t^2)t\exp(t^2/2)$$
$$= [(3t+t^3)]\exp(t^2/2)$$

$$G_Z^{(4)}(t) = \frac{d}{dt}\{[(3t+t^3)]\exp(t^2/2)\}$$
$$= (3+3t^2)\exp(t^2/2) + (3t+t^3)t\exp(t^2/2)$$
$$= (3++6t^3+t^4)\exp(t^2/2)$$

I momenti di *Z* risultano: $M(Z) = G'_Z(0) = 0$; $M(Z^2) = G''_Z(0) = 1$; $M(Z^3) = G'''_Z(0) = 0$ e $M(Z^4) = G_Z^{(4)}(0) = 3$. I risultati ottenuti sono coerenti con il fatto che la *v.c.* Z è “standardizzata” si presenta “simmetrica” rispetto all’origine.

Esempio 20.

Consideriamo la *v.c. X*, ottenuta dalla *v.c. Z* gaussiana standardizzata vista nell’esercizio precedente, si desideri ottenere la *F.G.M.* di $X = \mu + \sigma Z$ dove μ e σ sono rispettivamente la media e lo scarto quadratico medio di *X*. Dalla osservazione 3. Abbiamo

$$G_X(t) = \mathrm{e}^{\mu t} G_Z(\sigma t)$$

da cui, ricordando che la *F.G.M.* di $Z \sim N(0{,}1)$ è data da: $G_Z(t) = \exp(t^2/2)$ e quindi di ottiene

$$G_X(t) = \mathrm{e}^{\mu t}\exp(\sigma^2\, t^2/2) = \exp\left(\mu t + \frac{\sigma^2 t^2}{2}\right)$$

è possibile ottenere i principali momenti, sia mediante successive derivazioni di $G_X(t)$ oppure mediante lo sviluppo della funzione esponenziale. In questo ultimo caso abbiamo

$$G_X(t) = \exp\left(\mu t + \frac{\sigma^2 t^2}{2}\right) = \sum_{k=0}^{\infty} \frac{1}{k!}\left(\mu t + \frac{\sigma^2 t^2}{2}\right)^k$$

Per $k \leq 2$ abbiamo

$$G_X(t) = 1 + \frac{1}{1!}\left(\mu t + \frac{\sigma^2 t^2}{2}\right) + \frac{1}{2!}\left(\mu t + \frac{\sigma^2 t^2}{2}\right)^2 + \cdots$$
$$= 1 + \mu t + \frac{\sigma^2 + \mu^2}{2!} t^2 + \cdots$$
$$= 1 + G_x(0) t + \frac{G_x'(0)}{2!} t^2 + \cdots$$

Da cui abbiamo conferma che $M(X) = \mu$ e $M(X^2) = \sigma^2 + \mu^2$.

25. Diseguaglianza di Chebyshev

L'importanza dei momenti è evidenziata dalla equivalenza, se sussistono le condizioni di convergenza assoluta tra la *F.G.M.* e la *F.d.R.* di una *v.c. X*. Già la conoscenza dei momenti fino al secondo ordine, $\mu = M(X)$ e $\sigma^2 = Var(X)$ permettono di precisare, in forma di diseguaglianza, l'andamento distributivo di *X*.

La diseguaglianza di Chebyshev precisa che la probabilità di *X* con valori "esterni" all'intervallo centrato sulla media $\mu \pm h\sigma$, con $h > 1$, è inferiore a $1/h^2$.

$$[P_X(|X - \mu|) \geq h\sigma] \leq \frac{1}{h^2}$$

Consideriamo, ad esempio, una *v.c. X* di tipo continuo, abbiamo che la varianza σ^2 è data da

$$\sigma^2 = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)dx$$
$$= \int_{-\infty}^{\mu-h\sigma} (x-\mu)^2 f(x)dx + \int_{\mu-h\sigma}^{\mu+h\sigma} (x-\mu)^2 f(x)dx$$
$$+ \int_{\mu+h\sigma}^{\infty} (x-\mu)^2 f(x)dx =$$

trascurando il termine centrale e ponendo al posto di x rispettivamente gli estremi di integrazione $\mu - h\sigma$ e $\mu + h\sigma$ negli altri due integrali, si ha

$$\sigma^2 = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)dx$$
$$\geq \int_{-\infty}^{\mu-h\sigma} (x-\mu)^2 f(x)dx + \int_{\mu+h\sigma}^{\infty} (x-\mu)^2 f(x)dx$$
$$\geq \int_{-\infty}^{\mu-h\sigma} (-h\sigma)^2 f(x)dx + \int_{\mu+h\sigma}^{\infty} (h\sigma)^2 f(x)dx$$
$$= h^2\sigma^2 \left[\int_{-\infty}^{\mu-h\sigma} f(x)dx + \int_{\mu+h\sigma}^{\infty} f(x)dx\right]$$

essendo l'espressione entro parentesi quadrata pari alla probabilità di valori di X esterni all'intervallo $\mu \pm h\sigma$, ottenendo così

$$\sigma^2 \geq h^2\sigma^2[P_X(|X-\mu|) \geq h\sigma] \to [P_X(|X-\mu|) \geq h\sigma] \leq \frac{1}{h^2}$$

Da tale diseguaglianza si ottiene che la probabilità di valori "compresi" nell'intervallo centrato sulla media $\mu \pm h\sigma$ è maggiore di $1 - 1/h^2$.

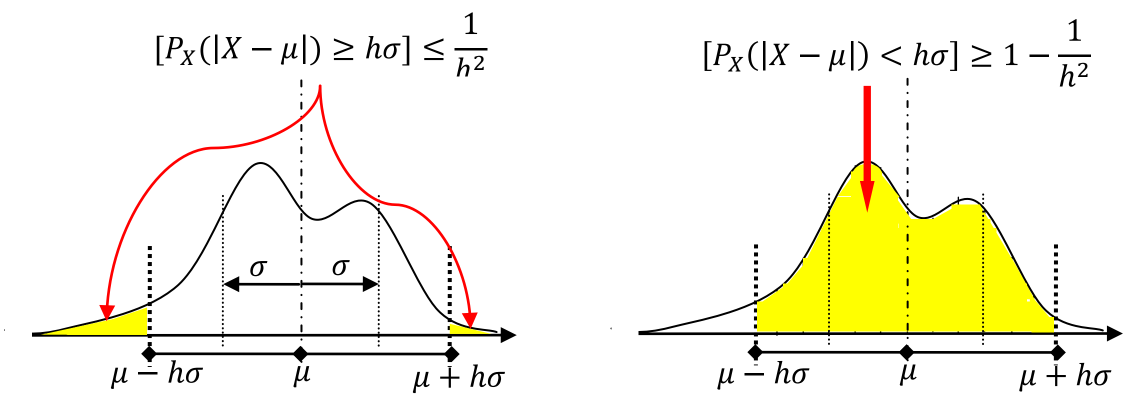
$$[P_X(|X-\mu|) < h\sigma] \geq 1 - \frac{1}{h^2}$$

Ad esempio per $h = 3$, nell'intervallo $\mu \pm 3\sigma$ si esaurisce circa o più del 90% della *v.c.* X. Infatti $[P_X(|X-\mu|) \geq 3\sigma] \leq \frac{1}{3^2} \cong 0{,}11$ da cui $[P_X(|X-\mu|) < 3\sigma] \geq 1 - \frac{1}{9} \cong 0{,}89$.

Qualora si consideri la *v.c.* standardizzata $Z = (X-\mu)/\sigma$, la diseguaglianza di Chebyshev può scriversi

$$[P_Z(|Z|) \geq h] \leq \frac{1}{h^2}$$

Scegliendo un valore opportuno di $h > 1$ ci si può assicurare che la *v.c.* oggetto di interesse sia concentrata in un intorno della sua media con una probabilità maggiore di un preassegnato valore $(1 - 1/h^2)$, è questo il ruolo importante della diseguaglianza di Chebyshev. In letteratura sono disponibili altre diseguaglianze che, sotto condizioni specifiche di conoscenza dei momenti, presentano soglie più ristrette di quella data dalla diseguaglianza di Chebyshev.



26. Modelli e variabili casuali di notevole interesse

Molte delle *v.c.* aventi notevole interesse sia per le applicazioni sia per quanto riguarda i modelli teorici sona già state presentate negli esempi precedenti. Si dà qui un riassunto proprietà delle principali *v.c.* discrete e continue.

1. <u>La *v.c.* “bernulliana”</u> $X \sim Be(p)$ con $0 < p < 1$

   È anche detta “indicatore” essendo associata al verificarsi o non verificarsi di un generico evento casuale A, avente nello spazio probabilistico considerato, la probabilità $p = P(A)$ e $q = 1 - p = P(A^c)$. Si tratta di una *v.c.* discreta che presenta la *f.d.p.* seguente

$$f(x) = \begin{cases} q & \text{per} \quad x = 0 \\ p & \text{per} \quad x = 1 \\ 0 & \text{per} \quad \Re - \{0,1\} \end{cases}$$

   La media di $X \sim Be(p)$ è $\mu_X = p$, la varianza è $\sigma_X^2 = pq$ e la *F.G.M.* è $G_X(t) = q + p\mathrm{e}^t$ da cui si ottengono tutti i momenti dall'origine $M(X^r) = p$.

2. <u>La *v.c.* “binomiale”</u> $Y \sim Bi(n, p)$ con $n = 0,1,2, \cdots$ e $0 < p < 1$

   La *v.c.* binomiale rappresenta il modello delle “prove ripetute” corrispondente all'esecuzione di n replicazioni “indipendenti” in ciascuna delle quali può o non può manifestarsi l'evento A a cui è associata la probabilità $p = P(A)$, (v. la *v.c.* bernoulliana). I valori assunti da Y sono dati dal numero di volte che si verifica l'evento A. Trattandosi di *v.c.* discreta la *f.d.p.* è data da

$$f_Y(y) = \begin{cases} \binom{n}{y} p^y q^{n-y} & \text{per } y = 0,1,2, \cdots, n \\ 0 & \text{altrove} \end{cases}$$

   La *v.c.* Y è data dalla somma di n *v.c.* bernoulliane *i.d.d.* $X_i \sim X \sim Be(p)$

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

   Consegue che la media e la varianza di Y sono: $\mu_Y = n\mu_X = np$, $\sigma_Y^2 = n\sigma_X^2 = npq$. La *F.G.M.* è $G_Y(t) = [G_Y(t)]^n = (q + p\mathrm{e}^t)^n$, dalla quale mediante derivazione o sviluppo in potenze di t si possono ottenere tutti i momenti.

Se $p = q = 1/2$ (esempio lancio di una moneta equilibrata) la *f.d.p* risulta $f_Y(y) = \binom{n}{x}\left(\frac{1}{2}\right)^n$ per $x = 0,1,2,\cdots,n$ e, inoltre, $\mu_Y = n/2$, $\sigma_Y^2 = n/4$.

3. La *v.c.* di Poisson $W \sim Po(\lambda)$ con $\lambda > 0$

   Si consideri una successione di *v.c.* binomiali $\{X_n \sim Bi(n,p); n = 1,2,\cdots\}$ aventi tutte la stessa media $\{M(X_n) = np = \lambda;\ n = 1,2,\cdots\}$ e quindi la successione di *v.c.* binomiali è data da $\{X_n \sim Bi(n, \lambda/n); n = 1,2,\cdots\}$. Al divergere di *n* la *v.c.* X_n tende alla seguente distribuzione discreta

$$f_W(w) = \lim_{n\to\infty} \binom{n}{w}\left(\frac{\lambda}{n}\right)^w \left(1 - \frac{\lambda}{n}\right)^{n-w}$$

$$= \lim_{n\to\infty} \frac{n!}{w!\,(n-w)!}\frac{\lambda^w}{n^w}\frac{\left(1-\frac{\lambda}{n}\right)^n}{\left(1-\frac{\lambda}{n}\right)^w} =$$

   Per un valore di *w* finito si dimostra facilmente che

$$\lim_{n\to\infty} \frac{n!}{(n-w)!\,n^w} = 1$$

$$\lim_{n\to\infty} \left(1-\frac{\lambda}{n}\right)^n = e^{-\lambda}$$

$$\lim_{n\to\infty} \left(1-\frac{\lambda}{n}\right)^w = 1$$

   quindi

$$f_W(w) = e^{-\lambda}\frac{\lambda^w}{w!} \quad \text{per} \quad w = 0, 1, 2, \cdots$$

   Si tratta di una *f.d.p.* in quanto $f_W(w) > 0$ e

$$\sum_{w=0}^{\infty} f_W(w) = \sum_{w=0}^{\infty} e^{-\lambda}\frac{\lambda^w}{w!} = e^{-\lambda}\sum_{w=0}^{\infty}\frac{\lambda^w}{w!} = e^{-\lambda}e^{\lambda} = 1$$

Il valor medio e la varianza di $W \sim Po(\lambda)$ possono ottenersi come limite della media e della varianza di $X_n \sim Bi(n, \lambda/n)$ al divergere di n

$$\mu_W = M(W) = \lim_{n\to\infty} M(X_n) = \lambda$$

$$\sigma_W^2 = Var(W) = \lim_{n\to\infty} Var(X_n) = \lim_{n\to\infty} n\left(\frac{\lambda}{n}\right)\left(1-\frac{\lambda}{n}\right) = \lambda$$

Similmente è possibile ottenere la *F.G.M.* di W

$$\begin{aligned} G_Y(t) &= \lim_{n\to\infty}\left(1-\frac{\lambda}{n}+\frac{\lambda}{n}\mathrm{e}^t\right)^n \\ &= \lim_{n\to\infty}\left(1+\frac{\lambda(\mathrm{e}^t-1)}{n}\right)^n = \exp\left(\lambda(\mathrm{e}^t-1)\right) \end{aligned}$$

espressione che esiste per ogni valore di *t* e quindi esistono tutti i momenti della *v.c.* W che possono ottenersi mediante lo sviluppo oppure mediante il calcolo delle derivate in $t = 0$.

4. <u>La *v.c.* "uniforme o rettangolare continua"</u> $U \sim Re(0,1)$

Sia *U* una *v.c.* continua con *f.d.d.* costante diversa da zero nell'intervallo $(0,1)$

$$f_U(u) = \begin{cases} k > 0 \text{ per } u \in (0,1) \\ 0 \qquad \text{altrimenti} \end{cases}$$

Il valore *k* è individuato dalla condizione

$$\int_{-\infty}^{\infty} f_U(u)du = \int_{-\infty}^{0} 0\, du = \int_0^1 k\, du = \int_1^{\infty} 0\, du = k\int_0^1 du = k = 1$$

La *F.d.R.* di *U* risulta

$$F_U(u) = \begin{cases} 0 & \text{per } u < 0 \\ u & \text{per } 0 \le u < 1 \\ 1 & \text{per } u \ge 1 \end{cases}$$

La media e la varianza di U sono date

$$\mu_U = M(U) = \int_0^1 u\, du = \left|\frac{u^2}{2}\right|_0^1 = 1/2$$

$$\sigma_U^2 = Var(U) = \int_0^1 u^2\, du - \mu_U^2 = \left|\frac{u^3}{3}\right|_0^1 - \frac{1}{4} = \frac{1}{3} - \frac{1}{4} = 1/12$$

La *F.G.M.* di U risulta

$$G_U(t) = M(\mathrm{e}^{Ut}) = \frac{1}{t}\int_0^1 \mathrm{e}^{(ut)}\, d(ut) = \frac{\mathrm{e}^t - 1}{t}$$

Espressione che per $t \to 0$ assume forma indeterminata del tipo $0/0$ e che conviene sviluppare in termini di potenze di t per ottenere i momenti

$$\begin{aligned} G_U(t) = {} & \frac{\mathrm{e}^t - 1}{t} \\ & = \frac{1}{t}\sum_{i=0}^{\infty} \frac{1}{i!} t^i - 1 = \sum_{i=1}^{\infty} \frac{1}{i!} t^{i-1} = \sum_{r=0}^{\infty} \frac{1/(r+1)}{r!} t^r \end{aligned}$$

Da cui si ottiene il generico momento dall'origine $M(U^r) = \frac{1}{r+1}$ confermando i risultati di $\mu_U = M(U) = 1/2$ e $M(U^2) = 1/3$.

5. <u>La *v.c.* "esponenziale negativa"</u> $X \sim Ex(\theta)$ con $\theta > 0$

In molte situazioni applicative quali quelle che riguardano l'affidabilità di sistemi o di analisi della sopravvivenza si considera la *v.c.* T che rappresenta il tempo che intercorre dalla messa in funzione del sistema e il manifestarsi di un evento "*A*" quale può essere un "guasto". Come modello interpretativo del fenomeno si può ricorrere a una condizione di omogeneità della probabilità del'evento "*A*" nel generico intervallo di tempo *t* considerato

$$P\big(A \in (u, u + du)\big) = \lambda du \;\text{ per }\; u \in (0, t]$$

Suddiviso l'intervallo $(0,t]$ in m $\{I_1, I_2, \cdots, I_m\}$ sotto-intervalli di ampiezza $\Delta u = t/m$, con m elevato, La probabilità che in $(0,t]$ non si manifesti l'evento A, e quindi la *v.c.* T assuma valori maggiori di t, è pari alla probabilità che negli m sotto-intervalli non si sia verificato l'evento A, probabilità quest'ultima data per $X = 0$ dalla distribuzione binomiale $X \sim Bi(m, \lambda\Delta u)$ con media $\mu_X = m \cdot \lambda\Delta u = \lambda t$. La distribuzione di X, dato l'elevato valore di m, può approssimarsi alla distribuzione di Poisson $X \sim Po(\lambda t)$, avendosi

$$P_T(T > t) = P_X(0) = \mathrm{e}^{-\lambda t}\frac{(\lambda t)^0}{0!} = e^{-\lambda t} \quad t > 0$$

Da cui otteniamo la *F.d.R.* di T come

$$F_T(t) = \begin{cases} 0 & \text{per} \quad t < 0 \\ 1 - \mathrm{e}^{-\lambda t} & \text{per} \quad t \geq 0 \end{cases}$$

Si tratta di una *v.c.* continua che si indica come $T \sim Es(\theta = \frac{1}{\lambda})$ denominata *v.c.* "esponenziale" o "esponenziale negativa". Il parametro $\theta > 0$ è pari alla media di T: $M(T) = \mu_T = \theta$.

Come *v.c.* esponenziale di base si considera $U \sim Es(1)$ avente *f.d.d.* e *F.d.R.* rispettivamente

$$f_U(u) = \begin{cases} 0 & \text{per} \quad u < 0 \\ \mathrm{e}^{-u} & \text{per} \quad u \geq 0 \end{cases}$$

$$F_U(u) = \begin{cases} 0 & \text{per} \quad u < 0 \\ 1 - \mathrm{e}^{-u} & \text{per} \quad u \geq 0 \end{cases}$$

La media di U si ottiene

$$\mu_U = M(U) = \int_0^{\infty} u\mathrm{e}^{-u} du = \Gamma(2) = 1! = 1$$

Si ricordi infatti che l'integrale euleriano di 2° specie vale

$$\int_0^{\infty} u^{\alpha-1}\mathrm{e}^{-u}du = \Gamma(\alpha)$$

dove $\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1)$ è la funzione "gamma". Se α e un numero intero abbiamo $\Gamma(\alpha) = (\alpha - 1)!$. In particolare si dimostra che $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$. Il momento dall'origine di ordine *r* risulta

$$M(U^r) = \int_0^{\infty} u^r e^{-u}du = \Gamma(r+1) = r! \quad \text{per } r = 0, 1, 2, \cdots$$

Da cui in particolare $M(U^2) = 2! = 2$ e quindi la varianza di *U* vale: $Var(U) = 2 - 1^2 = 1$.

Esistendo tutti i momenti esiste anche la *F.G.M.* che vale

$$G_U(t) = M(\mathrm{e}^{Ut}) = \int_0^{\infty} \mathrm{e}^{ut}\mathrm{e}^{-u}du = \int_0^{\infty} \mathrm{e}^{u(t-1)}du$$

Mediante una sostituzione di variabile $-w = u(t-1)$, da cui $du = [1/(1-t)]dw$, abbiamo

$$G_U(t) = (1-t)^{-1}\int_0^{\infty} \mathrm{e}^{-w}dw = (1-t)^{-1}$$

essendo $\int_0^{\infty} \mathrm{e}^{-w}dw = 1$. Svolgendo il calcolo delle derivate successive di $G_U(t)$ oppure sviluppando $G_U(t)$ in termini di potenze di *t* si ottengono tutti i momenti della *v.c.* $U \sim Es(1)$.

La *v.c.* esponenziale $Y \sim Es(\theta)$ con $\theta > 0$ parametro di "scala" può pensarsi ottenuta da *U* mediante la trasformazione lineare: $Y = \theta\, U$ risultando la *f.d.d.*, la *F.d.R.* e la *F.G.M.* date da

$$f_Y(y) = \begin{cases} 0 & \text{per } \; y < 0 \\ \dfrac{1}{\theta}\mathrm{e}^{-y/\theta} & \text{per } \; y \geq 0 \end{cases}$$

$$F_Y(y) = \begin{cases} 0 & \text{per} \quad y < 0 \\ 1 - \mathrm{e}^{-y/\theta} & \text{per} \quad t \geq 0 \end{cases}$$

$$G_Y(t) = G_Y(\theta t) = (1 - \theta t)^{-1}$$

Da cui i principali momenti di $Y \sim Es(\theta)$ risultano

$$\begin{cases} M(Y) = \theta M(U) = \theta \\ M(Y^2) = \theta^2 M(U^2) = 2\,\theta^2 \\ Var(Y) = \theta^2 Var(U) = \theta^2 \\ M(Y^r) = \theta^r M(U^r) = r!\,\theta^r \end{cases}$$

Osservazione

Oltre alla *v.c.* esponenziale che è continua e presenta supporto in $[0, \infty) \equiv \Re_+$ vi è una classe più ampia di *v.c.*, che anche la *v.c.* esponenziale ed è data dalle distribuzioni "gamma", di cui considereremo qui, in ordine, la classe "gamma a un parametro" e la classe "gamma a due parametri".

1. *Distribuzione Gamma a un parametro* (*di forma* $\alpha > 0$): $X \sim G_1(\alpha)$

   La *f.d.d.* è data da

   $$f_X(x) = \frac{1}{\Gamma(\alpha)} x^{\alpha-1} \mathrm{e}^{-x} \quad \text{per} \quad x \geq 0$$

   La *F.d.R.* non è esprimibile in forma analitica ma mediante la funzione "gamma incompleta" $\Gamma(x; \alpha) = \int_0^x w^{\alpha-1}\, \mathrm{e}^{-w} dw$

   $$F_X(x) = \frac{1}{\Gamma(\alpha)} \int_0^x w^{\alpha-1}\, \mathrm{e}^{-w} dw \quad \text{per} \quad x \geq 0$$

   Si dimostra che i momenti dall'origine di X sono dati da

$$M(X^r) = \frac{1}{\Gamma(\alpha)} \int_0^x w^r w^{\alpha-1}\, \mathrm{e}^{-w} dw$$
$$= \frac{1}{\Gamma(\alpha)} \int_0^x w^{(\alpha+r)-1}\, \mathrm{e}^{-w} dw$$

$$M(X^r) = \frac{\Gamma(\alpha + r)}{\Gamma(\alpha)} \quad \text{per } r = 0, 1, 2, \cdots$$

da cui, in particolare di ottiene

$$\begin{cases} \mu_X = M(X) = \alpha \\ M(X^2) = (\alpha + 1)\alpha \\ \sigma_X^2 = Var(X) = \alpha \end{cases}$$

La *F.G.M.* si ottiene come

$$G_X(t) = M(e^{Xt}) = \frac{1}{\Gamma(\alpha)} \int_0^\infty \mathrm{e}^{xt} x^{\alpha-1}\, \mathrm{e}^{-x} dx$$
$$= \frac{1}{\Gamma(\alpha)} \int_0^\infty x^{\alpha-1} \mathrm{e}^{-x(1-t)} dx = (1-t)^{-\alpha}$$

Per $\alpha = 1$ ritroviamo la *v.c.* esponenziale $G_1(1) \equiv Es(1)$.

La *f.d.d.* all'origine ($x = 0$) si presenta

$$f_X(0) = \begin{cases} \infty, \text{(asintoto verticale , monotona discendente)}: \alpha < 1 \\ 1 \qquad \text{(monotona discendente)}: \alpha = 1 \\ 0 \qquad \text{(campanulare asimmetrica)}: \alpha > 1 \end{cases}$$

Se $X_1 \sim G_1(\alpha_1)$ e $X_2 \sim G_1(\alpha_2)$ sono *v.c.* "indipendenti", la *v.c.* "somma" è ancora una *v.c.* "Gamma" con parametro $\alpha_1 + \alpha_2$

$$X_1 + X_2 \sim G_1(\alpha_1 + \alpha_2)$$

La *v.c.* $X \sim G_1(\alpha = n)$ è distribuita come la somma di *n* *v.c.* esponenziali *i.d.d.*: $X_i \sim G_1(1), i = 1, 2, \cdots, n$

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

2. *Distribuzione Gamma a due parametri* (*di forma* $\alpha > 0$, *di scala* $\theta > 0$): $Y \sim G_2(\alpha, \theta)$

   La *v.c.* Y è ottenuta da $X \sim G_1(\alpha)$ mediante la trasformazione lineare: $Y = \theta \cdot X$. Le *f.d.d.* e *F.d.R.* di *Y* si determinano da quelle di *X*, come

$$f_Y(y) = \frac{1}{\theta} f_X(y/\theta) = \frac{1}{\theta^\alpha \Gamma(\alpha)} y^{\alpha-1} \mathrm{e}^{-y/\theta} \ \text{ per } y \geq 0$$

$$F_Y(y) = F_X(y/\theta) \ \text{ per } y \geq 0$$

   La media e la varianza di *Y* risultano immediatamente date da quelle di *X*

$$\mu_Y = \theta \cdot \mu_X = \theta\alpha, \quad \sigma_Y^2 = \theta^2 \cdot \sigma_X^2 = \theta^2 \alpha$$

   e la *F.G.M.* di *Y* è data da quella di *X*, come

$$G_Y(t) = G_X(\theta t) = (1 - \theta t)^{-\alpha}$$

   L'andamento delle *f.d.d.* e *F.d.R.* di *Y* è del tutto analogo a quello della corrispondente *v.c.* "gamma a un parametro", dipendendo solo dal paramero di forma α.

6. La v.c. "gaussiana o normale" $X \sim N(\mu_X, \sigma_X^2)$ con $-\infty < \mu_X < \infty, 0 < \sigma_X^2$

È stato introdotto in precedenza (v. esempi 14, 14 e 19) la *v.c.* continua "normale standardizzata", indicata con $Z \sim N(0,1)$ che costituisce un modello di ampio impiego in statistica sia metodologica che applicata. Si tratta di una *v.c.* senza parametri e di semplice *f.d.d.*

$$\varphi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2} \ \text{ per } -\infty < z < \infty$$

L'andamento è campanulare, simmetrico rispetto all'origine, in cui presenta un punto di massimo e due flessi in corrispondenza di ± 1, tende a zero asintoticamente per $\pm\infty$, come è evidenziato in figura.

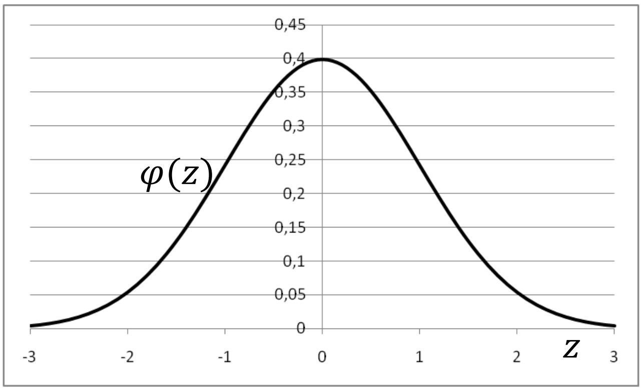
La *F.d.R.* di *Z*, non esprimibile in forma esplicita, è

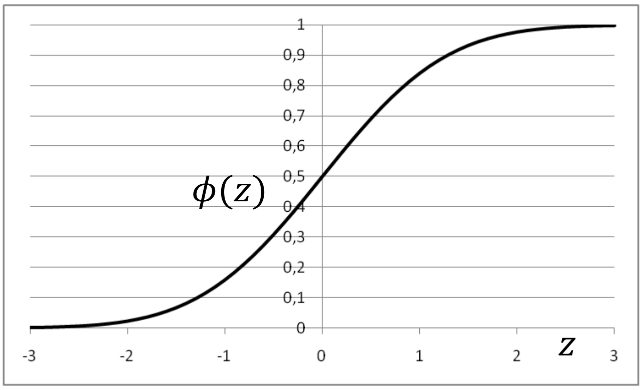
$$\phi(z) = \int_{-\infty}^{z} \varphi(u)\, du = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-u^2/2}\, du$$

La *F.G.M.* esiste, come esistono tutti i momenti, ed è pari a

$$G_Z(t) = \exp(t^2/2)$$

Da cui abbiamo: $\mu_Z = M(Z) = 0$; $\sigma_Z^2 = Var(Z) = M(Z^2) = 1$; $M(Z^3) = 0$ e $M(Z^4) = 3$. Il valore modale e quello mediano coincidono con l'origine ossia con il valor medio.





Se si considera una trasformazione lineare di Z si ottiene una *v.c.* $X = a + bZ$ per $b > 0$ avente *f.d.d.* e *F.d.R.*

$$f_X(x) = \frac{1}{b}\varphi\left(\frac{x-a}{b}\right) = \frac{1}{\sqrt{2\pi}\,b}\exp\left\{-\frac{1}{2}\left(\frac{x-a}{b}\right)^2\right\} \text{ per } \infty < x < \infty$$

$$F_X(x) = \phi\left(\frac{x-a}{b}\right) = \frac{1}{\sqrt{2\pi}\,b}\int_{-\infty}^{x}\exp\left\{-\frac{1}{2}\left(\frac{u-a}{b}\right)^2\right\} du$$

La *v.c.* X ha: $\mu_X = M(X) = a$; $\sigma_X^2 = Var(X) = b^2$, quindi $\mu_X \equiv a$, $\sigma_X^2 \equiv b^2$ che costituiscono i due parametri analiticamente indipendenti della distribuzione normale $X \sim N(\mu_X, \sigma_X^2)$.

La *F.G.M.* della *v.c. X* risulta

$$G_X(t) = \mathrm{e}^{\mu_X t} G_z(\sigma_X t) = \exp(\mu_X t + \sigma_X^2 t^2/2)$$

In particolare se si vogliono i momenti centrali di *X* è sufficiente, porre in $G_X(t)$, $\mu_X = 0$ e poi sviluppare in termini di potenze di *t* l'espressione $G_X(t) = \exp(\sigma_X^2 t^2/2)$.

Tra le proprietà della distribuzione normale ricordiamo che la classe di distribuzioni è "chiusa" rispetto all'operazione di "somma", proprietà che di può allargare ad ogni combinazione lineare di *v.c.* normali.

Nelle seguente tabella vengono riportati i valori di $\varphi(z)$ e di $\phi(z)$ per alcuni valori di *z*; si ricordi che, per la simmetria della distribuzione normale abbiano: $\varphi(-z) = \varphi(z)$; $\phi(-z) = 1 - \phi(z)$.

z	0	0,5	1	1,5	2	2,5	3	3,5	4
$\varphi(z)$	0,3989	0,3521	0,2420	0,1295	0,0540	0,0175	0,0044	0,0009	0,0001
$\phi(z)$	0,5000	0,6915	0,8413	0,9332	0,9772	0,9938	0,9987	0,9998	1,0000

Indicato con z_p il punto p-quantile della *v.c. Z*, ossia il tale che $P_Z(Z \le z_p) = \int_{-\infty}^{z_p}\varphi(z)\,dz = \phi(z_p)$ per alcuni prefissati valori di $p = 1 - \alpha$ abbiamo i dati della seguente tabella; si ricordi ancora che, per la simmetria della distribuzione normale abbiano: $z_{1-p} = z_\alpha = -z_p$.

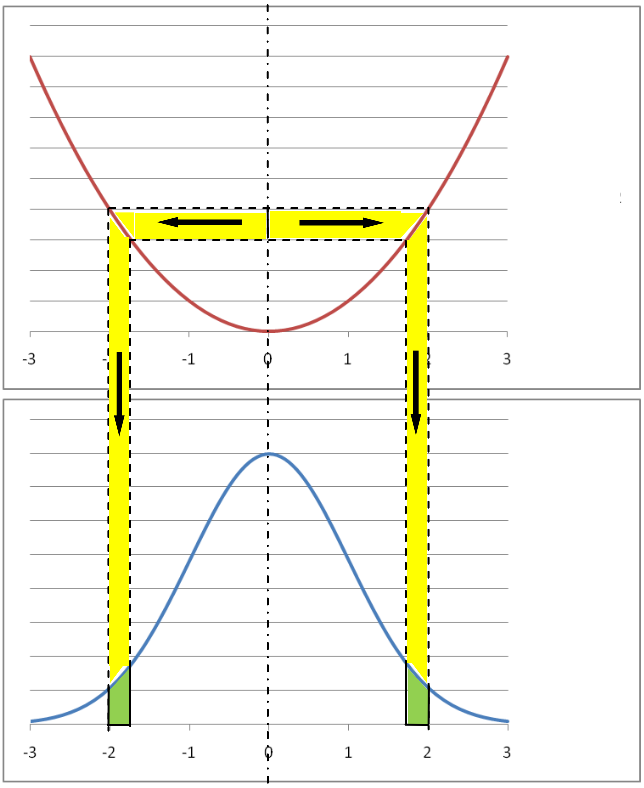
$p = \phi(z_p)$	0,5	0,6	0,7	0,75	0,8	0,85
z_p	0,000	0,253	0,524	0,674	0,842	1,036
$p = \phi(z_p)$	0,9	0,95	0,975	0,99	0,995	
z_p	1,282	1,645	1,960	2,326	2,576	

7. <u>La *v.c.* "chi quadrato"</u> $Y \sim \chi^2(g)$ con $g = 1, 2, \cdots$

Consideriamo la *v.c.* $Z \sim N(0,1)$ normale standardizzata con *f.d.d.* pari a

$$f_Z(z) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{z^2}{2}\right\} \text{ per} -\infty < z < \infty$$

Sia U la *v.c.* quadrato di *Z*: $U = Z^2$, essendo la *f.d.d.* di *Z* simmetrica rispetto all'origine è possibile ottenere la *f.d.d.* di *U* da quella di *Z*



$$f_U(u)du = 2f_Z(z)dz \quad con\ z = +u^{1/2}\ e\ dz = (1/2)u^{-1/2}du$$

da cui si ottiene

$$f_U(u) = 2\frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{u}{2}\right\}\left[(1/2)u^{-1/2}\right]$$

$$= \frac{1}{\sqrt{2}\sqrt{\pi}}u^{\frac{1}{2}-1}e^{-u/2} \qquad\qquad \text{per } u \geq 0$$

Confrontando quest'ultima espressione con la *f.d.d.* della *v.c.* "gamma a due parametri" $Y \sim G_2(\alpha, \theta)$

$$f_Y(y) = \frac{1}{\theta}f_X(y/\theta) = \frac{1}{\theta^\alpha \Gamma(\alpha)}y^{\alpha-1}\mathrm{e}^{-y/\theta} \ \text{ per } y \geq 0$$

abbiamo $U \sim G_2(\alpha = \frac{1}{2}, \theta = 2)$, essendo $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$.

La media e la varianza di *Y* risultano

$$\mu_Y = \alpha \cdot \theta = 1, \quad M(Y^2) = \alpha(\alpha+1)\theta^2 = 3, \qquad \sigma_Y^2 = 2$$

e la *F.G.M.* è

$$G_Y(t) = (1-2t)^{-1/2}$$

Si considerino ora g *v.c.* normali standardizzate $\{Z_i \sim N(0,1)\ i = 1,2,\cdots,g\}$ e *i.i.d.*. Le corrispondenti *v.c.* "quadrato" $\left\{U_i \sim G_2(\frac{1}{2}, 2)\ i = 1,2,\cdots,g\right\}$ sono *i.i.d.*.

La *v.c.* $V \sim \chi^2(g)$, detta "chi quadrato con parametro g" è data dalla "somma"

$$V = \sum_{i=1}^{g} U_i = \sum_{i=1}^{g} Z_i^2$$

Per la proprietà di additività delle *v.c.* "gamma", la *v.c.* V si distribuisce come una "gamma a due parametri" con: $\alpha = g \cdot \frac{1}{2} = g/2$ e $\theta = 2$.

La *f.d.d.* di $V \sim \chi^2(g) \sim G_2(\frac{g}{2}, 2)$ risulta

$$f_V(v) = \frac{1}{2^{g/2}\Gamma(g/2)} v^{\frac{g}{2}-1} \exp\left(-\frac{v}{2}\right) \quad \text{per} \quad v \geq 0$$

Si ricordi che: $\Gamma(1/2) = \sqrt{\pi}$, $\Gamma(3/2) = \frac{1}{2}\sqrt{\pi}$, $\Gamma(5/2) = \frac{3}{2}\Gamma(3/2), \cdots$

I principali momenti di $V \sim \chi^2(g)$ e la *F.G.M.*, risultano

$$\mu_V = M(V) = g; \;\; \sigma_V^2 = 2g; \;\; G_V(t) = (1-2t)^{-g/2}$$

Osservazioni

Oltre alla *v.c.* $V \sim \chi^2(g)$ sono di interesse in ambito statistico altre due *v.c.* strettamente collegate ad essa.

A) la *v.c.* $W \sim \chi^2(g)/g = V/g$ che può intendersi come media della successione di *v.c.* $\{U_i = Z_i^2;\; i = 1,2, \cdots, g\}$ o media dei quadrati della successione di *v.c.*, normali standardizzate, $\{Z_i;\; i = 1,2, \cdots, g\}$. Dato il legame lineare tra W e V, la *f.d.d.* di W si ottiene da quella di V mediante la relazione

$$f_w(w) = g \cdot f_V(gw) \quad 0 \leq w < \infty$$

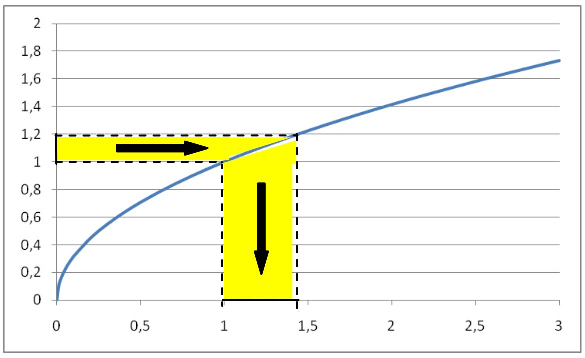
Il valor medio e la varianza di W risultano

$$M(W) = M\left(\frac{V}{g}\right) = \frac{M(V)}{g} = \frac{g}{g} = 1$$

$$Var(W) = Var\left(\frac{V}{g}\right) = \frac{Var(V)}{g^2} = \frac{2g}{g^2} = \frac{2}{g}$$

La media è costante, al variare di g, mentre la varianza, al divergere di g, tende a zero.

B) La $H \sim +\sqrt{W} = \sqrt{\chi^2(g)/g}$ che può intendersi come “media quadratica” (media potenziata di ordine due) della successione di *v.c.*, normali standardizzate, $\{Z_i;\ i = 1,2,\cdots,g\}$ o ancora "scarto quadratico medio" della stessa successione. Dato il legame monotono e biettivo tra H e W $(h = \sqrt{w} \rightarrow w = h^2)$ la *f.d.d.* di W si ottiene da quella di V mediante la relazione



$$f_H(h) = \left|\frac{dw}{dh}\right| f_W(h^2) = 2h f_W(h^2) \quad 0 \leq h < \infty$$

Sia nel caso A) sia nel caso B) la *F.d.R.*, rispettivamente, delle *v.c.* W e H si ottiene da quella della *v.c.* $V \sim \chi^2(g) \sim G_2(\frac{g}{2}, 2)$ che come è noto non è generalmente data in forma analitica ma richiede l'impiego di opportune tavole o algoritmi di calcolo

$$F_W(w) = F_V(v = g\, w)$$

$$F_H(h) = F_W(w = h^2) = F_V(v = g\, h^2)$$

Si osservi ancora come tutte tre le *v.c.* V, W e H presentano uno stesso parametro $g = 1, 2, \cdots$ che è denominato "gradi di libertà" (*g.d.l.*).

8. <u>La v.c. "t di Student"</u> $T \sim t(g)$ con $g = 1, 2, \cdots$

Consideriamo $(1 + g)$ *v.c.* normali standardizzate a componenti indipendenti $\{Z \sim N(0,1); Z_i \sim N(0,1)\ i = 1,2,\cdots,g\}$, dalle g componenti si ottenga la *v.c* $\chi^2(g) \sim \sum_{i=1}^{g} Z_i^2$ e da questa la *v.c* $H \sim \sqrt{\chi^2(g)/g}$ che indipendente da Z.

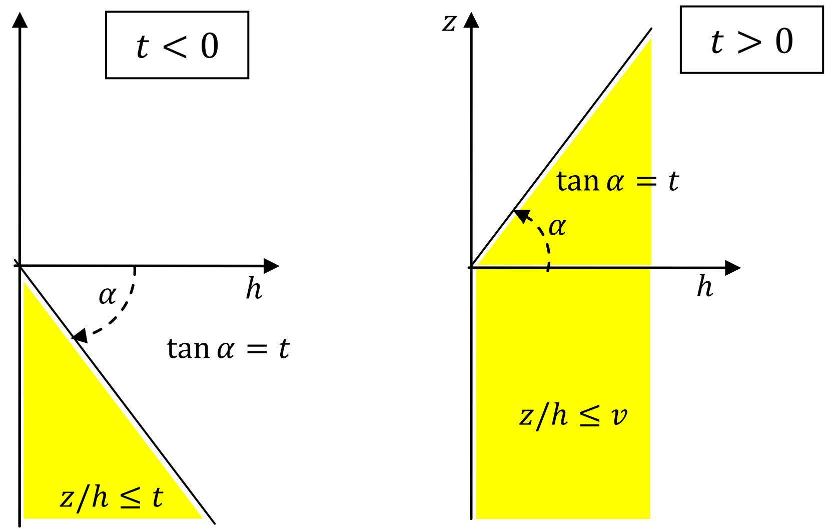
La *v.c* “t di Student” $T \sim t(g)$, dove “Student” è lo pseudonimo di W.S. Gosset (1876-1937) e definita come rapporto tra Z e H

$$T = t(g) \sim \frac{Z}{H} = \frac{N(0,1)}{\sqrt{\chi^2(g)/g}}$$

La *v.c* “t di Student” è una *v.c.* continua, campanulare, simmetrica, con mediana e moda nell’origine, similmente a quanto avviene per la $Z \sim N(0,1)$ e assume valori $-\infty < t < \infty$.

Lo studio della legge di distribuzione si può condurre ricordando che la Z e la H sono indipendenti con *f.d.d.* note e quindi la *f.d.d.* congiunta è data dal prodotto di quelle delle due componenti. Come è stato visto al paragrafo 21, mediante integrazione della *f.d.d* congiunta si ottiene la *F.d.R.* della *v.c.* T da cui si ottiene la corrispondente *f.d.d.* $f_T(t; g)$, con semplici passaggi algebrici !!!,

$$f_T(t; g) = \frac{\Gamma\big((g+1)/2\big)}{\sqrt{g\pi}\,\Gamma(g/2)}\left(1 + \frac{t^2}{g}\right)^{-(g+1)/2} \qquad -\infty < t < \infty$$



Per $g = 1$ non è definibile la media $M(T)$ in quanto $M(|T|)$ non converge a un numero finito. Anche in questo caso la *F.d.R.* non è generalmente data in forma analitica ma richiede l’impiego di opportune tavole o algoritmi di calcolo. Al divergere di g il denominatore tende, in probabilità a 1, ne consegue che $T \xrightarrow{g\to\infty} \sim Z = N(0,1)$, per scopi applicativi si può approssimare T alla normale standardizzata già per $g > 30$.

9. <u>La *v.c.* "F di Snedecor"</u> $U \sim F(g_1, g_2)$ con $g_1, g_2 = 1, 2, \cdots$

Si consideri due *v.c.* chi-quadrato $V_1 \sim \chi^2(g_1)$ e $V_2 \sim \chi^2(g_2)$ indipendenti e le due corrispondenti *v.c.* $W_1 \sim \chi^2(g_1)/g_1$ e . $W_2 \sim \chi^2(g_2)/g_2$. La *v.c.* "F di Snedecor", indicata $R \sim F_S(g_1, g_2)$ è data dal rapporto tra W_1 e W_2

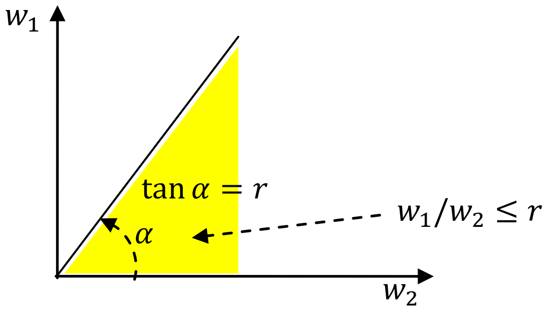
$$R \sim F_S(g_1, g_2) = \frac{W_1}{W_2} = \frac{\chi^2(g_1)/g_1}{\chi^2(g_2)/g_2}$$

La *v.c* "F di Snedecor" è una *v.c.* continua, campanulare in generale attorno al valore $1 \cong M(R) = g_2/(g_2 - 2) \;\; per \;\; g_2 > 2$ e assume valori $0 \le r < \infty$.

Lo studio della legge di distribuzione si può condurre ricordando che la W_1 e la W_2 sono indipendenti con *f.d.d.* note e quindi la *f.d.d.* congiunta è data dal prodotto di quelle delle due componenti. Come è stato visto al paragrafo 21, mediante integrazione della *f.d.d* congiunta si ottiene la *F.d.R.* della *v.c.* R da cui si ottiene la corrispondente *f.d.d.* $f_R(r; g_1, g_2)$:

$$f_R(r; g_1, g_2) = \frac{1}{k} \frac{r^{g_1/2-1}}{\left(1 + \frac{g_1}{g_2} r\right)^{(g_1+g_2)/2}} \quad 0 \le r < \infty$$

essendo k la costante di normalizzazione, dipendente da g_1 e g_2, che rende $\int_0^\infty f_R(r; g_1, g_2) dr = 1$.



Riferimenti bibliografici

Cifarelli D.M., (1988) *Elementi di calcolo delle probabilità*, Giappichelli, Torino.

Dall'Aglio G., (1987) *Calcolo delle probabilità*, Zanichelli, Bologna.

Parzen N.E., (1992) *Modern Probability Theory*, Wiley, N.Y (USA).

Ross S.M., (2007) *Calcolo delle probabilità*, seconda edizione, Apogeo, Milano.

Weiss N.A., (1992) *Calcolo delle probabilità*, Pearson Italia, Milano.

Sommario