

Corso di Laurea Triennale in Matematica

Università di Roma "La Sapienza"

Appunti di Probabilità

Testo ad uso degli studenti del Corso di Probabilità 2

A.A. 2015/2016

(Versione aggiornata il 5 Giugno 2016)

INDICE

Capitolo 1. Spazi di probabilità e variabili aleatorie reali: cenni su definizioni e proprietà fondamentali

- 1.1 Spazi di probabilità, variabili aleatorie e funzioni di ripartizione
- 1.2 Funzioni di ripartizione e distribuzioni di probabilità sulla retta
- 1.3 Trasformazioni di variabili aleatorie
- 1.4 I diversi tipi di distribuzioni di probabilità
- 1.5 Valori attesi di variabili aleatorie
- 1.6 Breve nota storica, critica, bibliografica

Capitolo 2. Distribuzioni di probabilità di vettori aleatori

- 2.1 Funzioni di distribuzione bidimensionali
- 2.2 Distribuzioni marginali e indipendenza stocastica
- 2.3 Caso assolutamente continuo
- 2.4 Densità condizionate e Formule di Bayes
- 2.5 Trasformazioni di densità congiunte
- 2.6 Somme di variabili aleatorie
- 2.7 Distribuzioni di probabilità in più dimensioni

Capitolo 3. Famiglie notevoli di distribuzioni di probabilità unidimensionali

- 3.1 Distribuzioni di probabilità discrete notevoli
- 3.2 Distribuzione uniforme
- 3.3 Distribuzione esponenziale
- 3.4 Distribuzione gamma
- 3.5 Distribuzione gaussiana (o normale)
- 3.6 Distribuzione del chi-quadro
- 3.7 Distribuzione t di Student
- 3.8 Distribuzione beta
- 3.9 Altre famiglie notevoli di distribuzioni di probabilità per variabili non negative

Capitolo 4. Questioni circa vettori di variabili gaussiane

- 4.1 Spazi di probabilità, variabili aleatorie e funzioni di ripartizione
- 4.2 Distribuzioni gaussiane in due dimensioni
- 4.3 Trasformazioni ortogonali di campioni gaussiani e Teorema di Cochran
- 4.4 Un'applicazione statistica del Teorema di Cochran

Capitolo 5. Il metodo della Funzione Caratteristica e il Teorema Centrale del Limite per variabili aleatorie i.i.d.

- 5.1 Definizione di Funzione Caratteristica e sue proprietà elementari
- 5.2 Funzioni caratteristiche di distribuzioni di probabilità notevoli
- 5.3 Proprietà fondamentali delle funzioni caratteristiche e Teorema di Lindberg-Lévy
- 5.4 Proprietà di stabilità rispetto alla somma

Capitolo 6. Processi di Poisson e alcune loro proprietà

- 6.1 Definizione e proprietà fondamentali del Processo di Poisson omogeneo
- 6.2 Statistiche ordinate di variabili i.i.d. e partizioni casuali di intervalli
- 6.3 Processi di Poisson omogenei e partizioni casuali di intervalli

1 Spazi di probabilità e variabili aleatorie reali: definizioni e proprietà fondamentali

Si presuppone che il lettore abbia già maturato, dal punto di vista euristico, l'intuizione circa nozioni fondamentali della Teoria della Probabilità, quali *spazio di probabilità*, *variabile aleatoria*, *distribuzione di probabilità*, *valore atteso e momento di una distribuzione*. In quanto segue verrà dato un breve cenno alla formalizzazione di tali nozioni da un punto di vista assiomatico. Ciò permetterà di fornire, in particolare, un quadro di riferimento per gli argomenti che verranno svolti nei successivi capitoli.

1.1 Spazi di probabilità, variabili aleatorie e funzioni di ripartizione

Sia $\Omega \equiv \{\omega\}$ un arbitrario spazio di punti e sia $\mathcal{P}(\Omega)$ la famiglia delle sue parti; sia inoltre $\mathcal{F} \subset \mathcal{P}(\Omega)$ una famiglia di sottoinsiemi di Ω .

Definizione 1.1

Diremo che \mathcal{F} costituisce una σ -algebra se sono verificate le seguenti condizioni

- a) $\Omega \in \mathcal{F}$
- b) $E \in \mathcal{F} \Rightarrow \tilde{E} \in \mathcal{F}$
- c) $E_1, E_2, \dots \in \mathcal{F} \Rightarrow \bigcup_{h=1}^{\infty} E_h \in \mathcal{F}$

Diremo invece che \mathcal{F} costituisce un'algebra se la condizione c), di chiusura rispetto alle unioni numerabili, viene sostituita dalla più debole condizione c') di chiusura rispetto alle unioni finite:

$$c') \quad n \in \mathbb{N}, E_1, E_2, \dots, E_n \in \mathcal{F} \Rightarrow \bigcup_{h=1}^n E_h \in \mathcal{F}$$

Osservazione 1.1: utilizzando le identità di De Morgan possiamo stabilire che, se \mathcal{F} è una σ -algebra, allora vale anche l'implicazione

$$d) \quad E_1, E_2, \dots \in \mathcal{F} \Rightarrow \bigcap_{h=1}^{\infty} E_h \in \mathcal{F}$$

Osservazione 1.2: $\mathcal{P}(\Omega)$ è ovviamente una σ -algebra

Osservazione 1.3: L'intersezione fra σ -algre costituisce ancora una σ -algebra

Definizione 1.2

Uno *spazio misurabile* è una coppia (Ω, \mathcal{F}) dove $\Omega \equiv \{\omega\}$ è un arbitrario spazio di punti e $\mathcal{F} \subset \mathcal{P}(\Omega)$ è una σ -algebra di suoi sottoinsiemi. Uno *spazio di probabilità* è una terna $(\Omega, \mathcal{F}, \mathbb{P})$ dove \mathbb{P} è una *misura di probabilità* su (Ω, \mathcal{F}) , cioè un'applicazione $\mathbb{P} : \mathcal{F} \rightarrow [0, 1]$ che soddisfa le condizioni:

- e) $P(\Omega) = 1$
 f) se $E_1, E_2, \dots \in \mathcal{F}$ sono insiemi a due a due disgiunti, allora

$$\mathbb{P}\left(\bigcup_{h=1}^{\infty} E_h\right) = \sum_{h=1}^{\infty} \mathbb{P}(E_h). \quad (1)$$

La proprietà in (1), detta *additività numerabile* o σ -*additività*, implica l'*additività finita*; non vale però, generalmente, l'implicazione inversa.

Sappiamo che, nella teoria della probabilità, gli *eventi (risultati)* in un esperimento aleatorio possono essere rappresentati come sottoinsiemi di un opportuno spazio di probabilità. Nel caso di uno spazio di probabilità finito, qualunque sottoinsieme di Ω può essere visto come *evento* e i sottoinsiemi costituiti dai singoli elementi di Ω sono gli *eventi elementari*. Nel caso generale, invece, vengono considerati *eventi* soltanto quei sottoinsiemi di Ω che appartengano ad \mathcal{F} , cioè quei sottoinsiemi E per cui sia effettivamente definita la probabilità $\mathbb{P}(E)$. Tali sottoinsiemi vengono anche detti *misurabili*.

Corrispondentemente, viene data la seguente definizione

Definizione 1.3

Una *variabile aleatoria* X su uno spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$ è un'applicazione $X : \Omega \rightarrow \mathbb{R}$ tale che, $\forall x \in \mathbb{R}$, resti soddisfatta la condizione

$$\{\omega \in \Omega | X(\omega) \leq x\} \in \mathcal{F}. \quad (2)$$

Notiamo che la validità della condizione (2) ci permette di considerare, al variare di $x \in \mathbb{R}$, le quantità $\mathbb{P}(\{\omega \in \Omega | X(\omega) \leq x\})$.

Nel linguaggio corrente scriveremo più semplicemente

$$\{\omega \in \Omega | X(\omega) \leq x\} = \{X \leq x\}$$

e

$$\mathbb{P}(\{\omega \in \Omega | X(\omega) \leq x\}) = \mathbb{P}\{X \leq x\}$$

Definizione 1.4

Sia X una variabile aleatoria su uno spazio di probabilità (Ω, \mathcal{F}, P) . Si chiama *funzione di ripartizione* di X la funzione di variabile reale (a valori in $[0, 1]$) definita da

$$F_X(x) := \mathbb{P}\{X \leq x\}.$$

Prima di proseguire è opportuno individuare quali condizioni, da un punto di vista analitico, debbano necessariamente essere soddisfatte da una funzione reale di variabile reale F che risulti la funzione di ripartizione di una qualche variabile aleatoria X .

A tale scopo si deve osservare preliminarmente che una misura di probabilità \mathbb{P} (\mathbb{P} σ -additiva) su uno spazio misurabile (Ω, \mathcal{F}) gode delle proprietà di *monotonia* e di *continuità*. Si ha cioè

Proposizione 1.1

Siano $A_1, A_2 \in \mathcal{F}$ due sottoinsiemi tali che $A_1 \subseteq A_2$. Allora $\mathbb{P}(A_1) \leq \mathbb{P}(A_2)$.

Dimostrazione

A_2 si può decomporre come unione disgiunta di A_1 e A_2/A_1 dove anche A_2/A_1 appartiene ad \mathcal{F} . Quindi

$$\mathbb{P}(A_2) = \mathbb{P}(A_1) + \mathbb{P}(A_2/A_1) \geq \mathbb{P}(A_1).$$

Proposizione 1.2

Siano $A_1 \subseteq A_2 \subseteq \dots$ e $B_1 \supseteq B_2 \supseteq \dots$ due *catene* di elementi di \mathcal{F} e sia \mathbb{P} una misura di probabilità su \mathcal{F} . Allora valgono le seguenti relazioni

a)

$$\mathbb{P}\left(\bigcup_{h=1}^{\infty} A_h\right) = \lim_{k \rightarrow \infty} \mathbb{P}(A_k)$$

b) $\bigcap_{h=1}^{\infty} B_h \in \mathcal{F}$ e

$$\mathbb{P}\left(\bigcap_{h=1}^{\infty} B_h\right) = \lim_{k \rightarrow \infty} \mathbb{P}(B_k)$$

Dimostrazione

a) Notiamo innanzitutto che il sottoinsieme $\bigcup_{h=1}^{\infty} A_h$ appartiene ad \mathcal{F} essendo

questa una σ -algebra; e quindi è ben definita la quantità $\mathbb{P}\left(\bigcup_{h=1}^{\infty} A_h\right)$. Inoltre, essendo $\mathbb{P}(A_1), \mathbb{P}(A_2), \dots$ una successione non decrescente e limitata ($\mathbb{P}(A_k) \leq 1$), certamente esiste il limite $\lim_{k \rightarrow \infty} \mathbb{P}(A_k)$. Consideriamo la successione di sottoinsiemi definiti come segue

$$C_1 = A_1, C_2 = A_2/A_1, C_3 = A_3/A_2, \dots$$

Tali insiemi sono a due a due disgiunti e tali che, per ogni $k = 2, 3, \dots$, vale

$$A_k = \bigcup_{h=1}^k C_h \text{ e } \bigcup_{h=1}^{\infty} A_h = \bigcup_{h=1}^{\infty} C_h.$$

Quindi

$$\mathbb{P}(A_k) = \sum_{h=1}^k \mathbb{P}(C_h), \mathbb{P}\left(\bigcup_{h=1}^{\infty} A_h\right) = \mathbb{P}\left(\bigcup_{h=1}^{\infty} C_h\right)$$

e, in virtù dell'ipotesi di σ -additività, risulta

$$\mathbb{P}\left(\bigcup_{h=1}^{\infty} C_h\right) = \sum_{h=1}^{\infty} \mathbb{P}(C_h).$$

D'altra parte, per definizione di somma della serie,

$$\sum_{h=1}^{\infty} \mathbb{P}(C_h) = \lim_{k \rightarrow \infty} \sum_{h=1}^k \mathbb{P}(C_h) = \lim_{k \rightarrow \infty} \mathbb{P}(A_k)$$

La dimostrazione del punto b) si può ottenere dalla precedente, attraverso la relazione di De Morgan.

Sia ora F una assegnata funzione reale di variabile reale.

Proposizione 1.3

Se F è la funzione di ripartizione di una qualche variabile aleatoria X (definita su un qualche spazio di probabilità (Ω, \mathcal{F}, P)), devono allora essere soddisfatte le seguenti condizioni

a) F è non decrescente:

$$x' < x'' \Rightarrow F(x') \leq F(x'')$$

b)

$$\lim_{x \rightarrow -\infty} F(x) = 0, \quad \lim_{x \rightarrow +\infty} F(x) = 1$$

c) F è continua da destra:

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

Dimostrazione

a) Per $x' < x''$

$$\{\omega \in \Omega | X(\omega) \leq x'\} \subseteq \{\omega \in \Omega | X(\omega) \leq x''\}.$$

Quindi $F(x') \leq F(x'')$ segue dalla monotonia della \mathbb{P} .

b) Consideriamo la catena decrescente dei sottoinsiemi della forma

$$B_k = \{\omega \in \Omega | X(\omega) \leq -k\}.$$

Allora $\bigcap_{k=1}^{\infty} B_k = \emptyset$ e, per la continuità di \mathbb{P} , risulta

$$\lim_{k \rightarrow \infty} \mathbb{P}(B_k) = \mathbb{P}\left(\bigcap_{k=1}^{\infty} B_k\right) = \mathbb{P}(\emptyset) = 0$$

La relazione $\lim_{x \rightarrow +\infty} F(x) = 1$ si verifica analogamente considerando la catena crescente dei sottoinsiemi della forma

$$A_k = \{\omega \in \Omega | X(\omega) \leq k\}$$

e tenendo conto che $\bigcup_{k=1}^{\infty} \{\omega \in \Omega | X(\omega) \leq k\} = \Omega$ con $\mathbb{P}(\Omega) = 1$.

c) Consideriamo ora, per ogni fissato valore $x_0 \in \mathbb{R}$, la catena decrescente dei sottoinsiemi della forma

$$B_k = \{\omega \in \Omega | X(\omega) \leq x_0 + \frac{1}{k}\}.$$

Osserviamo che risulta

$$\{\omega \in \Omega | X(\omega) \leq x_0\} = \bigcap_{k=1}^{\infty} B_k$$

e quindi

$$F(x_0) = \mathbb{P}\{\omega \in \Omega | X(\omega) \leq x_0\} = \lim_{k \rightarrow \infty} \mathbb{P}(B_k) = \lim_{k \rightarrow \infty} F(x_0 + \frac{1}{k}) = \lim_{x \rightarrow x_0^+} F(x).$$

Daremo successivamente un cenno alla dimostrazione del fatto che l'insieme delle condizioni a), b) e c) è anche sufficiente affinché F sia la funzione di ripartizione di una qualche variabile aleatoria X .

Consideriamo ora, quindi, una funzione F che sia la funzione di ripartizione di una qualche variabile aleatoria X e concentriamo l'attenzione sull'insieme

$$\mathcal{D}_F = \{x \in \mathbb{R} | x \text{ punto di discontinuità per } F\}.$$

Siccome F è monotona, sappiamo che \mathcal{D}_F - se non è vuoto - contiene al più un'infinità numerabile di punti x_1, x_2, \dots e che le discontinuità possono essere soltanto di prima specie. Più precisamente, dal momento che F è non decrescente, si avrà che un punto di discontinuità $x_i \in \mathcal{D}_F \subset \mathbb{R}$ deve essere tale che

$$F(x_i) = \lim_{x \rightarrow x_i^+} F(x) > \lim_{x \rightarrow x_i^-} F(x).$$

Osserviamo allora che, essendo

$$\bigcup_{k=1}^{\infty} \{\omega \in \Omega | X(\omega) \leq x_i - \frac{1}{k}\} = \{\omega \in \Omega | X(\omega) < x_i\},$$

si ha

$$\mathbb{P}\{X < x_i\} = \lim_{k \rightarrow \infty} \mathbb{P}\{X \leq x_i - \frac{1}{k}\} = \lim_{x \rightarrow x_i^-} F(x).$$

Possiamo dunque concludere che punti di discontinuità per F sono tutti e soli gli x_i tali che

$$\mathbb{P}\{X \leq x_i\} > \mathbb{P}\{X < x_i\}.$$

Cioè, siccome $\{X \leq x_i\}/\{X < x_i\} = \{X = x_i\}$, questi punti sono tali che

$$\mathbb{P}\{X = x_i\} > 0 :$$

l'ampiezza del salto della F in un suo punto di discontinuità x_i è uguale alla probabilità che X assuma il valore x_i .

Quando F è ovunque *continua* ciascun singolo valore $x \in \mathbb{R}$ si presenta con probabilità 0, e viceversa.

Le variabili aleatorie la cui funzione di distribuzione è *costante a tratti*, sono invece quelle con *distribuzione discreta*: con probabilità uguale a 1 esse assumono un valore su cui è concentrata una probabilità positiva.

Osserviamo infine che, dati due valori reali $a < b$, vale

$$\mathbb{P}\{a < X \leq b\} = F(b) - F(a).$$

Nel particolare caso in cui F è continua, avendo ciascun singolo valore probabilità di presentarsi nulla, si ha

$$\begin{aligned} F(b) - F(a) &= \mathbb{P}\{a < X \leq b\} = \mathbb{P}\{a \leq X \leq b\} \\ &= \mathbb{P}\{a \leq X < b\} = \mathbb{P}\{a < X < b\}. \end{aligned}$$

1.2 Funzioni di ripartizione e distribuzioni di probabilità sulla retta

La nozione di funzione di ripartizione ha un ruolo fondamentale nella teoria della probabilità. In particolare, sotto la condizione di σ -additività, possiamo dire che F_X individua univocamente la distribuzione di probabilità di X . Per formalizzare e motivare tale affermazione dobbiamo innanzitutto definire, da un punto di vista assiomatico, che cosa si possa intendere per *distribuzione di probabilità* di una variabile aleatoria. A tale scopo, si rendono necessari alcuni cenni circa definizioni e questioni fondamentali della teoria della misura.

Consideriamo due spazi *misurabili*, cioè due coppie $(\Omega, \mathcal{F}), (\Omega', \mathcal{F}')$, ed una applicazione $\varphi : \Omega \rightarrow \Omega'$.

Definizione 2.1

φ è detta \mathcal{F}/\mathcal{F}' -misurabile se, per ogni $E' \in \mathcal{F}'$, è soddisfatta la condizione

$$\varphi^{-1}(E') \equiv \{\omega \in \Omega \mid \varphi(\omega) \in E'\} \in \mathcal{F}.$$

Sia ora fissata una misura di probabilità \mathbb{P} su \mathcal{F} e sia $\varphi : \Omega \rightarrow \Omega'$ un'applicazione \mathcal{F}/\mathcal{F}' -misurabile.

Ha quindi senso considerare su \mathcal{F}' la funzione di insieme definita dalla posizione

$$\mathbb{P}_\varphi(E') = \mathbb{P}\{\omega \in \Omega \mid \varphi(\omega) \in E'\} = \mathbb{P}(\varphi^{-1}(E')), \quad E' \in \mathcal{F}'.$$

Si può verificare che \mathbb{P}_φ è anche essa una misura di probabilità. \mathbb{P}_φ viene detta *misura di probabilità indotta* dalla φ ; in base alla definizione stessa, risulta anche naturale indicarla con il simbolo $\mathbb{P} \circ \varphi^{-1}$.

Sia ancora (Ω, \mathcal{F}) uno spazio misurabile e sia $\mathcal{A} \subset \mathcal{F}$ una arbitraria famiglia di sottoinsiemi misurabili di Ω .

Siccome l'intersezione fra σ -algebre costituisce ancora una σ -algebra (Osservazione 1.3), ha senso considerare la σ -algebra ottenuta quale intersezione di tutte quelle σ -algebre \mathcal{G} (sotto σ -algebre di \mathcal{F}) che soddisfano la condizione $\mathcal{A} \subset \mathcal{G}$. Tale σ -algebra viene indicata con il simbolo $\sigma(\mathcal{A})$ e viene detta *σ -algebra generata da \mathcal{A}* .

Da tale definizione segue dunque:

sia $\mathcal{G} \subset \mathcal{F}$ una σ -algebra tale che $\mathcal{A} \subset \mathcal{G}$. Allora vale anche $\sigma(\mathcal{A}) \subset \mathcal{G}$.

E' ovvio che se \mathcal{A} è essa stessa una σ -algebra allora $\sigma(\mathcal{A}) = \mathcal{A}$. Altrimenti si ha necessariamente $\mathcal{A} \subset \sigma(\mathcal{A})$.

Definizione 2.2

Una famiglia di sottoinsiemi $\mathcal{A} \subset \mathcal{F}$ costituisce un *π -sistema* se è chiusa sotto l'operazione di intersezione, cioè se vale l'implicazione

$$A_1 \in \mathcal{A}, A_2 \in \mathcal{A} \Rightarrow A_1 \cap A_2 \in \mathcal{A}$$

Definizione 2.3

Una famiglia di sottoinsiemi $\mathcal{L} \subset \mathcal{F}$ costituisce un *λ -sistema* se soddisfa le seguenti condizioni:

a)

$$\Omega \in \mathcal{L}$$

b) per $A, B \in \mathcal{L}$ e tali che $A \subset B$

$$B/A \in \mathcal{L}$$

c) per ogni catena crescente $A_1 \subset A_2 \subset \dots$ di sottoinsiemi in \mathcal{L} , si ha

$$\bigcup_{k=1}^{\infty} A_k \in \mathcal{L}.$$

Una σ -algebra è, al tempo stesso, un λ -sistema ed π -sistema. Notiamo più precisamente che una famiglia di sottoinsiemi è una σ -algebra se e solo se è contemporaneamente un π -sistema e un λ -sistema.

Risulta varie volte utile, per due famiglie di sottoinsiemi $\mathcal{A}, \mathcal{G} \subset \mathcal{F}$ tali che $\mathcal{A} \subset \mathcal{G}$, poter stabilire se vale anche $\sigma(\mathcal{A}) \subset \mathcal{G}$. Ciò è automaticamente garantito se \mathcal{G} è una σ -algebra. Un risultato che risulta utile per diverse questioni rilevanti è il seguente *Teorema λ - π di Dynkin*. Questo permette di indebolire, in quanto detto sopra, la condizione che \mathcal{G} sia una σ -algebra, aggiungendo però la condizione che \mathcal{A} sia un π -sistema:

Teorema 2.1

Siano $\mathcal{A}, \mathcal{G} \subset \mathcal{F}$ tali che

- \mathcal{G} è un λ -sistema
- \mathcal{A} π -sistema
- $\mathcal{A} \subset \mathcal{G}$

Allora risulta $\sigma(\mathcal{A}) \subset \mathcal{G}$.

Come conseguenza del Teorema 2.1 abbiamo il seguente risultato che, a sua volta, si rivela utile in molte situazioni.

Teorema 2.2

Sia $\mathcal{A} \subset \mathcal{F}$ una famiglia di sottoinsiemi e siano \mathbb{P}_1 e \mathbb{P}_2 due misure di probabilità su \mathcal{F} . Se risulta

$$\mathbb{P}_1(A) = \mathbb{P}_2(A), \forall A \in \mathcal{A}$$

e \mathcal{A} è un π -sistema, allora $\mathbb{P}_1(A) = \mathbb{P}_2(A), \forall A \in \sigma(\mathcal{A})$.

Sulla base del Teorema 2.1., questo risultato si dimostra facilmente notando che la famiglia

$$\mathcal{G} = \{A \in \mathcal{F} | \mathbb{P}_1(A) = \mathbb{P}_2(A)\}$$

risulta essere un λ -sistema.

Consideriamo la famiglia \mathcal{I} costituita dagli intervalli della retta, della forma $(-\infty, x], x \in \mathbb{R}$. E' utile notare subito che \mathcal{I} costituisce, ovviamente, un π -sistema.

Quale *famiglia degli insiemi di Borel della retta reale*, si può intendere la σ -algebra (di sottoinsiemi della retta) generata dalla famiglia \mathcal{I} . Tale σ -algebra $\sigma(\mathcal{I})$ viene usualmente indicata con il simbolo \mathcal{B} .

La *misura di Lebesgue* su \mathcal{B} è definita come la funzione di insieme $\lambda : \mathcal{B} \rightarrow \mathbb{R}$, individuata dalle seguenti proprietà:

- i) λ è σ -additiva
- ii) $\lambda(B) = (b - a)$ per $B = (a, b]$.

Estendendo opportunamente il precedente Teorema 2.2 dal campo delle misure di probabilità σ -additive al campo più generale delle misure σ -additive non necessariamente finite, si dimostra che λ è individuata univocamente in base alle proprietà i) e ii).

Torniamo ora a considerare una variabile aleatoria X , definita su uno spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$.

Possiamo vedere X quale applicazione $X : \Omega \rightarrow \mathbb{R}$ \mathcal{F}/\mathcal{B} -misurabile e possiamo così dare una definizione assiomatica dell'oggetto che, nel linguaggio corrente della probabilità, è detto *distribuzione di probabilità* di X : quest'ultima può essere vista come una misura di probabilità sullo spazio $(\mathbb{R}, \mathcal{B})$ secondo la seguente

Definizione 2.4

Sia \mathbb{P}_X la misura di probabilità su $(\mathbb{R}, \mathcal{B})$ definita dalla relazione

$$\mathbb{P}_X(B) := \mathbb{P}\{X \in B\} = \mathbb{P}\{\omega \in \Omega | X(\omega) \in B\}$$

per $B \in \mathcal{B}$. \mathbb{P}_X è detta *misura di probabilità indotta* dall'applicazione X . Possiamo ora intendere la "distribuzione di probabilità di X " semplicemente come la misura \mathbb{P}_X su $(\mathbb{R}, \mathcal{B})$.

Vogliamo, a questo punto, rispondere a due domande.

Innanzitutto: nel precedente paragrafo avevamo definito una variabile aleatoria su (Ω, \mathcal{F}) come applicazione $X : \Omega \rightarrow \mathbb{R}$ tale che valga la condizione (2): $\forall x \in \mathbb{R}, \{\omega \in \Omega | X(\omega) \leq x\} \in \mathcal{F}$.

Per quale motivo possiamo vedere X , come citato poco fa, quale applicazione \mathcal{F}/\mathcal{B} -misurabile?

La risposta è che, anche se la condizione (2) appare più debole della \mathcal{F}/\mathcal{B} -misurabilità, le due condizioni risultano in effetti equivalenti.

Consideriamo a questo proposito la famiglia \mathcal{G} costituita da tutti gli insiemi $G \in \mathcal{B}$ tali che

$$X^{-1}(G) = \{\omega \in \Omega | X(\omega) \in G\} \in \mathcal{F}.$$

Dire che X è tale che valga la condizione (2) significa dunque affermare $\mathcal{I} \subset \mathcal{G}$.

E' possibile inoltre verificare che la famiglia \mathcal{G} costituisce un λ -sistema.

Essendo allora \mathcal{I} un π -sistema e \mathcal{G} un λ -sistema, il Teorema 2.1 assicura che anche

$$\mathcal{B} = \sigma(\mathcal{I}) \subset \mathcal{G};$$

cioè che X è una applicazione \mathcal{F}/\mathcal{B} -misurabile.

La seconda domanda è quella che ci eravamo posti all'inizio del paragrafo: in quale senso possiamo dire che F_X individua la distribuzione di probabilità di X ?

La risposta dipende ancora dal fatto che \mathcal{I} è un π -sistema ed è data dal precedente Teorema 2.2: la funzione di ripartizione $F_X(x)$ individua la misura di probabilità \mathbb{P}_X sugli insiemi appartenenti ad \mathcal{I} . Ciò, essendo \mathcal{I} un π -sistema, permette a sua volta di ricostruire univocamente la misura \mathbb{P}_X su tutta la famiglia \mathcal{B} .

1.3 Trasformazioni di variabili aleatorie

In questo paragrafo X indica una fissata variabile aleatoria su uno spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$, con funzione di distribuzione F_X . Indichiamo con \mathcal{S} il *supporto* della distribuzione di X , cioè l'intervallo $[x_{\min}, x_{\max}]$ con

$$x_{\min} = \inf\{x | F(x) > 0\}, x_{\sup} = \sup\{x | F(x) < 1\}.$$

Consideriamo poi un'arbitraria funzione $\phi : \mathbb{R} \rightarrow \mathbb{R}$, \mathcal{B}/\mathcal{B} misurabile; una tale funzione viene usualmente detta direttamente *misurabile*, senza ulteriori

specificazioni. Di tale tipo sono, in particolare, tutte le funzioni monotone e tutte le funzioni continue.

Vogliamo quindi considerare l'applicazione $Z : \Omega \rightarrow \mathbb{R}$ costruita come funzione composta

$$Z(\omega) := (\phi \circ X)(\omega) = \phi(X(\omega)).$$

Innanzitutto osserviamo che Z è, anch'essa, una variabile aleatoria su $(\Omega, \mathcal{F}, \mathbb{P})$. Infatti, come si verifica direttamente, è \mathcal{F}/\mathcal{B} -misurabile, in quanto tale è X .

La distribuzione di probabilità di Z , \mathbb{P}_Z , si può vedere come misura di probabilità indotta nelle due diverse accezioni $(\mathbb{P}_X)_\phi$ e $\mathbb{P}_{\phi \circ X}$.

Noi consideriamo ora il caso particolare in cui ϕ sia continua e strettamente monotona su $[x_{\min}, x_{\max}]$, così da ammettere su tale intervallo una funzione inversa $\psi = \phi^{-1}$, anch'essa continua e strettamente monotona. Iniziamo, per fissare le idee, con l'assumere ϕ strettamente crescente.

Per la funzione di ripartizione di Z vale allora la relazione

$$\begin{aligned} F_Z(z) &:= \mathbb{P}\{Z \leq z\} = \mathbb{P}\{\phi(X) \leq z\} \\ &= \mathbb{P}\{X \leq \psi(z)\} = F_X(\psi(z)). \end{aligned} \quad (3)$$

Nel caso ϕ strettamente decrescente avremmo invece

$$\begin{aligned} F_Z(z) &= \mathbb{P}\{\phi(X) \leq z\} \\ &= \mathbb{P}\{X \geq \psi(z)\} = \mathbb{P}\{X = \psi(z)\} + \mathbb{P}\{X > \psi(z)\} \\ &= \mathbb{P}\{X = \psi(z)\} + 1 - F_X(\psi(z)), \end{aligned}$$

e $F_Z(z)$ si ridurrebbe a $1 - F_X(\psi(z))$ nel caso in cui anche F_X è continua.

Prima di proseguire, è utile focalizzare l'attenzione su una distribuzione di probabilità notevole, che verrà incontrata spesso nel seguito.

Definizione 3.1

Diciamo che una variabile aleatoria U segue una *distribuzione uniforme* sull'intervallo $[0, 1]$ se la relativa funzione di distribuzione è data da

$$F_U(x) = \begin{cases} 0 & x < 0 \\ x & 0 \leq x \leq 1 \\ 1 & x > 1 \end{cases} .$$

Notiamo che tale F_U , oltre che essere continua su tutta la retta, è anche strettamente crescente sul supporto $[0, 1]$. In tale intervallo essa ammette inversa F_U^{-1} , che ovviamente coincide con la F_U stessa

Osservazione 3.2

Consideriamo il caso in cui F_X è continua e strettamente crescente sul suo supporto. La variabile aleatoria $U = F_X(X)$ prende, ovviamente, valori soltanto nell'intervallo $[0, 1]$. Risulta avere, inoltre, una distribuzione uniforme sull'intervallo $[0, 1]$. Infatti, per $0 \leq x \leq 1$,

$$F_U(x) = \mathbb{P}\{F_X(X) \leq x\} = \mathbb{P}\{X \leq F_X^{-1}(x)\} = x.$$

Sia ora F una funzione che soddisfa le ipotesi a),b),c) elencate nell'enunciato della Proposizione 3; aggiungiamo per il momento anche la condizione che sia continua e strettamente crescente nel suo supporto, cosicchè possiamo considerare la sua inversa F^{-1} , anch'essa continua e strettamente crescente.

Supponiamo ora che V sia una qualunque variabile aleatoria, definita su un qualunque spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$, tale che la sua distribuzione di probabilità sia uniforme su $[0, 1]$ e consideriamo, su $(\Omega, \mathcal{F}, \mathbb{P})$ la variabile aleatoria

$$Z = F^{-1}(V).$$

Vediamo allora, dalla (3) che,

$$F_Z(x) = F_V(F(x)) = F(x).$$

In altre parole, partendo da V , basta porre $Z = F^{-1}(V)$ per ottenere una variabile aleatoria, funzione di V , con funzione di distribuzione coincidente con la funzione F assegnata.

Ciò è utile dal punto di vista applicativo, in quanto ci permette (almeno in linea di principio) di "simulare" variabili con distribuzione assegnata a partire da variabili aleatorie con distribuzione uniforme su $[0, 1]$.

Ma ciò è anche importante dal punto di vista teorico: infatti ci mostra che una funzione F , che goda di tutte le proprietà sopra richieste, è certamente la funzione di distribuzione di una qualche variabile aleatoria.

Le stesse argomentazioni si possono estendere anche ai casi in cui si lasci cadere la condizione che F sia continua e strettamente crescente, purchè restino conservate le ipotesi a),b),c). Per trattare tali casi basta generalizzare la nozione di inversa di una funzione di distribuzione ponendo, per $u \in [0, 1]$:

$$F^{-1}(u) = \inf\{x \in \mathbb{R} \mid F(x) \geq u\}.$$

In tal modo è quindi possibile concludere che le condizioni a), b) e c) sono non soltanto necessarie, ma anche sufficienti affinché una funzione F sia una funzione di ripartizione (di una qualche variabile aleatoria).

1.4 I diversi tipi di distribuzioni di probabilità

In questo paragrafo analizziamo la struttura della famiglia \mathbb{F} costituita da tutte le possibili funzioni di ripartizione di distribuzioni di probabilità sulla retta, cioè di tutte le funzioni che soddisfano le condizioni a), b) e c) della Proposizione 1.2.

Innanzitutto notiamo che \mathbb{F} gode della proprietà di convessità

Proposizione 4.1

Siano $F_1, F_2 \in \mathbb{F}$ e α_1, α_2 due costanti positive tali che $\alpha_1 + \alpha_2 = 1$. Allora anche la combinazione convessa

$$F(x) = \alpha_1 F_1(x) + \alpha_2 F_2(x)$$

è un elemento di \mathbb{F} .

Dimostrazione

E' facile verificare che, essendo le condizioni a), b) e c) della Proposizione 1.2 soddisfatte per F_1 ed F_2 , allora valgono anche per F . E quindi la tesi segue dal fatto che tali condizioni sono anche sufficienti per l'appartenenza ad \mathbb{F} .

Come descritto nel primo paragrafo, vi sono due classi notevoli di elementi di \mathbb{F} : le funzioni di ripartizione costanti a tratti, corrispondenti alle distribuzioni discrete, e quelle ovunque continue (cioè tali che $\mathcal{D}_F = \emptyset$).

All'interno di tale ultima classe troviamo quelle che ammettono una *funzione di densità*, secondo la seguente definizione.

Definizione 4.1

La funzione $f(x)$ è una funzione di densità di probabilità per F se risulta

$$F(x) = \int_{-\infty}^x f(\xi)d\xi, \quad x \in \mathbb{R}. \quad (4)$$

Se esiste una densità, la F risulta derivabile quasi ovunque e vale la relazione

$$\frac{d}{dx}F(x) = f(x), \text{ q.o.}$$

Si ha inoltre, per una variabile aleatoria X con funzione di ripartizione F e per valori reali $a < b$,

$$\mathbb{P}\{a \leq X \leq b\} = F(b) - F(a) = \int_a^b f(\xi)d\xi.$$

In "ambiente σ -additivo" la scrittura $\int_{-\infty}^x f(\xi)d\xi$ viene intesa come un integrale di Lebesgue e, per completezza, va corrispondentemente citato che le varie relazioni relative alle funzioni di densità sono valide "quasi ovunque" rispetto alla misura di Lebesgue, cioè a meno di insiemi di misura di Lebesgue nulla.

Nelle usuali applicazioni, però, si può tranquillamente ammettere che la f sia ovunque continua o abbia, al più, un numero finito di punti di discontinuità e sia limitata. In tali casi l'integrazione in (4) può equivalentemente intendersi nel senso di Riemann.

Una F per cui valga la (4) viene anche detta *assolutamente continua*, in quanto la misura di probabilità da essa determinata sulla σ -algebra \mathcal{B} risulta assolutamente continua rispetto alla misura di Lebesgue, cioè tale che $\mathbb{P}_X(B) = 0$ per ogni insieme $B \in \mathcal{B}$ di misura di Lebesgue nulla.

Il termine "densità di probabilità" per la f è motivato dalla relazione

$$f(x) = \lim_{\Delta x \rightarrow 0} \frac{F(x + \Delta x) - F(x)}{\Delta x} = \lim_{\Delta x \rightarrow 0} \frac{\mathbb{P}\{x \leq X \leq x + \Delta x\}}{\Delta x},$$

cioè dal fatto che $f(x)$ si può vedere come il limite di un rapporto fra la probabilità contenuta in un intervallo e la lunghezza dell'intervallo stesso, al tendere a 0 di tale lunghezza.

Proposizione 4.2

Una funzione di densità di probabilità $f(x)$ deve soddisfare le condizioni

- $f(x) \geq 0$
- $\int_{-\infty}^{+\infty} f(\xi)d\xi = 1$ (*condizione di normalizzazione*)

Osservazione 4.1

Sia $g(x)$ una funzione non negativa ed integrabile assegnata. In base alla proprietà di normalizzazione, vale quanto segue: supponiamo di sapere che una data $f(x)$ è una funzione di densità di probabilità e risulta proporzionale alla $g(x)$, cioè esiste un numero positivo K tale che

$$f(x) = Kg(x).$$

(scriveremo brevemente tale relazione con il simbolo $f(x) \propto g(x)$).

Imponendo su $f(x)$ la condizione di normalizzazione possiamo concludere

$$K = \frac{1}{\int_{-\infty}^{+\infty} g(\xi)d\xi}, f(x) = \frac{g(x)}{\int_{-\infty}^{+\infty} g(\xi)d\xi}.$$

Nell'analisi della famiglia \mathbb{F} un punto importante è il seguente: fra le funzioni di distribuzioni ovunque continue, esiste una sottoclasse di funzioni F che non ammettono densità, cioè per le quali non vale la rappresentazione (4). Se una variabile aleatoria X è distribuita secondo una tale F , allora esiste un insieme $B \in \mathcal{B}$ che soddisfa contemporaneamente le seguenti condizioni:

- i) $\mathbb{P}\{X \in B\} = 1$;
- ii) la misura di Lebesgue di B è nulla.

Tali funzioni di ripartizione vengono dette *singolari*. Naturalmente le condizioni i) e ii) valgono anche per le distribuzioni discrete. La differenza fra i due casi sta nel fatto che, al contrario di quelle singolari, le funzioni di ripartizione discrete non sono ovunque continue ed anzi crescono soltanto per salti, in corrispondenza dei loro punti di discontinuità: siano ancora X una variabile aleatoria con funzione di ripartizione F e \mathcal{D}_F l'insieme dei punti di discontinuità di F ; nel caso F è costante a tratti risulta

$$\mathbb{P}(X \in \mathcal{D}_F) = 1,$$

mentre nel caso singolare risulta

$$\mathbb{P}(X \in \mathcal{D}_F) = 0.$$

Indichiamo rispettivamente con $\mathbb{F}_d, \mathbb{F}_{ac}, \mathbb{F}_s$ le tre classi costituite dalle distribuzioni discrete (F costante a tratti), dalle distribuzioni assolutamente continue (F con densità), e da quelle singolari (F continua ma senza densità).

Riguardo la caratterizzazione della forma più generale degli elementi in \mathbb{F} vale il seguente teorema *di decomposizione di Lebesgue*.

Teorema 4.3

Sia $F \in \mathbb{F}$ un'arbitraria funzione di ripartizione. Allora esistono $F_d \in \mathbb{F}_d, F_{ac} \in \mathbb{F}_{ac}, F_s \in \mathbb{F}_s$ e tre costanti non negative $\alpha_d, \alpha_{ac}, \alpha_s$, con $\alpha_d + \alpha_{ac} + \alpha_s = 1$, tali che

$$F(x) = \alpha_d F_d(x) + \alpha_{ac} F_{ac}(x) + \alpha_s F_s(x).$$

1.5 Valori attesi e momenti

Ricordiamo innanzitutto la definizione di valore atteso di una variabile aleatoria X definita su uno spazio di probabilità finito $(\Omega, \mathcal{P}(\Omega), \mathbb{P})$, con $\Omega = \{\omega_1, \dots, \omega_N\}$. In questo caso il valore atteso di X viene definito dalla posizione

$$\mathbb{E}(X) := \sum_{i=1}^N \mathbb{P}\{\omega_i\} \cdot X(\omega_i). \quad (5)$$

Supponiamo che i valori possibili per X siano x_1, x_2, \dots, x_n e che questi vengano assunti con probabilità rispettive p_1, p_2, \dots, p_n , cioè, per $j = 1, 2, \dots, n$,

$$p_j = \mathbb{P}\{\omega_i \in \Omega | X(\omega_i) = x_j\}.$$

In base a tali posizioni si verifica facilmente che valgono le seguenti proprietà:

i)

$$\mathbb{E}(X) = \sum_{j=1}^n p_j x_j; \quad (6)$$

ii)

$$\min_{j=1, \dots, n} x_j \leq \mathbb{E}(X) \leq \max_{j=1, \dots, n} x_j;$$

iii)

$$\mathbb{E}(X) = \bar{x}$$

se X è una variabile aleatoria *degenere* con distribuzione di probabilità concentrata nel valore \bar{x} , cioè se

$$X(\omega_i) = \bar{x}, i = 1, \dots, N.$$

Un'ulteriore proprietà di *linearità* verrà vista nel successivo capitolo.

Notiamo che (5) costituisce la *definizione formale* della nozione di valore atteso, mentre (6) può essere vista piuttosto come una *regola di calcolo* di $\mathbb{E}(X)$. E' importante notare che quest'ultima richiede soltanto la conoscenza della distribuzione di probabilità di X .

Nel caso di uno spazio di probabilità in generale la definizione (5) non potrà, ovviamente, più essere valida. Viene però formulata una definizione di integrale in modo tale che si possa porre

$$\mathbb{E}(X) := \int_{\Omega} X(\omega) \mathbb{P}(d\omega). \quad (7)$$

Il valore di $\int_{\Omega} X(\omega) \mathbb{P}(d\omega)$ mantiene comunque la forma del tipo $\sum_{j=1}^n p_j x_j$ nel caso in cui Ω risulta unione di n eventi disgiunti E_1, \dots, E_n tali che $\mathbb{P}(E_j) = p_j$ e $X(\omega) = x_j, \forall \omega \in E_j, j = 1, 2, \dots, n$.

Nella presente trattazione non ci tratteniamo ulteriormente su tale definizione di integrale e sulle proprietà relative. Possiamo però accennare al fatto che tale nozione generale si può vedere come una naturale estensione di quella relativa all'integrale di Lebesgue sulla retta reale.

In ogni caso, è possibile calcolare $\mathbb{E}(X)$ basandosi soltanto sulla conoscenza della distribuzione di probabilità \mathbb{P}_X , secondo quanto verrà accennato qui di seguito.

Per misure di probabilità ν su $(\mathbb{R}, \mathcal{B})$ e per funzioni misurabili $\phi : \mathbb{R} \rightarrow \mathbb{R}$, si può definire l'integrale

$$\int_{\mathbb{R}} \phi(x) \nu(dx)$$

in modo analogo alla definizione di $\int_{\Omega} X(\omega) \mathbb{P}(d\omega)$. E si dimostra inoltre che vale la relazione

$$\mathbb{E}(X) = \int_{\mathbb{R}} x \cdot \mathbb{P}_X(dx),$$

dove, ricordiamo, \mathbb{P}_X indica la misura di probabilità sulla retta, indotta dalla variabile aleatoria X .

Per quanto ci riguarda al momento, dobbiamo soltanto tener presente a che cosa si riduca l'espressione $\int_{\mathbb{R}} x \cdot \mathbb{P}_X(dx)$ nei casi particolari di distribuzioni di probabilità \mathbb{P}_X discrete e distribuzioni di probabilità assolutamente continue. Nel caso discreto abbiamo ancora

$$\mathbb{E}(X) = \sum_j p_j x_j$$

e, nel caso assolutamente continuo con funzione di densità di probabilità $f(x)$, si ha invece

$$\mathbb{E}(X) = \int_{-\infty}^{+\infty} x f(x) dx. \quad (8)$$

Dobbiamo comunque notare che la scrittura $\int_{\mathbb{R}} x \cdot \mathbb{P}_X(dx)$ non è soltanto un simbolo unificante per significare rispettivamente le quantità $\sum_j p_j x_j$ e $\int_{-\infty}^{+\infty} x f(x) dx$ nei due casi discreto e assolutamente continuo.

$\int_{\mathbb{R}} x \cdot \mathbb{P}_X(dx)$ indica infatti un oggetto ben definito, chiamato *integrale di Stieltjes*, che si riduce a $\int_{-\infty}^{+\infty} xf(x)dx$ oppure a $\sum_j p_j x_j$, nei due casi notevoli citati.

Siccome, come sappiamo, la distribuzione di probabilità \mathbb{P}_X è individuata dalla funzione di ripartizione F_X , si ha che anche l'integrale $\int_{\mathbb{R}} x \cdot \mathbb{P}_X(dx)$ è individuato da F_X ed infatti viene usualmente indicato con il simbolo

$$\int_{-\infty}^{+\infty} x \cdot dF_X(x).$$

Dobbiamo anche notare che l'integrale improprio che figura al secondo membro della (8) potrebbe valere $+\infty$, oppure $-\infty$, oppure non essere definito (quando si presenta la forma indeterminata $+\infty - \infty$). Vedremo nel seguito qualche esempio in proposito.

In base alla definizione (7) si possono dimostrare le proprietà i),ii), iii) elencate sopra e la proprietà di linearità che vedremo nel prossimo capitolo.

Ancora in base alla definizione (7) si può dimostrare anche che si ha una proprietà di linearità *rispetto alle distribuzioni di probabilità*. Cioè, se la funzione di ripartizione F di una variabile aleatoria X è combinazione convessa di altre funzioni di ripartizione, allora $\mathbb{E}(X)$ risulta uguale alla combinazione convessa dei valori attesi rispetto a tali distribuzioni. In altre parole, le condizioni

a)

$$F(x) = \sum_{h=1}^m \alpha_h F_h(x),$$

b)

$$\int_{-\infty}^{+\infty} x \cdot dF_h(x) = \mu_h, h = 1, \dots, m,$$

(con $\alpha_1, \dots, \alpha_m$ e μ_1, \dots, μ_m m -uple di numeri reali assegnati) implicano

$$\int_{-\infty}^{+\infty} x \cdot dF_X(x) = \sum_{h=1}^m \alpha_h \mu_h.$$

Tale ultima formula può risultare particolarmente utile, ricordando l'enunciato del teorema di decomposizione di Lebesgue (Teorema 4.3): il valore atteso di una qualunque variabile aleatoria reale, se esiste ed è un numero finito, si può sempre scrivere come combinazione lineare di valori attesi relativi ai tre casi base: discreto, assolutamente continuo e singolare.

Un'ulteriore proprietà fondamentale dei valori attesi è relativa alle trasformazioni di variabili aleatorie. Sia ancora X una variabile aleatoria con funzione di ripartizione F_X . Sia $\phi : \mathbb{R} \rightarrow \mathbb{R}$ una funzione misurabile e consideriamo la variabile aleatoria $Z = \phi(X)$. La funzione di ripartizione F_Z si ottiene a partire da F_X in base alla trasformazione ϕ , cioè come funzione di ripartizione della

nuova distribuzione di probabilità indotta su \mathcal{B} dalla ϕ . Consideriamo dunque l'integrale

$$\int_{-\infty}^{+\infty} z \cdot dF_Z(z)$$

che definisce il valore atteso $\mathbb{E}(Z)$. Possiamo considerare, d'altra parte, l'integrale

$$\int_{-\infty}^{+\infty} \phi(x) \cdot dF_X(x).$$

Come proprietà degli integrali di Stiltjes si può dimostrare allora che vale l'identità

$$\mathbb{E}(Z) = \int_{-\infty}^{+\infty} z \cdot dF_Z(z) = \int_{-\infty}^{+\infty} \phi(x) \cdot dF_X(x).$$

Tale proprietà consiste semplicemente in una sostituzione di variabile nell'integrale nel membro a destra e mostra dunque l'equivalenza fra due diverse formule per il calcolo di $\mathbb{E}(Z)$.

Per una variabile aleatoria X , con funzione di ripartizione F_X , si definisce *momento di ordine h* il valore atteso $\mathbb{E}(X^h)$, cioè l'integrale di Stiltjes

$$\int_{\mathbb{R}} x^h dF(dx).$$

Vediamo in particolare la *varianza* di X come il *momento centrato di ordine 2*:

$$Var(X) := \mathbb{E} \left[(X - \mu_X)^2 \right], \quad (9)$$

dove si è posto, per brevità, $\mu_X = \mathbb{E}(X)$.

L'integrale che definisce $Var(X)$ potrebbe essere uguale a $+\infty$. Ovviamente, nel caso invece in cui $Var(X) < +\infty$, si ha $Var(X) \geq 0$ e risulta

$$Var(X) = 0$$

se e soltanto se X è una variabile aleatoria degenere.

1.6 Breve nota storica, critica, bibliografica

Nei precedenti paragrafi sono state richiamate succintamente le nozioni fondamentali della teoria assiomatica della probabilità. Tale teoria risale sostanzialmente agli anni '30 del secolo scorso ed è stata principalmente introdotta dal matematico russo A.N. Kolmogorov. Essa è basata sulla Teoria della Misura, a quei tempi oramai consolidata, caratterizzata dall'assunzione che la misura, come funzione di insieme, sia numerabilmente additiva (o σ -additiva). In tale ambito - possiamo riassumere un pò semplicisticamente- era stata sviluppata la nozione di misura di Lebesgue come alternativa alla misura di Peano-Jordan e, corrispondentemente, quella di integrale di Lebesgue come superamento dell'integrale di Riemann.

Lo sviluppo, su tali basi, della teoria della probabilità permise di formalizzarne le nozioni peculiari in chiari termini assiomatici e di tradurre in termini probabilistici vari utili risultati di carattere analitico. Ciò favorì notevolmente l'interesse per la probabilità e la sua diffusione in ambiente matematico. La teoria della misura vi trovò, a sua volta, nuove problematiche e stimoli per successivi sviluppi. Anche a seguito di questo, la teoria della probabilità conobbe successi notevolissimi nei decenni successivi dando luogo, fra l'altro, ad un impetuoso sviluppo dello studio dei processi stocastici.

E' quindi a tale impostazione della probabilità cui si fa usualmente riferimento in ambito matematico e la conoscenza di tali argomenti ha un'importanza fondamentale per quanti vogliono approfondire lo studio della probabilità. Nelle trattazioni più elementari e in quelle di carattere maggiormente applicativo, d'altra parte, le motivazioni e le conseguenze corrispondenti a tale assiomatizzazione non sempre possono essere analizzate e messe in evidenza in forma adeguata. Nei precedenti paragrafi ne è stato quindi dato appena un cenno, per i lettori che non hanno avuto ancora modo di appropriarsi delle nozioni fondamentali della teoria della misura σ -additiva.

Si deve comunque far presente che, nonostante tale assiomatizzazione abbia indubbi meriti e sia quella quasi universalmente seguita in ambito matematico, essa non è l'unica ad essere stata proposta quale base per lo studio della probabilità. Ed essa può presentare sia vantaggi che alcuni punti deboli se confrontata con altre possibili impostazioni. Su tale tematica si è sviluppato nel corso del Novecento un vivace dibattito, che si è anche andato intrecciando con il parallelo dibattito circa i "fondamenti" della probabilità, cioè con le riflessioni e le proposte inerenti il significato euristico, sperimentale, e applicativo da attribuire a tale termine.

In proposito c'è da notare che, mentre l'assunzione dell'assioma di σ -additività può essere oggetto di discussione, la più debole condizione di additività finita risulta invece universalmente accettabile, in quanto del tutto conseguente alle proprietà della probabilità stessa, qualunque significato "pratico" possa venire a questa attribuito. In particolare, sempre a partire dalla fine degli anni venti e inizio degli anni trenta, una teoria della probabilità fondata sull'additività finita e sul rifiuto ad assumerne a priori la proprietà di σ -additività, fu sviluppata in particolare da Bruno de Finetti. Una tale teoria risulta "più debole" (in quanto con un minor numero di assiomi e, quindi, di risultati "automatici"), ma più generale e flessibile. E anche tale teoria ha portato, nei decenni successivi, a interessanti sviluppi. Numerosi matematici, filosofi della scienza e statistici, fra i quali in prima persona lo stesso de Finetti, hanno seguito e alimentato il dibattito circa benefici e punti critici delle varie impostazioni e circa le relazioni con le diverse possibili interpretazioni del termine "probabilità", per tutto il perdurare del secolo.

Per quanto riguarda l'assunzione di σ -additività una sua conseguenza fondamentale è la seguente: una probabilità σ -additiva può in generale essere assegnata soltanto su una qualche σ -algebra $\mathcal{F} \subset \mathcal{P}(\Omega)$, ma non su tutto $\mathcal{P}(\Omega)$. Consideriamo infatti una tale \mathbb{P} definita su un'algebra \mathcal{A} ; sappiamo allora che certamente esiste un modo, ed anzi un unico modo, di estendere \mathbb{P} alla σ -algebra

$\mathcal{F} = \sigma(\mathcal{A})$ in forma tale che venga mantenuta la σ -additività. Un punto cruciale risiede nella possibilità di costruire esempi nei quali:

- \mathcal{F} è strettamente inclusa in $\mathcal{P}(\Omega)$
- non è possibile, mantenendone la σ -additività, estendere la \mathbb{P} a tutti gli insiemi $E \in \mathcal{P}(\Omega)/\mathcal{F}$.

Problematiche di tale tipo, in particolare, emergono in modo alquanto naturale nello studio dei processi stocastici a tempo continuo.

Tale caratteristica, che è alla base delle definizioni di spazio di probabilità e di variabile aleatoria richiamate nel primo paragrafo di questo capitolo, viene dunque a determinare una notevole differenza fra le diverse impostazioni della teoria delle probabilità. Ma sussistono diversità ancora più fondamentali, ad esempio con l'impostazione sviluppata da de Finetti. In tale impostazione, infatti, non si considera ovvio il fatto che si debba assumere come primitiva la nozione di spazio di probabilità. Cioè non viene individuato a priori - ed una volta per tutte - uno spazio di eventi elementari e non si assume che le variabili aleatorie di interesse siano viste come funzioni definite su tale spazio.

Altre differenze fondamentali si incontrano nell'assiomatica della probabilità quantistica.

Cogliamo anche questa occasione per accennare al fatto che, in diverse applicazioni - in particolare in Economia- si è recentemente sviluppato lo studio di misure (cioè di funzioni di insieme) che, oltre a non essere σ -additive, non sono neanche semplicemente additive e mantengono però la proprietà di essere *monotone*.

La letteratura su tutti questi argomenti è vastissima sia a livello di saggi di ricerca che di libri di testo universitari, più o meno recenti. Fra i numerosissimi suggerimenti possibili, indicheremo qui di seguito soltanto alcuni riferimenti bibliografici di base o divenuti oramai classici, la cui lettura può risultare utile per avviare uno studio approfondito ed autonomo della probabilità e delle sue applicazioni.

Alcuni Riferimenti bibliografici

- P. Billingsley. *Probability and Measure*. John Wiley & Sons, varie edizioni
- B. de Finetti. *Teoria delle Probabilità*. Einaudi, 1970
- D. Denneberg. *Non-additive measure and integral*. Kluwer Academic Publishers Group, 1994.
- J. L. Doob. *Stochastic processes*. John Wiley & Sons, varie edizioni
- W. Feller. *An introduction to probability theory and its applications*. Vol. I e II. John Wiley & Sons., varie edizioni
- P. Halmos. *Measure Theory*. Van Nostrand, varie edizioni
- A. N. Kolmogorov. *Grundbegriffe der Wahrscheinlichkeitsrechnung*. Springer-Verlag, 1933
- P. A. Meyer. *Quantum probability for probabilists*. Springer-Verlag, 1995.
- I. Molchanov. *Theory of random sets*. Probability and its Applications (New York). Springer-Verlag, 2005.

2 Distribuzioni di probabilità di vettori aleatori

In questo capitolo consideriamo il caso di n ($n \geq 2$) variabili aleatorie definite su uno stesso spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$. In particolare, nel primo paragrafo, definiamo la *distribuzione di probabilità congiunta* di una coppia di variabili aleatorie $X, Y : \Omega \rightarrow \mathbb{R}$ ed introduciamo la nozione di *funzione di ripartizione congiunta*.

Qui notiamo che, essendo Ω il comune dominio delle applicazioni X, Y , è possibile definire delle operazioni fra X, Y a partire dalle corrispondenti operazioni fra due numeri reali. Ad esempio, la variabile aleatoria *somma* fra X ed Y coincide con l'applicazione $(X + Y) : \Omega \rightarrow \mathbb{R}$ definita da

$$(X + Y)(\omega) = X(\omega) + Y(\omega).$$

Pure se è naturale usare lo stesso simbolo $+$ sia a sinistra che a destra del segno di uguale, notiamo che tale simbolo vi assume due significati diversi: a sinistra stiamo definendo la somma fra le due variabili aleatorie e a destra stiamo calcolando la usuale somma fra i due numeri reali $X(\omega)$ ed $Y(\omega)$. Notiamo anche che la misurabilità di $(X + Y)$ rispetto ad \mathcal{F} è subito garantita dall'essere X ed Y delle variabili aleatorie. Analogamente definiremo il *prodotto* fra X ed Y come l'applicazione definita da

$$(X \cdot Y)(\omega) = X(\omega) \cdot Y(\omega),$$

e così via anche a partire dalle altre usuali operazioni fra due numeri reali.

Quanto appena detto costituisce dunque una fondamentale implicazione dell'aver definito le variabili aleatorie in questione come applicazioni su uno stesso spazio di probabilità. Potremmo anche dire che la nozione di spazio di probabilità mostra la sua effettiva rilevanza proprio quando dobbiamo considerare più di una variabile aleatoria, mentre tale nozione potrebbe apparire un pò artificiale fintanto che ci limitassimo a considerare una variabile aleatoria singola.

Una diversa implicazione risiede nella possibilità di definire, come vedremo nel seguito attraverso la nozione di *distribuzione condizionata*, che cosa si possa intendere per relazione probabilistica fra variabile aleatorie diverse.

2.1 Funzioni di ripartizione in due dimensioni

Siano date le variabili aleatorie $X, Y : \Omega \rightarrow \mathbb{R}$. Analogamente a quanto visto nel capitolo precedente per il caso di una singola variabile aleatoria, per *distribuzione di probabilità congiunta* di (X, Y) possiamo intendere la misura di probabilità su $(\mathbb{R}^2, \mathcal{B}^{(2)})$ definita dalla relazione

$$\mathbb{P}_{X,Y}(B) := \mathbb{P}\{(X, Y) \in B\} = \mathbb{P}\{\omega \in \Omega | (X(\omega), Y(\omega)) \in B\},$$

essendo $B \subset \mathbb{R}^2$ un generico elemento appartenente alla famiglia $\mathcal{B}^{(2)}$ degli *insiemi boreliani* del piano.

In modo analogo a quanto fatto per \mathcal{B} nel capitolo precedente, la famiglia $\mathcal{B}^{(2)}$ può venir definita come la σ -algebra generata dalla famiglia $I^{(2)}$ degli insiemi della forma (12). A questo proposito dobbiamo osservare che $I^{(2)}$ costituisce un π -sistema.

Definizione 1.1

La *funzione di ripartizione congiunta* di (X, Y) viene definita dalla relazione

$$F_{X,Y}(x, y) := \mathbb{P}\{(X \leq x) \cap (Y \leq y)\}, x \in \mathbb{R}, y \in \mathbb{R}. \quad (10)$$

Cioè

$$F_{X,Y}(x, y) = \mathbb{P}_{X,Y}(B_{x,y}), \quad (11)$$

avendo posto, per $(x, y) \in \mathbb{R}^2$,

$$B_{x,y} := \{(\xi, \eta) \in \mathbb{R}^2 \mid \xi \leq x, \eta \leq y\}. \quad (12)$$

Ovviamente la conoscenza della distribuzione di probabilità congiunta $\mathbb{P}_{X,Y}$ determina in particolare $F_{X,Y}$, come mostrato dalla (11). La $\mathbb{P}_{X,Y}$ è, d'altra parte, individuata univocamente dalla $F_{X,Y}$. Analogamente a quanto è stato accenato per il caso unidimensionale nel Paragrafo 1.2 del precedente capitolo, la validità di quest'ultima affermazione discende direttamente dal fatto che $\mathcal{B}^{(2)} = \sigma(\mathcal{I}^{(2)})$ e che $I^{(2)}$ costituisce appunto un π -sistema.

Osservazione 1.1

In termini della $F_{X,Y}$ risulta immediato scrivere la probabilità che la coppia (X, Y) cada dentro una *semistriscia* parallela ad uno degli assi coordinati o che (X, Y) cada dentro un *intervallo coordinato* del piano. Sia infatti, per $x \in \mathbb{R}$, $y \in \mathbb{R}$, $x' < x'' \in \mathbb{R}$ e $y' < y'' \in \mathbb{R}$,

$$\begin{aligned} A &= \{(\xi, \eta) \in \mathbb{R}^2 \mid x' < \xi \leq x'', \eta \leq y\} \\ B &= \{(\xi, \eta) \in \mathbb{R}^2 \mid \xi \leq x, y' < \eta \leq y''\} \\ D &= \{(\xi, \eta) \in \mathbb{R}^2 \mid x' < \xi \leq x'', y' < \eta \leq y''\}. \end{aligned} \quad (13)$$

In base alla definizione (10), possiamo scrivere

$$\mathbb{P}\{(X, Y) \in A\} = F_{X,Y}(x'', y) - F_{X,Y}(x', y)$$

$$\mathbb{P}\{(X, Y) \in B\} = F_{X,Y}(x, y'') - F_{X,Y}(x, y')$$

$$\mathbb{P}\{(X, Y) \in D\} = F_{X,Y}(x'', y'') - F_{X,Y}(x', y'') - F_{X,Y}(x'', y') + F_{X,Y}(x', y'). \quad (14)$$

Si vede facilmente che una funzione $F(x, y) : \mathbb{R}^2 \rightarrow \mathbb{R}$ può essere una funzione di ripartizione congiunta di una coppia di variabili aleatorie X, Y solo se soddisfa il seguente sistema di proprietà:

(i)

$$0 \leq F(x, y) \leq 1, \quad \forall x \in \mathbb{R}, \forall y \in \mathbb{R};$$

(ii)

$$\lim_{x \rightarrow -\infty} F(x, y) = 0, \forall y \in \mathbb{R}; \quad \lim_{y \rightarrow -\infty} F(x, y) = 0, \forall x \in \mathbb{R};$$

$$\lim_{x \rightarrow +\infty, y \rightarrow +\infty} F(x, y) = 1; \quad (15)$$

(iii) $F(x, y)$ è non decrescente in ciascuna variabile: per $x' < x'', y' < y''$, risulta

$$F(x, y) \leq F(x, y)$$

$$F(x, y) \leq F(x, y)$$

(iv) $F(x, y)$ è *non-decrescente in due variabili*: per $x' < x'', y' < y''$, risulta

$$F(x'', y'') - F(x', y'') - F(x', y') + F(x', y) \geq 0. \quad (16)$$

Osservazione 1.2

La condizione (iv) deriva dal fatto che la sommatoria a sinistra del segno \geq esprime una probabilità, come indicato nell'Osservazione 1.1.

Notiamo inoltre che le condizioni (i) - (iii) non sono sufficienti a garantire anche la (iv). Infatti è facile trovare funzioni che soddisfano (i) - (iii) e non (iv). Un semplice esempio si può ottenere considerando il rettangolo coordinato individuato dai valori $x' = 0, x'' = 2, y' = 0, y'' = 2$ e la funzione $F(x, y) = \mathbf{1}_C(x, y)$, dove

$$C = \{(\xi, \eta) \in \mathbb{R}^2 \mid \xi + \eta \geq 1\}.$$

2.2 Distribuzioni marginali e indipendenza stocastica

Consideriamo ancora la distribuzione di probabilità congiunta e la funzione di ripartizione congiunta $F_{X,Y}$ per la coppia di variabili aleatorie (X, Y) . Fissiamo ora l'attenzione sulle due distribuzioni di probabilità unidimensionali \mathbb{P}_X e \mathbb{P}_Y di X ed Y , considerate una separatamente dall'altra. A questo proposito, dati comunque I, J sottoinsiemi boreliani della retta, consideriamo gli insiemi

$$I \times \mathbb{R} := \{(x, y) \in \mathbb{R}^2 \mid x \in I\}, \quad \mathbb{R} \times J := \{(x, y) \in \mathbb{R}^2 \mid y \in J\}.$$

Notiamo in particolare che, nel caso $I = (-\infty, x]$ oppure $J = (-\infty, y]$, ricordando la notazione introdotta in (12) possiamo scrivere

$$I \times \mathbb{R} = \lim_{y \rightarrow +\infty} B_{x,y}, \quad \mathbb{R} \times J = \lim_{x \rightarrow +\infty} B_{x,y}.$$

Possiamo così vedere \mathbb{P}_X e \mathbb{P}_Y , rispettivamente, come le misure di probabilità sulla retta tali che

$$\mathbb{P}_X(I) := \mathbb{P}\{X \in I\} = \mathbb{P}_{X,Y}(I \times \mathbb{R})$$

$$\mathbb{P}_Y(J) := \mathbb{P}\{Y \in J\} = \mathbb{P}_{X,Y}(\mathbb{R} \times J).$$

In tale ottica, per le relative funzioni di ripartizione F_X, F_Y possiamo scrivere

$$F_X(x) = \mathbb{P}\{X \leq x\} = \lim_{y \rightarrow +\infty} F_{X,Y}(x, y)$$

$$F_Y(y) = \mathbb{P}\{Y \leq y\} = \lim_{x \rightarrow +\infty} F_{X,Y}(x, y).$$

\mathbb{P}_X e \mathbb{P}_Y vengono anche dette *distribuzioni marginali* della $\mathbb{P}_{X,Y}$. F_X è detta inoltre funzione di ripartizione marginale di X ; analogamente per quanto riguarda F_Y .

Veniamo ora alla nozione di *indipendenza stocastica* fra X ed Y . Nel caso di variabili aleatorie discrete, come sappiamo, tale definizione si riporta alla condizione di indipendenza stocastica per le coppie di eventi del tipo $(X = x_i), (Y = y_j)$.

Nel caso generale quest'ultima condizione non è necessariamente significativa e la nozione di indipendenza viene piuttosto espressa dall'indipendenza stocastica per tutte le coppie di eventi del tipo $(X \in I), (Y \in J)$, essendo I, J generici sottoinsiemi boreliani della retta.

In particolare, I, J potrebbero essere intervalli oppure unioni di intervalli disgiunti. Notiamo quindi che l'indipendenza stocastica fra X ed Y implica l'indipendenza fra le coppie di eventi della forma $(X \leq x), (Y \leq y)$, cioè

$$F_{X,Y}(x, y) = F_X(x) \times F_Y(y). \quad (17)$$

Si può però dimostrare che in effetti tale ultima condizione non soltanto è necessaria ma anche sufficiente per l'indipendenza fra le coppie di eventi del tipo $(X \in I), (Y \in J)$, per ogni coppia di insiemi di Borel I, J . A tale fine si parte dall'osservazione che la famiglia dei sottoinsiemi $B_{x,y}$ del piano della forma (12) costituisce un π -sistema e si può poi pervenire alla conclusione desiderata tenendo conto del Teorema 2.1 del precedente capitolo, ragionando in modo analogo a quanto fatto nella dimostrazione del Teorema 2.2.

Tutto ciò porta ad assumere la relazione (17) quale definizione di indipendenza stocastica fra X ed Y . Per indicare tale condizione useremo anche il simbolo $X \underline{II} Y$.

2.3 Caso assolutamente continuo

Diciamo che la distribuzione congiunta della coppia (X, Y) ammette *densità* (o che $F_{X,Y}$ ammette *densità*) quando esiste una funzione $f_{X,Y}(x, y)$ tale che, $\forall x \in \mathbb{R}, \forall y \in \mathbb{R}$,

$$F_{X,Y}(x, y) = \int_{-\infty}^x \int_{-\infty}^y f_{X,Y}(\xi, \eta) d\xi d\eta. \quad (18)$$

$f_{X,Y}$ viene detta *funzione di densità (di probabilità) congiunta* di (X, Y) . Useremo semplicemente il simbolo f , quando non è necessario specificare gli indici X, Y .

Come immediata conseguenza della condizione (16), risulta

$$f_{X,Y}(x, y) \geq 0 \quad (19)$$

e, vista la (15), $f_{X,Y}$ soddisfa la condizione di normalizzazione

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f_{X,Y}(x, y) dx dy = 1. \quad (20)$$

Se vale la (18) allora, quasi ovunque, esiste la derivata seconda mista $\frac{\partial^2}{\partial x \partial y} F_{X,Y}$ e possiamo porre

$$f_{X,Y}(x, y) = \frac{\partial^2}{\partial x \partial y} F_{X,Y}(x, y). \quad (21)$$

A partire dalla (18) e dalla (14) è immediato verificare che per qualunque rettangolo coordinato, cioè della forma (13), si può scrivere

$$\mathbb{P}\{(X, Y) \in D\} = \int \int_D f_{X,Y}(\xi, \eta) d\xi d\eta = \int_{x'}^{x''} \int_{y'}^{y''} f_{X,Y}(\xi, \eta) d\xi d\eta. \quad (22)$$

A partire dalla (22), a sua volta, si può dimostrare che per un qualunque sottoinsieme di Borel $B \subset \mathbb{R}^2$, $B \in \mathcal{B}^{(2)}$, vale la relazione

$$\mathbb{P}\{(X, Y) \in B\} = \int \int_B f_{X,Y}(\xi, \eta) d\xi d\eta. \quad (23)$$

La validità delle relazioni (23) al variare di $B \in \mathcal{B}^{(2)}$ costituisce quindi una condizione anche necessaria affinché la distribuzione congiunta di (X, Y) ammetta funzione di densità congiunta e questa sia uguale a $f_{X,Y}$.

Ciò mostra la semplificazione insita nella condizione di esistenza della densità congiunta: *la probabilità, per la coppia (X, Y) , di cadere in una regione regolare del piano diventa un integrale doppio, esteso a quella regione, della funzione di densità congiunta.*

Un'ulteriore fondamentale semplificazione verrà vista nel paragrafo successivo, trattando di distribuzioni condizionate.

Osserviamo anche che, come conseguenza della (23), necessariamente risulta

$$\mathbb{P}\{(X, Y) \in B\} = 0$$

per qualunque sottoinsieme $B \in \mathcal{B}^{(2)}$ che abbia misura di Lebesgue nulla. Cioè la misura di probabilità $\mathbb{P}_{X,Y}$ è *assolutamente continua* rispetto alla misura di Lebesgue su $\mathcal{B}^{(2)}$. Si ha d'altra parte (*Teorema di Radon-Nykodim*) che la condizione di assoluta continuità è non soltanto necessaria ma anche sufficiente per l'esistenza di una funzione di densità di probabilità $f_{X,Y}$ (*misurabile*). Per tale motivo diciamo comunemente che $\mathbb{P}_{X,Y}$ è assolutamente continua, per indicare il fatto che essa ammette una funzione di densità di probabilità congiunta.

Vediamo ora quale conseguenza abbia tale condizione di assoluta continuità sul calcolo delle distribuzioni marginali e sulla condizione di indipendenza stocastica. Possiamo scrivere

$$F_X(x) = \mathbb{P}\{X \leq x\} = \int_{-\infty}^x \int_{-\infty}^{+\infty} f_{X,Y}(\xi, \eta) d\xi d\eta,$$

$$F_Y(y) = \mathbb{P}\{Y \leq y\} = \int_{-\infty}^{+\infty} \int_{-\infty}^y f_{X,Y}(\xi, \eta) d\xi d\eta.$$

Da tali relazioni segue immediatamente che le due distribuzioni marginali (quali distribuzioni unidimensionali) sono assolutamente continue e le relative funzioni di densità di probabilità sono date da

$$f_X(x) = \frac{d}{dx} \left[\int_{-\infty}^x \int_{-\infty}^{+\infty} f_{X,Y}(\xi, \eta) d\xi d\eta \right] = \int_{-\infty}^{+\infty} f_{X,Y}(x, \eta) d\eta,$$

$$f_Y(y) = \frac{d}{dy} \left[\int_{-\infty}^{+\infty} \int_{-\infty}^y f_{X,Y}(\xi, \eta) d\xi d\eta \right] = \int_{-\infty}^{+\infty} f_{X,Y}(\xi, y) d\xi. \quad (24)$$

Qui, e nel seguito, la possibilità di scambiare l'ordine di integrazione negli integrali doppi e di portare il segno di differenziazione sotto il segno di integrale, è garantita dalla validità delle condizioni (19) e (20).

Per quanto riguarda la condizione di indipendenza stocastica $X \perp\!\!\!\perp Y$ notiamo, tenendo conto della (21), che nel caso assolutamente continuo essa si traduce nella relazione

$$f_{X,Y}(x, y) = f_X(x) \cdot f_Y(y). \quad (25)$$

Viceversa la condizione che \mathbb{P}_X e \mathbb{P}_Y ammettano densità, $f_X(x)$ e $f_Y(y)$ rispettivamente, e che X, Y siano indipendenti garantisce che anche $\mathbb{P}_{X,Y}$ sia assolutamente continua e che valga la relazione (25). Ciò si può verificare immediatamente notando che possiamo scrivere

$$F_{X,Y}(x, y) = F_X(x) \cdot F_Y(y) =$$

$$= \left[\int_{-\infty}^x f_X(\xi) d\xi \right] \cdot \left[\int_{-\infty}^y f_Y(\eta) d\eta \right] = \int_{-\infty}^x \int_{-\infty}^y f_X(\xi) \cdot f_Y(\eta) d\xi d\eta.$$

Dobbiamo porre attenzione alla differenza che in generale sussiste fra la condizione di assoluta continuità per una distribuzione bidimensionale $\mathbb{P}_{X,Y}$ e la condizione di assoluta continuità per (entrambe) le distribuzioni unidimensionali $\mathbb{P}_X, \mathbb{P}_Y$, ottenute quali marginali di $\mathbb{P}_{X,Y}$: la prima condizione implica la seconda, ma non è vero il viceversa. Possiamo infatti costruire esempi di distribuzioni congiunte non assolutamente continue le cui marginali ammettano la densità.

Notiamo inoltre che, come è facile verificare (provare per esercizio), la condizione (25) è non soltanto sufficiente ma anche necessaria perchè valga una fattorizzazione della forma

$$f_{X,Y}(x, y) = \alpha(x) \cdot \beta(y), \quad (26)$$

essendo α e β due funzioni non negative, non necessariamente normalizzate (si tratta cioè di verificare che (26) implica (25)).

2.4 Densità condizionate e Formule di Bayes

Consideriamo di nuovo una coppia di variabili aleatorie X, Y con funzione di ripartizione congiunta F ; ci poniamo il problema di definire ed esprimere la *distribuzione di probabilità condizionata* di una delle due variabili, ad esempio X , dato un valore osservato per l'altra.

Tale problema è risolto immediatamente nel caso in cui X, Y siano variabili aleatorie discrete e ci si possa quindi limitare a considerare le probabilità condizionate $P\{X = x_i | Y = y_j\}$, per quei valori y_j tali che $P\{Y = y_j\} > 0$.

Viene richiesta però una trattazione più sofisticata per affrontare il caso generale in cui occorrerebbe dare un significato, ad esempio, ad espressioni del tipo

$$\mathbb{P}\{X \leq x | Y = y\},$$

anche quando y è un valore tale che $\mathbb{P}(Y = y) = 0$.

In quanto segue mostreremo che il problema ammette comunque una soluzione semplice non solo nel caso discreto, ma anche nel caso congiuntamente assolutamente continuo.

Iniziamo a considerare la distribuzione condizionata di X dato un evento del tipo $(y \leq Y \leq y + \Delta y)$.

Per un fissato $\Delta y > 0$, ci limiteremo a considerare quei valori y tali che

$$\mathbb{P}\{y \leq Y \leq y + \Delta y\} = F_Y(y + \Delta y) - F_Y(y) > 0.$$

Si ha

$$\mathbb{P}\{X \leq x | y \leq Y \leq y + \Delta y\} = \frac{\mathbb{P}\{(X \leq x) \cap (y \leq Y \leq y + \Delta y)\}}{\mathbb{P}\{y \leq Y \leq y + \Delta y\}} = \frac{F(x, y + \Delta y) - F(x, y)}{F_Y(y + \Delta y) - F_Y(y)}.$$

Il rapporto $\frac{F(x, y + \Delta y) - F(x, y)}{F_Y(y + \Delta y) - F_Y(y)}$ definisce una funzione di distribuzione unidimensionale nella variabile x , che indicheremo con il simbolo

$$F_X(x | y \leq Y \leq y + \Delta y).$$

Nel caso in cui F ammette una funzione di densità congiunta f , potremo scrivere

$$F_X(x | y \leq Y \leq y + \Delta y) = \frac{\int_{-\infty}^x \int_y^{y + \Delta y} f(\xi, \eta) d\xi d\eta}{\int_y^{y + \Delta y} f_Y(\eta) d\eta},$$

da cui anche

$$F_X(x | y \leq Y \leq y + \Delta y) = \int_{-\infty}^x f_X(\xi | y \leq Y \leq y + \Delta y) d\xi,$$

ponendo

$$f_X(\xi | y \leq Y \leq y + \Delta y) := \frac{\int_y^{y + \Delta y} f(\xi, \eta) d\eta}{\int_y^{y + \Delta y} f_Y(\eta) d\eta}.$$

Possiamo a questo punto *definire*

$$f_X(x | Y = y) := \lim_{\Delta y \rightarrow 0} f_X(\xi | y \leq Y \leq y + \Delta y)$$

(quando il limite esiste) quale *funzione di densità condizionata* di X , dato ($Y = y$).

Nel caso in cui $f_Y(x) > 0$ si verifica immediatamente, applicando il teorema della media sia al numeratore che al denominatore, che il limite di fatto esiste e che risulta

$$f_X(x | Y = y) = \frac{f(x, y)}{f_Y(y)}. \quad (27)$$

Analogamente, per x fissato e tale che $f_X(x) > 0$, definiremo *funzione di densità condizionata* di Y , dato ($X = x$) la funzione della variabile y

$$f_Y(y | X = x) = \frac{f(x, y)}{f_X(x)} \quad (28)$$

Osservazione 4.1

Confrontando le definizioni in (27) e (28), otteniamo immediatamente anche

$$f_X(x|Y=y) = f_X(x) \frac{f_Y(y|X=x)}{f_Y(y)}$$

Quest'ultima formula prende il nome di *Formula di Bayes* per il caso assolutamente continuo.

Osservazione 4.2

Notiamo che nella precedente formula di Bayes il fattore moltiplicativo $\frac{1}{f_Y(y)}$ è costante rispetto alla variabile x . Dunque $\frac{1}{f_Y(y)}$ ha il ruolo di *costante di normalizzazione* ed in effetti risulta

$$f_Y(y) = \int_{-\infty}^{+\infty} f_X(x) f_Y(y|X=x) dx.$$

Si veda anche, a questo proposito, l'Osservazione (4.2) del precedente capitolo. Possiamo dunque, equivalentemente, scrivere anche

$$f_X(x|Y=y) \propto f_X(x) f_Y(y|X=x)$$

oppure

$$f_X(x|Y=y) \propto f(x, y).$$

Osservazione 4.3

Nel caso di indipendenza stocastica $X \underline{II} Y$, definito dalla condizione (25), risulta ovviamente

$$f_X(x|Y=y) = f_X(x), f_Y(y|X=x) = f_Y(y),$$

$\forall x, y$ tali che $f_X(x) > 0, f_Y(y) > 0$.

Osservazione 4.4

Abbiamo dunque visto che il problema di definire la distribuzione condizionata di una variabile aleatoria, dato il valore assunto da un'altra variabile, ammette una soluzione diretta sia nel caso discreto che nel caso di distribuzione congiunta assolutamente continua.

A prima vista questi due casi possono apparire molto diversi fra loro. In verità, relativamente a tale problema, essi possono essere visti come casi particolari della seguente situazione più generale: la distribuzione congiunta risulta assolutamente continua rispetto alla misura prodotto delle distribuzioni marginali. Ciò significa che, dato un qualunque insieme $B \in \mathcal{B}^{(2)}$ e posto

$$B_X := \{x \in R \mid (x, \eta) \in B, \text{ per qualche } \eta \in \mathbb{R}\}$$

$$B_Y := \{y \in R \mid (\xi, y) \in B, \text{ per qualche } \xi \in \mathbb{R}\},$$

risulta

$$\mathbb{P}((X, Y) \in B) = 0$$

se

$$\mathbb{P}(X \in B_X) \cdot \mathbb{P}(Y \in B_Y) = 0.$$

2.5 Trasformazioni di densità congiunte

Consideriamo ancora due variabili aleatorie che, per ragioni di convenienza nelle notazioni, indichiamo ora con i simboli X_1, X_2 . Concentriamo l'attenzione sul caso congiuntamente assolutamente continuo, in cui la distribuzione congiunta è individuata da una funzione di densità congiunta $f_{X_1, X_2}(x_1, x_2)$.

Consideriamo una trasformazione biunivoca $\mathcal{T} : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ e definiamo una nuova coppia Y_1, Y_2 tramite la posizione

$$(Y_1, Y_2) := \mathcal{T}(X_1, X_2),$$

oppure, indicando con $\mathcal{T}_1, \mathcal{T}_2$ le coordinate della trasformazione \mathcal{T}

$$\begin{aligned} Y_1 &:= \mathcal{T}_1(X, Y) \\ Y_2 &:= \mathcal{T}_2(X, Y) \end{aligned} .$$

Indichiamo con \mathcal{V} la trasformata inversa della \mathcal{T} , cosicchè possiamo anche scrivere

$$\begin{aligned} X_1 &:= \mathcal{V}_1(Y_1, Y_2) \\ X_2 &:= \mathcal{V}_2(Y_1, Y_2) \end{aligned} .$$

Anche Y_1, Y_2 possono essere viste come variabili aleatorie sullo stesso spazio di probabilità su cui sono definite X_1, X_2 e, per ogni sottoinsieme $B \in \mathcal{B}^{(2)}$, vale la relazione

$$\mathbb{P}\{(Y_1, Y_2) \in B\} = \mathbb{P}\{(X_1, X_2) \in \mathcal{V}(B)\}. \quad (29)$$

La relazione (29) può essere utilizzata per analizzare la distribuzione di probabilità congiunta della coppia (Y_1, Y_2) a partire dalla conoscenza di quella di (X_1, X_2) .

Proposizione 5.1

Supponiamo che, per $i, j = 1, 2$, esistano continue le derivate parziali $\frac{\partial}{\partial x_j} \mathcal{V}_i(y_1, y_2)$ ed indichiamo con $J_{\mathcal{V}}(y_1, y_2)$ il determinante della matrice jacobiana della trasformazione \mathcal{V} . Allora anche la distribuzione di probabilità congiunta di (Y_1, Y_2) risulta assolutamente continua e la relativa funzione di densità congiunta è data dalla formula

$$f_{Y_1, Y_2}(y_1, y_2) = f_{X_1, X_2}(\mathcal{V}_1(y_1, y_2), \mathcal{V}_2(y_1, y_2)) \cdot |J_{\mathcal{V}}(y_1, y_2)|.$$

Dimostrazione

Ricordando quanto già affermato nel precedente Paragrafo 2.4 si tratta di verificare che, per ogni $B \in \mathcal{B}^{(2)}$, si abbia

$$\mathbb{P}\{(Y_1, Y_2) \in B\} = \int \int_B f_{X_1, X_2}(\mathcal{V}_1(y_1, y_2), \mathcal{V}_2(y_1, y_2)) \cdot |J_{\mathcal{V}}(y_1, y_2)| dy_1 dy_2. \quad (30)$$

Tenendo conto della (29) e del fatto che la distribuzione di probabilità congiunta di (X_1, X_2) è assolutamente continua con funzione di densità congiunta f_{X_1, X_2} , possiamo scrivere

$$\mathbb{P}\{(Y_1, Y_2) \in B\} = \mathbb{P}\{(X_1, X_2) \in \mathcal{V}(B)\} = \int_{\mathcal{V}(B)} f_{X_1, X_2}(x_1, x_2) dx_1 dx_2.$$

La relazione (30) si ottiene allora semplicemente tramite il cambiamento di variabili

$$\begin{aligned}x_1 &= \mathcal{V}_1(y_1, y_2) \\x_2 &= \mathcal{V}_2(y_1, y_2)\end{aligned}$$

e ricordando che $|J_{\mathcal{V}}(y_1, y_2)|$ indica il modulo del determinante della matrice jacobiana di tale trasformazione.

La Proposizione 5.1 costituisce il caso particolare (relativo a $n = 2$) di un risultato che verrà formulato nel successivo Paragrafo 2.7, e di cui vedremo nel seguito diverse applicazioni.

2.6 Somme di variabili aleatorie

Date due variabile aleatorie X_1, X_2 definite su uno stesso spazio di probabilità, consideriamo ora la distribuzione di probabilità della variabile aleatoria somma $S = X_1 + X_2$. Analizziamo in particolare i due casi specifici in cui la distribuzione congiunta di X_1, X_2 sia discreta o assolutamente continua.

Per quanto riguarda il primo caso, assumiamo che siano $\{x_1^{(1)}, x_2^{(1)}, \dots\}$ e $\{x_1^{(2)}, x_2^{(2)}, \dots\}$ gli insiemi dei valori possibili per X_1 e X_2 rispettivamente. Possiamo allora scrivere

$$\begin{aligned}\mathbb{P}\{S = s\} &= \sum_i \mathbb{P}\{(X_1 = x_i^{(1)}) \cap (S = s)\} = \\ &= \sum_i \mathbb{P}\{(X_1 = x_i^{(1)}) \cap (X_2 = s - x_i^{(1)})\}.\end{aligned}$$

Nel caso particolare in cui X, Y sono stocasticamente indipendenti, tale formula diventa quindi

$$\mathbb{P}\{S = s\} = \sum_i \mathbb{P}(X_1 = x_i^{(1)}) \mathbb{P}(X_2 = s - x_i^{(1)}).$$

Vedremo subito che formule analoghe valgono anche nel caso in cui la distribuzione congiunta di (X, Y) sia assolutamente continua.

In un tal caso, indicando con $f(x_1, x_2)$ la funzione di densità congiunta, la funzione ripartizione di S ,

$$F_S(s) = \mathbb{P}(X_1 + X_2 \leq s),$$

potrà essere scritta nella forma

$$F_S(s) = \int_{-\infty}^{+\infty} \left[\int_{-\infty}^{s-x_1} f_{X_1, X_2}(x_1, x_2) dx_2 \right] dx_1.$$

Ciò mostra che F_S è assolutamente continua e la relativa funzione di densità è data da

$$f_S(s) = \int_{-\infty}^{+\infty} f_{X_1, X_2}(x_1, s - x_1) dx_1. \quad (31)$$

Nel caso in cui X_1, X_2 sono stocasticamente indipendenti, tale formula a sua volta diventa

$$f_S(s) = \int_{-\infty}^{+\infty} f_{X_1}(x_1) f_{X_2}(s - x_1) dx_1. \quad (32)$$

Notiamo che nel caso in cui valga

$$f_{X_1}(x_1) = 0, f_{X_2}(x_2) = 0$$

per $x_1 < 0, x_2 < 0$, otteniamo poi

$$f_S(s) = \int_0^s f_{X_1}(x_1) f_{X_2}(s - x_1) dx_1. \quad (33)$$

La formula (32) definisce un'operazione binaria sullo spazio delle funzioni di densità di probabilità. Tale operazione è detta *prodotto di convoluzione* o, direttamente, *convoluzione*. Si può facilmente verificare (come d'altra parte risulta chiaro dal suo significato probabilistico) che tale operazione è commutativa e associativa.

Notiamo che alla formula (31) si può anche giungere attraverso un percorso diverso (pur se sostanzialmente equivalente al precedente). Mantenendo le notazioni del precedente paragrafo, consideriamo la trasformazione biunivoca $\mathcal{T} : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ definita dalle relazioni

$$\begin{aligned} Y_1 &= X_1 \\ Y_2 &= X_1 + X_2 \end{aligned} ,$$

la cui inversa è la trasformazione \mathcal{V} data da

$$\begin{aligned} X_1 &= Y_1 \\ X_2 &= Y_2 - Y_1 \end{aligned} .$$

Ricordiamo allora la precedente Proposizione 5.1 ed osserviamo che il modulo del determinante della matrice jacobiana della \mathcal{V} è costante e uguale a 1. Otteniamo dunque che la funzione di densità congiunta della coppia (Y_1, Y_2) è data da

$$f_{Y_1, Y_2}(y_1, y_2) = f_{X_1, X_2}(y_1, y_2 - y_1).$$

Dalla formula (24), vediamo che la densità marginale di $S = Y_2 = X_1 + X_2$ è data da

$$f_S(s) = \int_{-\infty}^{+\infty} f_{X_1, X_2}(\xi, s - \xi) d\xi.$$

Ritroviamo dunque la formula (31).

Veniamo ora all'analisi del valore atteso e della varianza della somma $S = X_1 + X_2$.

Ricordiamo dal capitolo precedente (formula (7)) che il valore atteso $\mathbb{E}(X)$ di una generica variabile aleatoria X definita su $(\Omega, \mathcal{F}, \mathbb{P})$ è dato dall'integrale $\int_{\Omega} X(\omega) \mathbb{P}(d\omega)$. In effetti non avevamo riportato la definizione di tale integrale, ma possiamo affermare che tale definizione mantiene tutte le proprietà basilari dell'integrale definito (nel senso di Lebesgue, negli spazi euclidei). In particolare, l'operatore che associa alla funzione integranda (variabile aleatoria) X il numero reale $\mathbb{E}(X)$ è un operatore lineare:

dati cioè α, β due numeri reali e due variabili aleatorie X_1, X_2 tali che $\int_{\Omega} X_1(\omega) \mathbb{P}(d\omega)$ e $\int_{\Omega} X_2(\omega) \mathbb{P}(d\omega)$ sono numeri finiti, allora anche

$$\mathbb{E}(\alpha X_1 + \beta X_2) = \int_{\Omega} [\alpha X_1(\omega) + \beta X_2(\omega)] \mathbb{P}(d\omega)$$

è un numero finito e risulta

$$\begin{aligned} \mathbb{E}(\alpha X_1 + \beta X_2) &= \int_{\Omega} [\alpha X_1(\omega) + \beta X_2(\omega)] \mathbb{P}(d\omega) = \\ &= \alpha \int_{\Omega} X_1(\omega) \mathbb{P}(d\omega) + \beta \int_{\Omega} X_2(\omega) \mathbb{P}(d\omega) = \alpha \mathbb{E}(X_1) + \beta \mathbb{E}(X_2). \end{aligned}$$

che possiamo riguardare quale effetto di un cambiamento di variabile nell'integrale (7).

In virtù della proprietà di linearità, per la variabile aleatoria S possiamo in particolare scrivere

$$\mathbb{E}(S) = \mathbb{E}(X_1) + \mathbb{E}(X_2). \quad (34)$$

Possiamo però guardare alla formula (34) anche da un diverso punto di vista: considerando per semplicità il caso in cui la distribuzione congiunta di (X_1, X_2) ammette una densità congiunta $f(x_1, x_2)$, possiamo scrivere

$$\begin{aligned} \mathbb{E}(S) &= \int_{-\infty}^{+\infty} s \cdot \left[\int_{-\infty}^{+\infty} f(\xi, s - \xi) d\xi \right] ds = \\ &= \int_{-\infty}^{+\infty} \left[\int_{-\infty}^{+\infty} s \cdot f(\xi, s - \xi) d\xi \right] ds \end{aligned}$$

e, ponendo $s = \xi + \eta$,

$$\begin{aligned} \mathbb{E}(S) &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (\xi + \eta) f(\xi, \eta) d\xi d\eta \\ &= \int_{-\infty}^{+\infty} \xi \left[\int_{-\infty}^{+\infty} f(\xi, \eta) d\eta \right] d\xi + \int_{-\infty}^{+\infty} \eta \left[\int_{-\infty}^{+\infty} f(\xi, \eta) d\xi \right] d\eta = \\ &= \int_{-\infty}^{+\infty} \xi \cdot f_{X_1}(\xi) d\xi + \int_{-\infty}^{+\infty} \eta \cdot f_{X_2}(\eta) d\eta = \mathbb{E}(X_1) + \mathbb{E}(X_2). \end{aligned}$$

Sottolineiamo comunque che la formula (34) vale in ogni caso, indipendentemente dal fatto che la distribuzione congiunta di (X_1, X_2) ammetta una densità congiunta $f(x_1, x_2)$, purché X_1 ed X_2 ammettano valori attesi finiti.

Consideriamo ora il caso in cui X_1 ed X_2 ammettano anche varianze finite. Tenendo presente la definizione di varianza (formula (9)) e la proprietà di linearità del valore atteso, si ottiene immediatamente

$$\text{Var}(X) = \mathbb{E}(X^2) - \mu_X^2,$$

$$\text{Var}(aX + b) = a^2 \text{Var}(X),$$

per a e b numeri reali,

$$\text{Var}(S) = \text{Var}(X_1) + \text{Var}(X_2) + 2[\mathbb{E}(X_1 \cdot X_2) - \mathbb{E}(X_1) \cdot \mathbb{E}(X_2)]. \quad (35)$$

La *covarianza* fra X_1 e X_2 , usualmente indicata con $\text{Cov}(X_1, X_2)$, è definita come il valore atteso

$$\text{Cov}(X_1, X_2) = \mathbb{E}[(X_1 - \mathbb{E}(X_1)) \cdot (X_2 - \mathbb{E}(X_2))].$$

Sviluppando il prodotto al secondo membro e applicando ancora una volta la proprietà di linearità dei valori attesi, si ottiene immediatamente anche l'espressione

$$\text{Cov}(X_1, X_2) = \mathbb{E}(X_1 \cdot X_2) - \mathbb{E}(X_1) \cdot \mathbb{E}(X_2).$$

Ovviamente, quindi, $\text{Cov}(X_1, X_2) = \text{Cov}(X_2, X_1)$. In virtù della (35) otteniamo inoltre

$$\text{Var}(S) = \text{Var}(X_1) + \text{Var}(X_1) + 2\text{Cov}(X_1, X_2). \quad (36)$$

Ancora tenendo conto della linearità dei valori attesi, si verifica immediatamente la seguente proprietà di *bilinearità* della covarianza: per un'arbitraria coppia di costanti a' e a'' e per variabili aleatorie X'_1, X''_1, X_2 , si ha ad esempio

$$\text{Cov}(a'X'_1 + a''X''_1, X_2) = a'\text{Cov}(X'_1, X_2) + a''\text{Cov}(X''_1, X_2).$$

Proposizione 6.1

Nel caso in cui X_1 e X_2 sono stocasticamente indipendenti risulta

$$\mathbb{E}(X_1 \cdot X_2) = \mathbb{E}(X_1) \cdot \mathbb{E}(X_2).$$

Dimostrazione

Nel caso assolutamente continuo la tesi si verifica scrivendo

$$\begin{aligned} \mathbb{E}(X_1 \cdot X_2) &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (x_1 \cdot x_2) f(x_1, x_2) dx_1 dx_2 = \\ &= \left[\int_{-\infty}^{+\infty} x f_{X_1}(x) dx \right] \cdot \left[\int_{-\infty}^{+\infty} x_2 f_{X_2}(x_2) dx_2 \right] = \mathbb{E}(X) \cdot \mathbb{E}(Y). \end{aligned}$$

Analogamente, possiamo procedere nel caso discreto.

Immediatamente otteniamo dunque

Proposizione 6.2

Se X_1, X_2 sono stocasticamente indipendenti si ha

$$Var(S) = Var(X_1) + Var(X_2).$$

2.7 Distribuzioni di probabilità in più dimensioni

Le varie nozioni fin qui illustrate per il caso di due variabili aleatorie ammettono un'estensione alquanto diretta nello studio della distribuzione congiunta di una n -upla di variabili aleatorie X_1, \dots, X_n . In quanto segue ci limitiamo a ripercorrere in modo schematico le nozioni essenziali, anche allo scopo di fissare il simbolismo che useremo nei capitoli successivi.

Indichiamo con $F_{X_1, \dots, X_n}(x_1, \dots, x_n)$ la *funzione di ripartizione congiunta* di X_1, \dots, X_n , definita dalla relazione

$$F_{X_1, \dots, X_n}(x_1, \dots, x_n) := \mathbb{P}\{X_1 \leq x_1, \dots, X_n \leq x_n\}.$$

Scriveremo semplicemente F invece che F_{X_1, \dots, X_n} nei casi in cui è chiaro a quali variabili aleatorie ci si stia riferendo.

La *distribuzione marginale* della variabile X_j ($j = 1, \dots, n$) è individuata dalla funzione di ripartizione (unidimensionale)

$$F_{X_j}(x_j) := \lim_{\substack{x_i \rightarrow +\infty \\ i \neq j}} F_{X_1, \dots, X_n}(x_1, \dots, x_n).$$

La distribuzione marginale del vettore $(X_{j_1}, \dots, X_{j_k})$ è individuata dalla funzione di ripartizione (k -dimensionale)

$$F_{X_{j_1}, \dots, X_{j_k}}(x_{j_1}, \dots, x_{j_k}) := \lim_{\substack{x_i \rightarrow +\infty \\ i \neq j_1, \dots, j_k}} F_{X_1, \dots, X_n}(x_1, \dots, x_n).$$

Diciamo che X_1, \dots, X_n sono *stocasticamente indipendenti* se risulta

$$F_{X_1, \dots, X_n}(x_1, \dots, x_n) = F_{X_1}(x_1) \cdot F_{X_2}(x_2) \cdot \dots \cdot F_{X_n}(x_n).$$

Diciamo che X_1, \dots, X_n sono *stocasticamente indipendenti ed identicamente distribuite*, secondo una distribuzione G (e abbrevieremo con il simbolo X_1, \dots, X_n i.i.d. $\sim G$) se risulta

$$F_{X_1, \dots, X_n}(x_1, \dots, x_n) = G(x_1) \cdot G(x_2) \cdot \dots \cdot G(x_n).$$

essendo G una funzione di ripartizione unidimensionale.

Diciamo che la distribuzione congiunta di X_1, \dots, X_n è *assolutamente continua* quando esiste una funzione di densità congiunta. Ciò significa che è possibile rappresentare F_{X_1, \dots, X_n} nella forma

$$F_{X_1, \dots, X_n}(x_1, \dots, x_n) = \int_{-\infty}^{x_1} \dots \int_{-\infty}^{x_n} f(\xi_1, \dots, \xi_n) d\xi_1 \dots d\xi_n.$$

Se una tale rappresentazione è possibile allora esiste quasi ovunque la derivata parziale mista

$$\frac{\partial^n}{\partial x_1 \dots \partial x_n} F_{X_1, \dots, X_n}(x_1, \dots, x_n)$$

ed essa costituisce una possibile scelta quale funzione f ; tale f è la *funzione di densità di congiunta* di F_{X_1, \dots, X_n} .

Una funzione di densità di probabilità n -dimensionale f deve necessariamente soddisfare le due ovvie condizioni di non-negatività e di normalizzazione:

$$\begin{aligned} f(x_1, \dots, x_n) &\geq 0, \\ \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} f(\xi_1, \dots, \xi_n) d\xi_1 \dots d\xi_n &= 1. \end{aligned} \quad (37)$$

Per tutto quanto qui ora segue ci manteniamo nel caso assolutamente continuo.

Per $J = \{j_1, \dots, j_k\} \subset \{1, \dots, n\}$, poniamo $\mathbf{X}_J := (X_{j_1}, \dots, X_{j_k})$ e indichiamo con $\mathbf{X}_{\bar{J}} := (X_i, i \notin J)$ il vettore costituito dalle rimanenti $(n-k)$ variabili. Anche la distribuzione marginale (k -dimensionale) di \mathbf{X}_J ammette una funzione di densità di probabilità. Con qualche abuso di simbolismo, possiamo scrivere che quest'ultima è data dalla formula

$$f_{\mathbf{X}_J}(\mathbf{x}_J) = \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} f(\xi_J, \xi_{\bar{J}}) d\xi_{\bar{J}}.$$

La *distribuzione condizionata* del vettore $\mathbf{X}_{\bar{J}}$ dato ($\mathbf{X}_J = \mathbf{x}_J$) è definita per \mathbf{x}_J tale che $f_{\mathbf{X}_J}(\mathbf{x}_J) > 0$ ed è la distribuzione di probabilità assolutamente continua k -dimensionale che ammette una funzione di densità congiunta data da

$$f_{\mathbf{X}_{\bar{J}}}(\mathbf{x}_{\bar{J}} | \mathbf{X}_J = \mathbf{x}_J) = \frac{f(\mathbf{x}_J, \mathbf{x}_{\bar{J}})}{f_{\mathbf{X}_J}(\mathbf{x}_J)}.$$

Per $\mathbf{x}_J, \mathbf{x}_{\bar{J}}$ tali che $f_{\mathbf{X}_J}(\mathbf{x}_J) > 0, f_{\mathbf{X}_{\bar{J}}}(\mathbf{x}_{\bar{J}}) > 0$, si ottiene immediatamente la formula di Bayes:

$$f_{\mathbf{X}_{\bar{J}}}(\mathbf{x}_{\bar{J}} | \mathbf{X}_J = \mathbf{x}_J) = f_{\mathbf{X}_{\bar{J}}}(\mathbf{x}_{\bar{J}} | \mathbf{X}_J = \mathbf{x}_J) \frac{f_{\mathbf{X}_{\bar{J}}}(\mathbf{x}_{\bar{J}})}{f_{\mathbf{X}_J}(\mathbf{x}_J)} f_{\mathbf{X}_J}(\mathbf{x}_J | \mathbf{X}_{\bar{J}} = \mathbf{x}_{\bar{J}}).$$

X_1, \dots, X_n sono stocasticamente indipendenti se e solo se risulta

$$f(x_1, \dots, x_n) = \prod_{j=1}^n f_{X_j}(x_j),$$

essendo f_{X_j} , per $j = 1, \dots, n$, la funzione di densità di probabilità (unidimensionale) di F_{X_j} . In tal caso si ha, per ogni k , per ogni $J = \{j_1, \dots, j_k\}$, e per ogni \mathbf{x}_J tale che $f_{\mathbf{x}_J}(\mathbf{x}_J) > 0$

$$f_{\mathbf{x}_{\bar{J}}}(\mathbf{x}_{\bar{J}} | \mathbf{X}_J = \mathbf{x}_J) = f_{\mathbf{x}_{\bar{J}}}(\mathbf{x}_{\bar{J}}).$$

X_1, \dots, X_n sono i.i.d. $\sim G$ se e solo se risulta

$$f(x_1, \dots, x_n) = \prod_{j=1}^n g(x_j)$$

essendo g la funzione di densità di probabilità (unidimensionale) di G .

Per descrivere la distribuzione congiunta di alcuni vettori aleatori che verranno studiati successivamente, faremo uso della seguente definizione. Con il simbolo $\lambda_n(B)$ indichiamo la misura di Lebesgue di un insieme $B \in \mathcal{B}^{(n)}$.

Definizione 7.1

Sia $B \in \mathcal{B}^{(n)}$ con $\lambda_n(B) > 0$. $\mathbf{X} \equiv (X_1, \dots, X_n)$ ha una *distribuzione uniforme* su B se, per una costante positiva k , risulta

$$f_{\mathbf{X}}(\mathbf{x}) = k \cdot \mathbf{1}_B(\mathbf{x}) = \begin{cases} k & \text{per } \mathbf{x} \in B \\ 0 & \text{per } \mathbf{x} \notin B \end{cases}.$$

Ciò verrà anche indicato con il simbolo $\mathbf{X} \sim \mathcal{U}_n(B)$. Ovviamente, per la condizione di normalizzazione (37), deve essere $k = \frac{1}{\lambda_n(B)}$ e tale definizione è equivalente alla seguente condizione

$$P\{\mathbf{X} \in C\} = \frac{\lambda_n(B \cap C)}{\lambda_n(B)}, \forall C \in \mathcal{B}^{(n)}.$$

Indichiamo con $\Gamma^{(n)}$ il cubo unitario

$$\Gamma^{(n)} = \{x \equiv (x_1, \dots, x_n) \in \mathbb{R}^n | x_j \in [0, 1], j = 1, \dots, n\}.$$

Osserviamo che un vettore $\mathbf{X} \equiv (X_1, \dots, X_n)$ ha una distribuzione uniforme su $\Gamma^{(n)}$ se e solo se le sue componenti X_1, \dots, X_n sono variabili aleatorie indipendenti identicamente distribuite con distribuzione (uni-dimensionale) uniforme sull'intervallo $[0, 1]$, cioè

$$f_{X_j}(x_j) = \begin{cases} 1 & \text{per } x_j \in [0, 1] \\ 0 & \text{per } x_j \notin [0, 1] \end{cases}.$$

Enunciamo ora un risultato fondamentale circa trasformazioni di vettori aleatori con distribuzione congiunta assol. continua. Questo risultato costituisce una diretta generalizzazione della precedente Proposizione 5.1 e verrà utilizzato ripetutamente nel seguito.

Siano $E, E' \subset \mathbb{R}^n$ due insiemi aperti ed \mathbf{X} un vettore aleatorio a valori in E . Sia inoltre $\mathcal{T} : E \rightarrow E'$ una trasformazione biunivoca e consideriamo il vettore di variabile aleatorie $\mathbf{Y} = \mathcal{T}(\mathbf{X})$. Ci si pone allora il problema di ottenere la distribuzione di probabilità congiunta di \mathbf{Y} . A tale scopo indichiamo con $\mathcal{V}_1, \dots, \mathcal{V}_n$ le componenti della trasformazione \mathcal{V} inversa della trasformazione \mathcal{T} , cosicchè si abbia

$$\begin{aligned} X_1 &= \mathcal{V}_1(Y_1, \dots, Y_n) \\ &\dots\dots\dots \\ &\dots\dots\dots \\ X_n &= \mathcal{V}_n(Y_1, \dots, Y_n) \end{aligned}$$

Assumiamo che \mathcal{V} sia dotata di derivate parziali prime continue $\frac{\partial \mathcal{V}_i}{\partial y_j}$ ed indichiamo con $J_{\mathcal{V}}(\mathbf{y})$ il determinante della relativa matrice jacobiana:

$$\begin{pmatrix} \frac{\partial \mathcal{V}_1}{\partial y_1} & \dots\dots & \frac{\partial \mathcal{V}_1}{\partial y_n} \\ \dots & \dots\dots & \dots \\ \dots & \dots\dots & \dots \\ \frac{\partial \mathcal{V}_n}{\partial y_1} & \dots\dots & \frac{\partial \mathcal{V}_n}{\partial y_n} \end{pmatrix}.$$

Si ha allora la seguente

Proposizione 7.1

Se \mathbf{X} ammette una funzione di densità congiunta $f_{\mathbf{X}}(\mathbf{x})$, allora \mathbf{Y} ammette una funzione di densità congiunta $f_{\mathbf{Y}}(\mathbf{y})$ e risulta, per $\mathbf{y} \equiv (y_1, \dots, y_n) \in E'$,

$$f_{\mathbf{Y}}(\mathbf{y}) = f_{\mathbf{X}}(\mathcal{V}_1(\mathbf{y}), \dots, \mathcal{V}_n(\mathbf{y})) |J_{\mathcal{V}}(\mathbf{y})|. \quad (38)$$

Dimostrazione

Per ogni sottoinsieme $B \subset E'$, con $B \in \mathcal{B}^{(n)}$, possiamo scrivere

$$\begin{aligned} \mathbb{P}\{\mathbf{Y} \in B\} &= \mathbb{P}\{\mathcal{V}(\mathbf{Y}) \in \mathcal{V}(B)\} = \mathbb{P}\{\mathbf{X} \in \mathcal{V}(B)\} \\ &= \int \dots \int_{\mathcal{V}(B)} f_{\mathbf{X}}(\mathbf{x}) dx_1 \dots dx_n. \end{aligned}$$

Eseguendo in quest'ultimo integrale la trasformazione di variabili $\mathbf{X} = \mathcal{V}(\mathbf{Y})$, si ha

$$\int \dots \int_{\mathcal{V}(B)} f_{\mathbf{X}}(\mathbf{x}) dx_1 \dots dx_n = \int \dots \int_B f_{\mathbf{X}}(\mathcal{V}_1(\mathbf{y}), \dots, \mathcal{V}_n(\mathbf{y})) |J_{\mathcal{V}}(\mathbf{y})| dy_1 \dots dy_n.$$

La (38) segue allora dall'arbitrarietà di B e dalla definizione di funzione di densità congiunta.

3 Famiglie notevoli di distribuzioni di probabilità unidimensionali

In questo capitolo verranno elencate le più notevoli famiglie di distribuzioni di probabilità unidimensionali. Si tratta delle distribuzioni che vengono più spesso incontrate nello studio della probabilità e della statistica matematica. Cominceremo innanzitutto con il richiamare succintamente le più importanti famiglie di distribuzioni di probabilità discrete.

3.1 Distribuzioni di probabilità discrete notevoli

3.1.1 Distribuzioni di Bernoulli e binomiali

X ha una distribuzione di *Bernoulli* con parametro θ ($0 < \theta < 1$) se risulta

$$P\{X = 1\} = \theta; P\{X = 0\} = 1 - \theta.$$

Come immediatamente si verifica,

$$\mathbb{E}(X) = \theta; \text{Var}(X) = \theta(1 - \theta).$$

Y è una variabile con distribuzione *binomiale* con parametri n e θ (scriveremo brevemente $Y \sim b(n, \theta)$) se risulta

$$p_k \equiv P\{Y = k\} = \binom{n}{k} \theta^k (1 - \theta)^{n-k}, k = 0, 1, \dots, n.$$

Consideriamo gli indicatori di n prove bernoulliane

$$X_1, X_2, \dots, X_n$$

(cioè variabili bernoulliane indipendenti con uguale parametro θ); allora si ha

$$Y = \sum_{h=1}^n X_h \sim b(n, \theta).$$

Quindi, per $Y \sim b(n, \theta)$, risulta

$$\mathbb{E}(Y) = n\theta; \text{Var}(Y) = n\theta(1 - \theta).$$

Il valore *modale* di Y è dato da $[(n+1)\theta]$: come si verifica facilmente, infatti

$$\frac{p_k}{p_{k+1}} \leq 1 \text{ per } k \leq [(n+1)\theta], \frac{p_k}{p_{k+1}} \geq 1 \text{ per } k \geq [(n+1)\theta]$$

se $(n+1)\theta$ è intero, si ha che $(n+1)\theta$ e $(n+1)\theta - 1$ sono entrambi valori modali.

3.1.2 Distribuzione di Poisson

Sia $\lambda > 0$. X segue una distribuzione di Poisson con parametro λ se, per $k = 0, 1, 2, \dots$,

$$P\{X = k\} = e^{-\lambda} \frac{\lambda^k}{k!}.$$

In tal caso scriveremo $X \sim \mathcal{P}(\lambda)$.

Risulta

$$\mathbb{E}(X) = \sum_{k=0}^{\infty} k P\{X = k\} = e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^k}{(k-1)!} = \lambda e^{-\lambda} \sum_{h=0}^{\infty} \frac{\lambda^h}{h!} = \lambda.$$

$$\begin{aligned} \mathbb{E}(X^2) &= \sum_{k=0}^{\infty} k^2 P\{X = k\} = \sum_{k=1}^{\infty} k^2 e^{-\lambda} \frac{\lambda^k}{k!} = \\ &= \sum_{h=0}^{\infty} (h+1) e^{-\lambda} \frac{\lambda^{h+1}}{h!} = \lambda \left[\sum_{h=0}^{\infty} h e^{-\lambda} \frac{\lambda^h}{h!} + \sum_{h=0}^{\infty} e^{-\lambda} \frac{\lambda^h}{h!} \right] = \lambda(\lambda + 1). \end{aligned}$$

Da cui

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

Per un fissato $\lambda > 0$, consideriamo la sottofamiglia delle distribuzioni binomiali $b(n, \frac{\lambda}{n})$ e calcoliamo il limite $\lim_{n \rightarrow \infty} p_k^{(n)}$, dove si è posto

$$p_k^{(n)} = \begin{cases} \binom{n}{k} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} & k = 0, 1, \dots, n \\ 0 & k = n+1, n+2, \dots \end{cases}.$$

Si ha

$$\lim_{n \rightarrow \infty} p_k^{(n)} = e^{-\lambda} \frac{\lambda^k}{k!}, k = 0, 1, \dots \quad (39)$$

La (39) giustifica quindi, per n abbastanza grande e p abbastanza piccolo, la seguente approssimazione (*approssimazione di Poisson*):

$$\binom{n}{k} p^k (1-p)^{n-k} \approx e^{-np} \frac{np^k}{k!}.$$

Come vedremo successivamente, la distribuzione di Poisson si presenta come distribuzione di probabilità del valore assunto da un processo di Poisson ad un fissato tempo t .

3.1.3 Distribuzione geometrica e di Pascal

Sia X_1, X_2, \dots una successione di variabili bernoulliane indipendenti con uguale parametro θ .

e sia T_1 la variabile aleatoria definita da

$$T_1 = \min\{n > 0 | X_n = 1\}.$$

Interpretando le variabili X_1, X_2, \dots come indicatori di prove bernoulliane, T_1 viene ad assumere il significato di "tempo d'attesa fino al primo successo".

Come è immediato verificare, T_1 segue una distribuzione *geometrica* di parametro θ :

$$P\{T_1 = k\} = (1 - \theta)^{k-1} \theta, k = 1, 2, \dots$$

Si verifica facilmente che risulta

$$\mathbb{E}(T_1) = \sum_{k=1}^{\infty} k P\{T_1 = k\} = \theta \sum_{k=1}^{\infty} k (1 - \theta)^{k-1} = \frac{1}{\theta}.$$

Consideriamo ora la variabile aleatoria T_r (*tempo d'attesa fino all' r -esimo successo*) definita da

$$T_r = \min\{n > 0 | \sum_{h=1}^n X_h = r\}.$$

T_r segue una distribuzione *di Pascal* di parametri r e θ : per $k = r, r + 1, \dots$, si ha

$$P\{T_r = k\} = \binom{k-1}{r-1} \theta^r (1 - \theta)^{k-r}.$$

T_r si può vedere come la somma di r variabili aleatorie i.i.d., con distribuzione geometrica di parametro θ e quindi risulta

$$\mathbb{E}(T_r) = \frac{r}{\theta}$$

3.1.4 Distribuzione ipergeometrica

X ha una distribuzione *ipergeometrica* con parametri n, N, N_1 ($N \geq \max(N_1, n)$), se risulta, per

$$\max(0, n + N_1 - N) \leq k \leq \min(n, N_1),$$

$$P\{X = k\} = \frac{\binom{N_1}{k} \binom{N-N_1}{n-k}}{\binom{N}{n}}.$$

Come è noto, questa distribuzione si presenta come la distribuzione di probabilita' del numero X di elementi, aventi una proprieta' C , estratti in n

estrazioni casuali senza reimbussolamento da una popolazione costituita da N elementi, di cui esattamente N_1 aventi la caratteristica C stessa.

In caso, invece, di estrazioni casuali con reimbussolamento, le diverse estrazioni risultano indipendenti, dando luogo a prove bernoulliane con $\theta = \frac{N_1}{N}$, per cui la corrispondente distribuzione per X sarebbe $b\left(n, \frac{N_1}{N}\right)$.

Osserviamo, a tale proposito, che risulta, per $0 < \theta < 1$, e $k \leq n$ fissati,

$$\lim_{N \rightarrow \infty} \frac{\binom{|N\theta|}{k} \binom{|N(1-\theta)|}{n-k}}{\binom{N}{n}} = \binom{n}{k} \theta^k (1-\theta)^{n-k}.$$

3.2 Distribuzione uniforme

Una distribuzione di probabilita' *uniforme* sull'intervallo (a, b) ($a < b$) e' una distribuzione assolutamente continua con funzione di densita'

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{per } a \leq x \leq b \\ 0 & \text{altrove} \end{cases};$$

la relativa funzione di ripartizione e' quindi data da

$$F(x) = \begin{cases} 0 & \text{per } x \leq a \\ \frac{x-a}{b-a} & \text{per } a \leq x \leq b \\ 1 & \text{per } x \geq b \end{cases}.$$

Se $\in e'$ una variabile aleatoria con distribuzione uniforme in (a, b) (scriveremo $X \sim R(a, b)$), risulta

$$\mathbb{E}(X) = \int_a^b \frac{x}{b-a} dx = \frac{a+b}{2},$$

$$\mathbb{E}(X^2) = \int_a^b \frac{x^2}{b-a} dx = \frac{1}{3} \frac{b^3 - a^3}{b-a};$$

da cui

$$Var(X) = \frac{(b-a)^2}{12}.$$

3.3 Distribuzione esponenziale

Sia $\lambda > 0$ una quantita' assegnata. La funzione di densita' di probabilita' di una distribuzione *esponenziale* di parametro λ ha la forma

$$f(x) = \begin{cases} \lambda \exp\{-\lambda x\} & \text{per } x \geq 0 \\ 0 & \text{per } x < 0 \end{cases}.$$

Quindi la corrispondente funzione di ripartizione e' data da

$$F(x) = \begin{cases} 1 - \exp\{-\lambda x\} & \text{per } x \geq 0 \\ 0 & \text{per } x < 0 \end{cases}$$

Diremo che una variabile aleatoria X segue una distribuzione *esponenziale* di parametro λ (e scriveremo brevemente $X \sim \mathcal{E}(\lambda)$) se la sua distribuzione è assolutamente continua e la densità è della forma qui specificata.

Se $X \sim \mathcal{E}(\lambda)$, risulta

$$\mathbb{E}(X) = \int_0^{\infty} \lambda x \exp\{-\lambda x\} dx = \frac{1}{\lambda}, \mathbb{E}(X^2) = \int_0^{\infty} \lambda x^2 \exp\{-\lambda x\} dx = \frac{2}{\lambda^2},$$

da cui

$$\text{Var}(X) = \frac{1}{\lambda^2}.$$

Come vedremo successivamente, la distribuzione esponenziale si presenta come la distribuzione di probabilità degli *intertempi* in un processo di Poisson.

La *funzione di intensità* di una variabile aleatoria X , con distribuzione di probab. ass. continua concentrata su $[0, +\infty)$, è definita dalla relazione

$$r(x) = \frac{f(x)}{1 - F(x)} = \lim_{\Delta t \rightarrow 0} \frac{P\{X \leq x + \Delta t | X > t\}}{\Delta t}.$$

Essendo $f(x) = \frac{d}{dx} F(x)$ e tenendo conto della condizione iniziale $F(0) = 0$, vale la relazione

$$F(x) = 1 - \exp\left\{-\int_0^x r(\xi) d\xi\right\}.$$

Si verifica allora immediatamente che le distribuzioni esponenziali sono caratterizzate dalla proprietà

$$r(x) = \text{costante}, x \geq 0.$$

Tale proprietà è equivalente alla seguente proprietà, detta di *manca di memoria*:

Se X è una variabile aleatoria con distribuzione esponenziale e t, s sono due arbitrarie quantità positive, allora risulta

$$P\{X > t + s | T > s\} = P\{X > t\} = \text{costante in } s.$$

3.4 Distribuzione gamma

Siano α e β quantità positive assegnate. Una distribuzione gamma di parametri α e β (indicata con il simbolo $G(\alpha, \beta)$) è individuata dalla funzione di densità:

$$f(x) = \begin{cases} K x^{\alpha-1} \exp\{-\beta x\} & \text{per } x \geq 0 \\ 0 & \text{per } x < 0 \end{cases}.$$

Il valore della costante di normalizzazione K dipende ovviamente dai parametri α e β ed è dato da

$$K = \left[\int_0^{\infty} x^{\alpha-1} \exp\{-\beta x\} dx \right]^{-1} =$$

$$\left[\int_0^\infty \left(\frac{z}{\beta} \right)^{\alpha-1} \exp\{-z\} \frac{dz}{\beta} \right]^{-1} = \left[\frac{1}{\beta^\alpha} \int_0^\infty z^{\alpha-1} \exp\{-z\} dz \right]^{-1},$$

(dove si e' posto $z = \beta x$, da cui $dx = \frac{dz}{\beta}$).

Per valori α reali positivi risulta

$$\int_0^\infty z^{\alpha-1} \exp\{-z\} dz = \Gamma(\alpha)$$

dove $\Gamma(\cdot)$ indica la *funzione gamma di Eulero*; scriviamo quindi

$$K = \frac{\beta^\alpha}{\Gamma(\alpha)}.$$

Tramite integrazione per parti si ottiene, per la funzione Γ , la relazione

$$\Gamma(\alpha) = (\Gamma(\alpha - 1)) \Gamma(\alpha - 1), \alpha \geq 1. \quad (40)$$

Risultando

$$\Gamma(1) = \int_0^\infty \exp\{-z\} dz = 1,$$

si verifica immediatamente che, per $\alpha \in \mathbb{N}$, si ha

$$\Gamma(\alpha) = (\alpha - 1)!.$$

Se $X \sim G(\alpha, \beta)$ allora, per $k = 1, 2, \dots$, risulta

$$\mathbb{E}(X^h) = \frac{\alpha(\alpha + 1) \dots (\alpha + h - 1)}{\beta^h},$$

infatti

$$\begin{aligned} \mathbb{E}(X^h) &= \int_0^\infty x^h \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} \exp\{-\beta x\} dx = \frac{\beta^\alpha}{\Gamma(\alpha)} \int_0^\infty x^{\alpha+h-1} \exp\{-\beta x\} dx = \\ &= \frac{\beta^\alpha}{\Gamma(\alpha)} \int_0^\infty \frac{z^{\alpha+h-1}}{\beta^{\alpha+h-1}} \exp\{-z\} \frac{dz}{\beta} = \frac{\Gamma(\alpha + h)}{\beta^h \Gamma(\alpha)}. \end{aligned}$$

Ne deriva, in particolare,

$$\mathbb{E}(X) = \frac{\alpha}{\beta}, \text{Var}(X) = \frac{\alpha}{\beta^2}.$$

Nel caso in cui sia $\alpha = k \in \mathbb{N}$, la distribuzione $G(\alpha, \beta)$ prende anche il nome di *distribuzione di Erlang*.

In tal caso e' possibile esprimere in forma esplicita la funzione di distribuzione: tramite un procedimento di induzione e integrando successivamente per parti si può mostrare che risulta, per $x > 0$,

$$F(x) = \exp\{-\beta x\} \sum_{h=k}^{\infty} \frac{(\beta x)^h}{h!}. \quad (41)$$

Ricaveremo però, successivamente, tale formula in modo alternativo considerando la distribuzione tempo d'attesa fino al k -esimo arrivo per un processo di Poisson.

Il seguente risultato mostra una proprietà di chiusura, della famiglia delle distribuzioni gamma, rispetto alla somma di variabili aleatorie indipendenti; tale risultato ha un'importanza fondamentale, come in parte vedremo, per molti aspetti, sia teorici che applicativi, del calcolo delle probabilità.

Proposizione 4.1

Siano X, Y due variabili aleatorie e poniamo $S = X + Y$.

Se X, Y sono tali che

$$X \sim G(\alpha_1, \beta), Y \sim G(\alpha_2, \beta),$$

e

X, Y stocasticamente indipendenti,

allora risulta

$$S \sim G(\alpha_1 + \alpha_2, \beta). \quad (42)$$

Dimostrazione

Ricordando la formula (33) del precedente capitolo, possiamo scrivere

$$f_S(s) = \int_0^s f_X(x) f_Y(s-x) dx,$$

Imponendo

$$f_X(x) = \frac{\beta^{\alpha_1}}{\Gamma(\alpha_1)} x^{\alpha_1-1} \exp\{-\beta x\} \mathbf{1}_{\{x>0\}}, f_Y(y) = \frac{\beta^{\alpha_2}}{\Gamma(\alpha_2)} y^{\alpha_2-1} \exp\{-\beta y\} \mathbf{1}_{\{y>0\}},$$

ottendiamo, per $s > 0$,

$$\begin{aligned} f_S(s) &= \int_0^s \frac{\beta^{\alpha_1+\alpha_2}}{\Gamma(\alpha_1)\Gamma(\alpha_2)} x^{\alpha_1-1} (s-x)^{\alpha_2-1} \exp\{-\beta s\} ds = \\ &= \frac{\beta^{\alpha_1+\alpha_2}}{\Gamma(\alpha_1)\Gamma(\alpha_2)} \exp\{-\beta s\} \int_0^s x^{\alpha_1-1} (s-x)^{\alpha_2-1} dx. \end{aligned} \quad (43)$$

Tramite la trasformazione $x = ts$ otteniamo

$$\int_0^s x^{\alpha_1-1} (s-x)^{\alpha_2-1} dx = \mathbf{I}(\alpha_1, \alpha_2) s^{\alpha_1+\alpha_2-1},$$

avendo indicato con il simbolo $\mathbf{I}(\alpha_1, \alpha_2)$ l'integrale

$$\mathbf{I}(\alpha_1, \alpha_2) \equiv \int_0^1 t^{\alpha_2-1} (1-t)^{\alpha_1-1} dy,$$

che risulta essere dunque una quantità costante, indipendente da s .

Per $s > 0$ abbiamo quindi, dalla (43),

$$f_S(s) = \frac{\mathbf{I}(\alpha_1, \alpha_2)}{\Gamma(\alpha_1)\Gamma(\alpha_2)} \beta^{\alpha_1+\alpha_2} s^{\alpha_1+\alpha_2-1} \exp\{-\beta s\},$$

e dunque la (42), e si è contemporaneamente dimostrata, $\forall a, b > 0$, l'identità

$$\int_0^1 t^{a-1} (1-t)^{b-1} dt = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}. \quad (44)$$

Corollario 4.2

Siano X_1, X_2, \dots, X_n variabili aleatorie indipendenti con

$$X_h \sim \mathcal{E}(\lambda), h = 1, 2, \dots, n.$$

Allora

$$\sum_{h=1}^n X_h \sim G(n, \lambda).$$

Dimostrazione

Basta notare che una distribuzione esponenziale di parametro λ coincide con una distribuzione gamma con parametri $\alpha = 1, \beta = \lambda$.

3.5 Distribuzione gaussiana (o normale)

Diremo che una variabile aleatoria Z segue una distribuzione *gaussiana standard*, e scriveremo brevemente,

$$Z \sim \mathcal{N}(0, 1),$$

se la distribuzione di probabilità di Z è assolutamente continua ed ammette una funzione di densità di probabilità data da

$$f_Z(x) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{x^2}{2}\right\}. \quad (45)$$

Viene usualmente indicata con il simbolo Φ la funzione di ripartizione di Z ; cioè si pone

$$\Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{\xi^2}{2}\right\} d\xi.$$

E' noto che non si sa scrivere in modo esplicito tale funzione in termini di altre funzioni elementari.

La quantità $\frac{1}{\sqrt{2\pi}}$ ha il ruolo di costante di normalizzazione, risulta infatti

$$I \equiv \int_{-\infty}^{+\infty} \exp\left\{-\frac{x^2}{2}\right\} dx = \sqrt{2\pi}, \quad (46)$$

o, equivalentemente,

$$\lim_{x \rightarrow +\infty} \Phi(x) = 1.$$

L'identità (46) si dimostra come segue. Scriviamo innanzitutto

$$I^2 = \left[\int_{-\infty}^{+\infty} \exp\left\{-\frac{x^2}{2}\right\} dx \right] \left[\int_{-\infty}^{+\infty} \exp\left\{-\frac{y^2}{2}\right\} dy \right].$$

Per il Teorema di Fubini sarà anche

$$I^2 = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \exp\left\{-\frac{x^2 + y^2}{2}\right\} dx dy.$$

Eseguendo in tale integrale un cambiamento delle coordinate cartesiane in coordinate polari, segue

$$I^2 = \int_{-\pi}^{+\pi} \int_0^{+\infty} \exp\left\{-\frac{\rho^2}{2}\right\} \rho d\rho d\theta,$$

ed essendo

$$\int_0^{+\infty} \exp\left\{-\frac{\rho^2}{2}\right\} \rho d\rho = 1,$$

possiamo concludere

$$I^2 = \int_{-\pi}^{+\pi} d\theta = 2\pi,$$

da cui otteniamo la (46).

Eseguendo un opportuno cambiamento di variabile, possiamo ricavare da quest'ultima un'ulteriore identità, che si rivelerà utile varie volte in quanto segue: consideriamo la trasformazione

$$u = \frac{x^2}{2} \tag{47}$$

(cosicché $dx = \frac{\sqrt{2}}{2} u^{-\frac{1}{2}} du$), risulta

$$I = 2 \int_0^{+\infty} \exp\left\{-\frac{x^2}{2}\right\} dx = \int_0^{+\infty} \exp\{-u\} \sqrt{2} u^{-\frac{1}{2}} du = \sqrt{2} \Gamma\left(\frac{1}{2}\right),$$

da cui

$$\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}. \tag{48}$$

Passiamo, a questo punto, a ricavare i momenti della distribuzione gaussiana standardizzata.

Si può facilmente mostrare che essa ammette momenti di qualunque ordine, cioè che sono finiti tutti gli integrali del tipo

$$\int_0^{+\infty} x^k \exp\left\{-\frac{x^2}{2}\right\} dx, k = 1, 2, \dots$$

Essendo la funzione di densità in (45) una funzione pari, $x^k \exp\{-\frac{x^2}{2}\}$ risulterà quindi una funzione pari o una funzione dispari a seconda che sia pari o dispari l'esponente k . Allora potremo scrivere subito

$$\mathbb{E}(Z^k) = 0, \text{ per } k = 1, 3, 5, \dots$$

Per calcolare i momenti di ordine pari, consideriamo di nuovo la trasformazione definita in (47) e teniamo conto della (48):

$$\begin{aligned} \mathbb{E}(Z^{2h}) &= \int_{-\infty}^{+\infty} x^{2h} \frac{1}{\sqrt{2\pi}} \exp\{-\frac{x^2}{2}\} dx = \frac{\sqrt{2}}{\sqrt{\pi}} \int_0^{+\infty} x^{2h} \exp\{-\frac{x^2}{2}\} dx = \\ &= \frac{\sqrt{2}}{\sqrt{\pi}} \int_0^{+\infty} (2u)^h \exp\{-u\} \frac{\sqrt{2}}{2} u^{-\frac{1}{2}} du = \frac{2^h}{\sqrt{\pi}} \int_0^{+\infty} u^{h-\frac{1}{2}} \exp\{-u\} du = \frac{2^h}{\sqrt{\pi}} \Gamma\left(\frac{h+1}{2}\right) = \\ &= \frac{2^h}{\sqrt{\pi}} \frac{2h-1}{2} \frac{2h-3}{2} \dots \frac{1}{2} \Gamma\left(\frac{1}{2}\right) = (2h-1)(2h-3) \dots \cdot 3 \cdot 1. \end{aligned}$$

Osserviamo che in particolare risulta

$$\mathbb{E}(Z) = 0, \text{Var}(Z) = 1;$$

cioè Z è una variabile aleatoria *standardizzata*.

Fissiamo ora due arbitrari numeri reali μ e σ , con $\sigma > 0$, e consideriamo la variabile aleatoria

$$X = \sigma Z + \mu.$$

Viene chiamata distribuzione *gaussiana* di parametri μ e σ^2 la distribuzione di probabilità di X e scriveremo brevemente

$$X \sim \mathcal{N}(\mu, \sigma^2).$$

E' immediato verificare che risulta

$$\mathbb{E}(Z) = \mu, \text{Var}(Z) = \sigma^2.$$

Vogliamo ora calcolare la funzione di densità di X . Si ha

$$\begin{aligned} F_X(x) &= P\{X \leq x\} = P\{\sigma Z + \mu \leq x\} = \\ &= P\{Z \leq \frac{x-\mu}{\sigma}\} = \Phi\left(\frac{x-\mu}{\sigma}\right); \end{aligned}$$

$$f_X(x) = \frac{d}{dx} \Phi\left(\frac{x-\mu}{\sigma}\right) = \frac{1}{\sigma} f_Z\left(\frac{x-\mu}{\sigma}\right) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2} \left(\frac{x-\mu}{\sigma}\right)^2\right\}.$$

I valori della funzione $\Phi(x)$ (per $x > 0$) si leggono sulle apposite "tavole della distribuzione gaussiana"; per ottenere dalle tavole i valori di $\Phi(x)$ per $x < 0$ basta osservare che risulta

$$\Phi(x) = 1 - \Phi(-x), \forall x \in \mathbb{R}.$$

Il ruolo fondamentale della distribuzione gaussiana è in parte spiegato dai risultati circa la convergenza, verso la distribuzione gaussiana stessa, delle distribuzioni di somme standardizzate di variabili aleatorie. Di tale corpo di risultati del Calcolo delle Probabilità ricordiamo in particolare il seguente

Teorema 5.1 (Lindberg-Levy)

Sia

$$X_1, X_2, \dots$$

una successione di variabili aleatorie indipendenti identicamente distribuite, con distribuzione $F(x)$ tale che

$$\mathbb{E}(X_h) = \mu, \text{Var}(X_h) = \sigma^2 < \infty$$

e, per il resto, arbitraria. Poniamo, per $n = 2, 3, \dots$

$$Y_n = \frac{\sum_{h=1}^n X_h - n\mu}{\sqrt{n}\sigma}.$$

Si ha allora, per arbitrari $a < b$,

$$\lim_{n \rightarrow \infty} P\{a < Y_n < b\} = \Phi(b) - \Phi(a) = \int_a^b \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{\xi^2}{2}\right\} d\xi.$$

3.6 Distribuzione del chi-quadro

Vogliamo ottenere la distribuzione di probabilità della variabile aleatoria

$$\Upsilon_n = \sum_{h=1}^n X_h^2,$$

dove

$$X_1, X_2, \dots, X_n$$

sono variabili aleatorie indipendenti con distribuzione gaussiana standard.

Per $n = 1$, otteniamo, $\forall y > 0$,

$$\begin{aligned} F_{\Upsilon_1}(y) &= P\{\Upsilon_1 \leq y\} = P\{\sqrt{y} \leq X_1 \leq \sqrt{y}\} = \Phi(\sqrt{y}) - \Phi(-\sqrt{y}) = \\ &= 2\Phi(\sqrt{y}) - 1. \end{aligned}$$

Ne segue, per $y > 0$,

$$f_{\Upsilon_1}(y) = \frac{d}{dy} F_{\Upsilon_1}(y) = \frac{1}{\sqrt{2\pi y}} \exp\left\{-\frac{y}{2}\right\},$$

mentre, per $y < 0$, risulta ovviamente $f_{\Upsilon_1}(y) = 0$. Possiamo scrivere dunque

$$\Upsilon_1 \sim G\left(\frac{1}{2}, \frac{1}{2}\right).$$

Dunque Υ_n risulta essere la somma di variabili aleatorie indipendenti e, ricordando la **Proposizione 4.1**, possiamo concludere scrivendo

$$\Upsilon_n \sim G\left(\frac{n}{2}, \frac{1}{2}\right).$$

Tale distribuzione è nota nella letteratura statistica con il nome di distribuzione *del chi-quadro con n gradi di libertà*.

3.7 Distribuzione t di Student

Consideriamo ora una famiglia di distribuzioni di probabilità che riveste di una grande importanza nella Statistica matematica. Iniziamo considerando una variabile aleatoria

$$Z = \frac{X\sqrt{n}}{\sqrt{Y}},$$

dove X, Y sono stocasticamente indipendenti e tali che $X \sim \mathcal{N}(0, 1)$, $Y \sim \text{chi-quadro}(n)$, $n = 1, 2, \dots$

La funzione di ripartizione di Z è data da

$$\begin{aligned} F_Z(z) &= P\left\{X \leq \frac{z\sqrt{Y}}{\sqrt{n}}\right\} = \int_0^\infty \left[\int_{-\infty}^{\frac{z\sqrt{y}}{\sqrt{n}}} \phi_X(x) dx \right] f_Y(y) dy = \\ &= \int_0^\infty \Phi\left(z\left(\frac{y}{n}\right)^{\frac{1}{2}}\right) f_Y(y) dy, \end{aligned}$$

da cui

$$\begin{aligned} f_Z(z) &= \frac{d}{dz} F_Z(z) = \int_0^\infty \left(\frac{y}{n}\right)^{\frac{1}{2}} \phi\left(\frac{z\sqrt{y}}{\sqrt{n}}\right) f_Y(y) dy = \frac{1}{\sqrt{2\pi}} \frac{1}{2^n} \exp\left\{-\frac{z^2 y}{2n}\right\} \left(\frac{y}{n}\right)^2 \\ &= \frac{1}{\sqrt{2\pi}} \int_0^\infty \left(\frac{y}{n}\right)^{\frac{1}{2}} \exp\left\{-\frac{z^2 y}{2n}\right\} \left(\frac{1}{2}\right)^{\frac{n}{2}} \frac{1}{\Gamma\left(\frac{n}{2}\right)} y^{\frac{n}{2}-1} \exp\left\{-\frac{y}{2}\right\} dy = \\ &\quad \left(\frac{1}{n\pi}\right)^{\frac{1}{2}} \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} \frac{1}{\left(1 + \frac{z^2}{n}\right)^{\frac{n+1}{2}}}. \end{aligned}$$

Tale distribuzione è detta *distribuzione t di Student con n gradi di libertà*.

Per $n > 1$ risulta finito l'integrale

$$\int_0^\infty \frac{z}{\left(1 + \frac{z^2}{n}\right)^{\frac{n+1}{2}}} dz;$$

in tal caso esiste dunque il valore atteso $\mathbb{E}(Z)$ e risulta ovviamente

$$\mathbb{E}(Z) = 0.$$

Per $n > 2$ si ha inoltre, introducendo la trasformazione $z = \left(\frac{nv}{1-v}\right)^{\frac{1}{2}}$

$$\int_0^\infty \frac{z^2}{\left(1 + \frac{z^2}{n}\right)^{\frac{n+1}{2}}} dz = \frac{n^{\frac{3}{2}}}{2} \int_0^1 v^{\frac{1}{2}} (1-v)^{\frac{n-2}{2}} dv = \frac{n^{\frac{3}{2}}}{2} \frac{\Gamma\left(\frac{3}{2}\right) \Gamma\left(\frac{n-2}{2}\right)}{\Gamma\left(\frac{n+1}{2}\right)},$$

da cui, ricordando (40), (44) e (48)

$$Var(Z) = \mathbb{E}(Z^2) = \frac{n}{n-2}.$$

Ponendo in particolare $n = 1$ otteniamo la distribuzione di Cauchy, che ha quindi funzione di densità

$$f(x) = \frac{1}{\pi} \frac{1}{1+x^2}$$

e funzione di distribuzione

$$F(x) = \frac{1}{2} + \frac{1}{\pi} \operatorname{arctg}(x).$$

Per μ e τ numeri reali assegnati, con $\tau > 0$, e Z con distribuzione t di Student con n gradi libertà, consideriamo ora la variabile aleatoria

$$W = \frac{Z}{\sqrt{\tau}} + \mu.$$

La funzione di densità di W sarà quindi data da

$$f_W(x) = \left(\frac{\tau}{n\pi}\right)^{\frac{1}{2}} \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} \left[1 + \frac{\tau(x-\mu)^2}{n}\right]^{-\frac{n+1}{2}}.$$

La relativa distribuzione di probabilità viene detta *distribuzione t di Student* con n gradi libertà, parametro di posizione μ e parametro di precisione τ .

3.8 Distribuzione beta

Siano a e b due costanti positive assegnate. Diciamo che X segue una distribuzione *beta* con parametri a e b , e scriveremo brevemente

$$X \sim \text{beta}(a, b),$$

se la sua distribuzione di probabilità è assolutamente continua, con una funzione di densità della forma

$$f(x) = \begin{cases} Kx^{a-1}(1-x)^{b-1} & \text{per } 0 < x < 1 \\ 0 & \text{altrimenti} \end{cases}.$$

Il valore della costante di normalizzazione

$$K = \left[\int_0^1 x^{a-1} (1-x)^{b-1} dx \right]^{-1}$$

è fornito dalla (44):

$$K = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}.$$

Otteniamo inoltre

$$\mathbb{E}(X) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \int_0^1 x^a (1-x)^{b-1} dx = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \frac{\Gamma(a+1)\Gamma(b)}{\Gamma(a+b+1)} = \frac{a}{a+b}.$$

Analogamente possiamo ottenere

$$\mathbb{E}(X^2) = \frac{a(a+1)}{(a+b)(a+b+1)},$$

da cui

$$Var(X) = \frac{ab}{(a+b)^2(a+b+1)}.$$

3.9 Altre famiglie notevoli di distribuzioni di probabilità per variabili non negative

In questo paragrafo vengono elencati alcuni altri tipi notevoli di distribuzioni di probabilità che, come nel caso delle distribuzioni gamma, sono concentrate sul semiasse positivo. Si tratta di distribuzioni di probabilità adatte a descrivere modelli probabilistici per variabili aleatorie il cui valore è non negativo con probabilità 1, quali ad esempi quelle con significato di *tempi d'attesa*, *tempi di durata* di apparecchiature, *reddito familiare annuo* di contribuenti, etc ...

3.9.1 Distribuzione di Weibull

La *distribuzione di Weibull* è caratterizzata da una funzione di distribuzione della forma

$$F(x) = \begin{cases} 0 & \text{per } x < 0 \\ 1 - \exp\{-\lambda x^\beta\} & \text{per } x > 0 \end{cases},$$

dove λ e β sono parametri positivi. Si ha, per $x > 0$,

$$f(x) = \frac{d}{dx} F(x) = \lambda \beta x^{\beta-1} \exp\{-\lambda x^\beta\},$$

$$r(x) = \frac{f(x)}{1 - F(x)} = \lambda \beta x^{\beta-1}.$$

E' interessante notare che $r(x)$ risulta essere una funzione monotona (crescente se $\beta > 1$, decrescente se $\beta < 1$). Ovviamente le distribuzioni esponenziali si possono vedere come quei casi particolari in cui $r(x)$ è costante.

Per quanto riguarda il valore atteso

$$\mathbb{E}(X) = \int_0^{+\infty} x \lambda \beta x^{\beta-1} \exp\{-\lambda x^\beta\} dx,$$

otteniamo, attraverso il cambiamento di variabile dato da $z = x^\beta$,

$$\mathbb{E}(X) = \int_0^{+\infty} \lambda \beta z \exp\{-\lambda z\} \frac{dz}{\beta z^{\frac{\beta-1}{\beta}}} = \int_0^{+\infty} \lambda z^{\frac{1}{\beta}} \exp\{-\lambda z\} = \frac{1}{\lambda^{\frac{1}{\beta}}} \Gamma\left(1 + \frac{1}{\beta}\right).$$

3.9.2 Distribuzione lognormale

Si tratta della distribuzione di probabilità di una variabile aleatoria X tale

$$Y = \log X$$

segue una distribuzione gaussiana. Quindi, per $x > 0$,

$$F_X(x) = P\{e^Y \leq x\} = P\{Y \leq \log x\} = F_Y(\log x);$$

$$f_X(x) = \frac{1}{x} f_Y(\log x) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2} \left(\frac{\log x - \mu}{\sigma}\right)^2\right\}.$$

3.9.3 Distribuzione di Pareto

Diciamo che X segue una distribuzione *di Pareto* con parametri a e x_0 se la sua distribuzione di probabilità è assolutamente continua, con una funzione di densità della forma

$$f(x) = \begin{cases} ax_0^a x^{-(a+1)} & \text{per } x > x_0 \\ 0 & \text{altrimenti} \end{cases}.$$

Se $a > 2$, risulta

$$\mathbb{E}(X) = \frac{ax_0^a}{a-1}, \text{Var}(X) = \frac{ax_0^a}{(a-1)^2(a-2)}.$$

4 Questioni circa vettori di variabili gaussiane

Variabili aleatorie con distribuzioni gaussiane sono alla base di numerosissimi modelli, risultati, problemi nell'ambito della teoria della probabilità, dei processi stocastici e della statistica matematica e la letteratura in proposito è vastissima. In quanto segue ci limitiamo a riportare pochissime questioni basilari. Ciò ci darà occasione di vedere, fra l'altro, alcune semplici applicazioni della Proposizione 7.1 del Capitolo 2 o, più in particolare, della Proposizione 5.1 dello stesso capitolo. Inizieremo con l'illustrare un metodo ben noto per generare variabili aleatorie gaussiane a partire da altre variabili indipendenti, con distribuzione uniforme sull'intervallo $[0, 1]$.

4.1 Generazione di variabili aleatorie gaussiane

In questo paragrafo la Proposizione 5.1 del Capitolo 2 verrà applicata al problema di generare variabili aleatorie gaussiane. Sia U_1, U_2 una coppia di variabili aleatorie la cui distribuzione congiunta è uniforme nel quadrato unitario $\Gamma^{(2)}$, cioè con funzione di densità congiunta

$$f_{U_1, U_2}(u_1, u_2) = \begin{cases} 1 & \text{per } 0 \leq u_1, u_2 \leq 1 \\ 0 & \text{altrove} \end{cases}.$$

In altre parole, U_1, U_2 sono variabili i.i.d. con distribuzione uniforme $R(0, 1)$. Consideriamo ora la trasformazione $\mathbf{T} : \Gamma^{(2)} \rightarrow \mathbb{R}^2$ (detta "Trasformazione di Box e Muller") definita dalle relazioni seguenti:

$$\begin{aligned} z_1(u_1, u_2) &= (-2 \log u_1)^{\frac{1}{2}} \cos(2\pi u_2), \\ z_2(u_1, u_2) &= (-2 \log u_1)^{\frac{1}{2}} \sin(2\pi u_2). \end{aligned}$$

Tale trasformazione è invertibile e la trasformazione inversa della \mathbf{T} è la trasformazione $\mathcal{U} : \mathbb{R}^2 \rightarrow \Gamma^{(2)}$ data da

$$\begin{aligned} u_1(z_1, z_2) &= \exp\left\{-\frac{z_1^2 + z_2^2}{2}\right\} \\ u_2(z_1, z_2) &= \frac{1}{2\pi} \operatorname{arctg} \frac{z_2}{z_1}. \end{aligned}$$

Lo jacobiano della trasformazione \mathcal{U} è dato da

$$\begin{aligned} J(z_1, z_2) &= \begin{vmatrix} -z_1 \exp\left\{-\frac{z_1^2 + z_2^2}{2}\right\} & -z_2 \exp\left\{-\frac{z_1^2 + z_2^2}{2}\right\} \\ \frac{-z_2/z_1}{2\pi \left(1 + \frac{z_2^2}{z_1^2}\right)} & \frac{-1/z_1}{2\pi \left(1 + \frac{z_2^2}{z_1^2}\right)} \end{vmatrix} = \\ &= \frac{1}{2\pi} \exp\left\{-\frac{z_1^2 + z_2^2}{2}\right\}. \end{aligned} \quad (49)$$

Consideriamo ora la coppia di variabili aleatorie

$$Z_1 = (-2 \log U_1)^{\frac{1}{2}} \cos(2\pi U_2), \quad Z_2 = (-2 \log U_1)^{\frac{1}{2}} \sin(2\pi U_2), \quad (50)$$

cioè $(Z_1, Z_2) = \mathbf{T}(U_1, U_2)$. In virtù della (49), (Z_1, Z_2) ammette una funzione di densità congiunta data da

$$f_{Z_1, Z_2}(z_1, z_2) = f_{U_1, U_2}(u_1(z_1, z_2), u_2(z_1, z_2)) \frac{1}{2\pi} \exp\left\{-\frac{z_1^2 + z_2^2}{2}\right\}.$$

Notando che, per qualunque scelta di $(z_1, z_2) \in \mathbb{R}^2$, $(u_1(z_1, z_2), u_2(z_1, z_2)) \in \Gamma^{(2)}$ e quindi che

$$f_{U_1, U_2}(u_1(z_1, z_2), u_2(z_1, z_2)) = 1,$$

otteniamo

$$f_{Z_1, Z_2}(z_1, z_2) = \frac{1}{2\pi} \exp\left\{-\frac{z_1^2 + z_2^2}{2}\right\}.$$

Possiamo concludere cioè che le variabili aleatorie Z_1, Z_2 definite dalla (50) hanno distribuzione gaussiana standard e sono indipendenti.

Accenniamo ad un motivo di interesse per questa trasformazione. Sia

$$\mathfrak{U} \equiv \{U_1, U_2, \dots\}$$

una successione di variabili aleatorie indipendenti identicamente distribuite, con distribuzione uniforme sull'intervallo $[0, 1]$.

Si presenta comunemente il problema di costruire, a partire da \mathfrak{U} e tramite semplici trasformazioni sulle variabili U_j , una successione di variabili aleatorie indipendenti con una funzione di distribuzione assegnata $F(x)$. A questo proposito torniamo su quanto già visto nel Paragrafo 1.3. Poniamo

$$\inf F = \sup\{x \in \mathbb{R} | F(x)\}, \sup F = \inf\{x \in \mathbb{R} | F(x) = 1\}$$

ed assumiamo che $F(x)$ sia continua e strettamente crescente nel dominio $I \equiv (\inf F, \sup F)$. Consideriamo inoltre la funzione $F^{-1} : [0, 1] \rightarrow I$ definita da

$$F^{-1}(u) = \begin{cases} \inf F & \text{se } u = 0 \\ x & \text{se } F(x) = u \\ \sup F & \text{se } u = 1 \end{cases}$$

e consideriamo le variabili aleatorie

$$X_1 = F^{-1}(U_1), X_2 = F^{-1}(U_2), \dots$$

X_1, X_2, \dots sono variabili aleatorie indipendenti identicamente distribuite, inoltre si ha

Proposizione 1.1

$$F_{X_j}(x) = F(x), \forall x \in \mathbb{R}.$$

Esempio 1.1

Sia U una variabile aleatoria con distribuzione uniforme sull'intervallo $[0, 1]$ e supponiamo di voler ottenere una variabile aleatoria con distribuzione esponenziale $\mathcal{E}(\lambda)$, con $\lambda > 0$. In tal caso si ha

$$\begin{aligned} F(x) &= [1 - \exp\{-\lambda x\}] \mathbf{1}_{\{x \geq 0\}} \\ \inf F &= 0, \sup F = \infty \\ F^{-1}(u) &= -\frac{1}{\lambda} \log(1 - u), 0 < u < 1. \end{aligned}$$

Dunque $X = -\frac{1}{\lambda} \log(1 - U)$ segue la distribuzione $\mathcal{E}(\lambda)$.

Nei casi in cui i valori della funzione F^{-1} possono essere calcolati soltanto mediante procedure numeriche, può risultare conveniente seguire metodi di costruzione alternativi a quello qui illustrato. Questo è appunto il caso, in particolare, quando si vogliono costruire variabili aleatorie con distribuzione gaussiana. Ciò sottolinea l'interesse della trasformazione di Box e Muller che permette di ricavare una coppia di variabili aleatorie indipendenti gaussiane standard a partire da una coppia di variabili aleatorie indipendenti distribuite uniformemente in $[0, 1]$.

4.2 Distribuzioni gaussiane in due dimensioni

Data una coppia di variabili aleatorie X, Y , vogliamo definire la condizione che la loro distribuzione congiunta sia una *gaussiana bidimensionale* e ne vogliamo illustrare alcune proprietà peculiari.

Definizione 2.1

Diremo che la coppia (X, Y) segue una *distribuzione gaussiana bidimensionale* di parametri $\mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2, \rho$, se essa ammette una funzione di densità congiunta e questa risulta essere della forma

$$f(x, y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}Q(x, y)\right\} \quad (51)$$

con

$$Q(x, y) \equiv \left(\frac{x - \mu_X}{\sigma_X}\right)^2 - 2\rho\left(\frac{x - \mu_X}{\sigma_X}\right)\left(\frac{y - \mu_Y}{\sigma_Y}\right) + \left(\frac{y - \mu_Y}{\sigma_Y}\right)^2.$$

Per indicare tale condizione scriveremo brevemente $(X, Y) \sim \mathcal{N}(\mu_X, \mu_Y; \sigma_X^2, \sigma_Y^2; \rho)$.

Prima di illustrare le proprietà implicate da tale condizione, si rendono opportune alcune semplici premesse. Concentriamo più in generale l'attenzione su coppie (X, Y) di variabili aleatorie dotate di valori attesi e di momenti del secondo ordine finiti. Tali condizioni ci garantiscono, in particolare, la possibilità

di considerare la covarianza $Cov(X, Y)$ ed il coefficiente di correlazione $\rho(X, Y)$ fra X e Y . Ricordiamo che vale l'identità

$$Cov(X, Y) = \mathbb{E}(X \cdot Y) - \mathbb{E}(X) \cdot \mathbb{E}(Y)$$

e che $\rho(X, Y)$ è definito dalla relazione

$$\rho(X, Y) := \frac{Cov(X, Y)}{\sqrt{Var(X)}\sqrt{Var(Y)}}.$$

Possiamo dunque scrivere anche

$$\mathbb{E}(X \cdot Y) = \rho(X, Y)\sqrt{Var(X)}\sqrt{Var(Y)} + \mathbb{E}(X) \cdot \mathbb{E}(Y). \quad (52)$$

Ammettiamo inoltre che la distribuzione congiunta ammetta una funzione di densità. Possiamo quindi parlare di densità condizionate, ad esempio, di Y dato ($X = x$) (per i valori x tali che $f_X(x) > 0$) e di *valori attesi condizionati* di Y dato ($X = x$):

$$\mathbb{E}(Y|X = x) = \int_{-\infty}^{+\infty} y f_Y(y|X = x) dy = \int_{-\infty}^{+\infty} y \frac{f(x, y)}{f_X(x)} dy. \quad (53)$$

La funzione della variabile x , definita da

$$\phi_Y(x) \equiv \mathbb{E}(Y|X = x)$$

ha un ruolo notevole, in particolare nella statistica, e viene detta *funzione di regressione* di Y rispetto alla variabile X .

Nel caso particolare in cui $\phi_Y(x)$ risulti essere della forma

$$\phi_Y(x) = ax + b \quad (54)$$

si parlerà di *retta di regressione* (ovviamente i ruoli delle due variabili X e Y potrebbero essere scambiati).

Il seguente risultato mette in luce alcune specifiche caratteristiche di distribuzioni bidimensionali in cui valga la relazione (54).

Proposizione 2.1

Supponiamo che la distribuzione congiunta di (X, Y) sia tale che valga la relazione (54). Allora risulta

i)

$$\mathbb{E}(Y) = a\mathbb{E}(X) + b, \quad (55)$$

ii)

$$\rho(X, Y) = a \frac{\sqrt{Var(X)}}{\sqrt{Var(Y)}}, \quad (56)$$

Dimostrazione

Osserviamo innanzitutto che, tenendo conto della (53), la (54) è equivalente

a

$$\int_{-\infty}^{+\infty} yf(x, y)dy = f_X(x)(ax + b) \quad (57)$$

i) si dimostra immediatamente integrando da $-\infty$ a $+\infty$ entrambi i membri della identità (57), rispetto alla variabile x .

Per quanto riguarda ii), moltiplichiamo per x entrambi i membri della identità (57) e successivamente integriamo da $-\infty$ a $+\infty$, ancora rispetto alla variabile x . In tal modo otteniamo

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} xyf(x, y)dy = \int_{-\infty}^{+\infty} ax^2 f_X(x)dx + b \int_{-\infty}^{+\infty} xf_X(x)dx$$

cioè

$$\mathbb{E}(X \cdot Y) = a\mathbb{E}(X^2) + b\mathbb{E}(X).$$

Ricordando la (52), quest'ultima relazione può essere anche riscritta come

$$\rho(X, Y)\sqrt{Var(X) \cdot Var(Y)} + \mathbb{E}(X)\mathbb{E}(Y) = aVar(X) + a(\mathbb{E}(X))^2 + b\mathbb{E}(X),$$

da cui ii) si ottiene immediatamente tenendo conto della (55).

Notiamo dunque che i coefficienti a e b nella relazione (54) hanno il seguente significato probabilistico:

$$a = \rho(X, Y) \frac{\sqrt{Var(Y)}}{\sqrt{Var(X)}}, \quad b = \mathbb{E}(Y) - \rho(X, Y) \frac{\sqrt{Var(Y)}}{\sqrt{Var(X)}} \mathbb{E}(X)$$

e che quindi possiamo anche scrivere

$$\phi_Y(x) = \mathbb{E}(Y) + \rho(X, Y) \frac{\sqrt{Var(Y)}}{\sqrt{Var(X)}} (x - \mathbb{E}(X)).$$

Osservazione 2.1

Notiamo che un caso particolarissimo in cui è soddisfatta la condizione (54), ma in cui la distribuzione congiunta non è assolutamente continua, è quando si ha, quasi certamente,

$$Y = aX + b.$$

In tal caso risulta $\rho(X, Y) = \pm 1$, a seconda del segno del coefficiente a .

Torniamo ora ad analizzare le proprietà delle distribuzioni gaussiane bidimensionali.

Proposizione 2.2

Supponiamo che (X, Y) segua una *distribuzione gaussiana bidimensionale* di parametri $\mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2, \rho$. Allora

- i) La densità marginale di X è una gaussiana di parametri μ_X, σ_X^2
La densità marginale di Y è una gaussiana di parametri μ_Y, σ_Y^2
- ii) La densità condizionata di Y , dato ($X = x$) è una gaussiana con valore atteso $\mu_Y + \rho \frac{\sigma_Y}{\sigma_X}(x - \mu_X)$ e varianza $\sigma_Y^2(1 - \rho^2)$
- iii) $\rho(X, Y) = \rho$

Dimostrazione

Osserviamo innanzitutto che vale la decomposizione

$$Q(x, y) = \left[\left(\frac{y - \mu_Y}{\sigma_Y} \right) + \rho \left(\frac{x - \mu_X}{\sigma_X} \right) \right]^2 + (1 - \rho^2) \left(\frac{x - \mu_X}{\sigma_X} \right)^2.$$

Possiamo anche scrivere, quindi,

$$Q(x, y) = \left(\frac{y - \beta(x)}{\sigma_Y} \right)^2 + (1 - \rho^2) \left(\frac{x - \mu_X}{\sigma_X} \right)^2$$

avendo posto

$$\beta(x) := \mu_Y + \rho \frac{\sigma_Y}{\sigma_X}(x - \mu_X).$$

- i) Per ogni fissato $x \in \mathbb{R}$, si ha

$$\begin{aligned} f_X(x) &= \int_{-\infty}^{+\infty} f(x, y) dy = \\ &= \frac{1}{\sigma_X \sqrt{2\pi}} \exp\left\{-\frac{1}{2\sigma_X^2}(x - \mu_X)^2\right\} \int_{-\infty}^{+\infty} \frac{1}{\sigma_Y \sqrt{2\pi} \sqrt{1 - \rho^2}} \exp\left\{-\frac{1}{2\sigma_Y^2(1 - \rho^2)}(y - \beta(x))^2\right\} dy \end{aligned}$$

Notiamo ora che, per $x \in \mathbb{R}$, la funzione integranda che compare sotto il segno di integrale nel membro a destra è l'espressione di una densità gaussiana (di parametri $\beta(x)$ e $\sigma_Y^2(1 - \rho^2)$) e quindi l'integrale stesso è uguale ad 1; si ottiene dunque

$$f_X(x) = \frac{1}{\sigma_X \sqrt{2\pi}} \exp\left\{-\frac{1}{2\sigma_X^2}(x - \mu_X)^2\right\}.$$

Tramite considerazioni analoghe possiamo verificare che la distribuzione marginale della variabile Y è $\mathcal{N}(\mu_Y, \sigma_Y^2)$.

- ii) Tenendo conto del precedente punto i), otteniamo subito

$$\begin{aligned} f_Y(y|X = x) &= \frac{f(x, y)}{f_X(x)} = \\ &= \frac{1}{\sigma_Y \sqrt{2\pi} \sqrt{1 - \rho^2}} \exp\left\{-\frac{1}{2\sigma_Y^2(1 - \rho^2)}(y - \beta(x))^2\right\}. \end{aligned}$$

Analoghe considerazioni possono essere sviluppate per verificare che la distribuzione condizionata di X dato ($Y = y$) è gaussiana e per determinarne i relativi parametri.

iii) Dal precedente punto ii), a sua volta si ricava in particolare che la distribuzione congiunta di X e Y è tale che vale la relazione (54) in quanto $\mathbb{E}(Y|X = x)$ coincide con $\beta(x)$ e cioè

$$\mathbb{E}(Y|X = x) = \mu_Y + \rho \frac{\sigma_Y}{\sigma_X} (x - \mu_X).$$

Dunque, dalla (56), otteniamo

$$\rho(X, Y) = \rho.$$

Osservazione 2.2

La funzione non negativa $f(x, y)$ definita nella (51) risulta effettivamente essere una funzione di densità congiunta. Per verificare tale proprietà basta mostrare che risulta

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

Tale identità segue immediatamente da quanto già sviluppato in merito al punto i) nella precedente dimostrazione. Possiamo infatti scrivere

$$\begin{aligned} & \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y) dx dy = \\ &= \int_{-\infty}^{+\infty} \frac{1}{\sigma_X \sqrt{2\pi}} \exp\left\{-\frac{1}{2\sigma_X^2} (x - \mu_X)^2\right\} dx \int_{-\infty}^{+\infty} \frac{1}{\sigma_Y \sqrt{2\pi} \sqrt{1 - \rho^2}} \exp\left\{-\frac{1}{2\sigma_Y^2 (1 - \rho^2)} (y - \beta)^2\right\} dy \\ &= \int_{-\infty}^{+\infty} \frac{1}{\sigma_X \sqrt{2\pi}} \exp\left\{-\frac{1}{2\sigma_X^2} (x - \mu_X)^2\right\} dx = 1. \end{aligned}$$

Osservazione 2.3

In generale la condizione di *non correlazione*, cioè $\rho(X, Y) = 0$, è condizione necessaria ma non sufficiente per l'indipendenza stocastica fra due variabili aleatorie X ed Y . Invece nel caso di distribuzione congiunta gaussiana tale condizione, eliminando il termine misto nelle due variabili x, y , fa sì che la funzione di densità congiunta in (51) risulti uguale al prodotto di una funzione della sola variabile x per una funzione della sola variabile y e dunque comporta l'indipendenza stocastica fra X ed Y .

Osservazione 2.4

Assumiamo una distribuzione congiunta gaussiana bidimensionale per le variabili aleatorie X, Y e consideriamo una nuova coppia di variabili aleatorie X' e Y' definite come combinazioni lineari di X ed Y :

$$X' = a_{11}X + a_{12}Y, Y' = a_{21}X + a_{22}Y$$

(supponiamo invertibile la matrice $A = (a_{ij})$).

Una diretta applicazione della Proposizione 5.1 del Capitolo 2 mostra che anche la coppia (X', Y') segue una distribuzione congiunta gaussiana bidimensionale. I parametri di tale distribuzione possono essere ricavati direttamente in base alle regole di calcolo dei valori attesi, delle varianze e dei coefficienti di correlazione.

Osservazione 2.5

Nella precedente osservazione si è supposto che la matrice A sia invertibile. Notiamo a questo proposito che, se esistesse una relazione di proporzionalità fra le sue righe $(a_{11}, a_{1,2})$ e $(a_{21}, a_{2,2})$ si avrebbe, con probabilità 1, Y' uguale a X' moltiplicata per una costante. Quindi non potrebbe esistere una funzione di densità congiunta per la coppia (X', Y') e quindi non verrebbe mantenuta, per la nuova coppia di variabili, la proprietà di avere ancora una distribuzione congiunta gaussiana bidimensionale.

Osservazione 2.6

Sappiamo che la somma di due variabili aleatorie gaussiane indipendenti segue ancora una distribuzione gaussiana. Come particolare conseguenza di quanto sin qui detto, possiamo dedurre che anche una variabile aleatoria (scalare), ottenuta come combinazione lineare di una coppia di variabili aleatorie X, Y , non necessariamente indipendenti, segue una distribuzione gaussiana, purchè la distribuzione congiunta della coppia (X, Y) sia una gaussiana bidimensionale.

Facciamo d'altra parte attenzione al fatto che, come si potrebbe mostrare con opportuni controesempi, una combinazione lineare di variabili aleatorie (marginalmente) gaussiane non indipendenti (e la cui distribuzione congiunta non sia una gaussiana bidimensionale) non è necessariamente una variabile aleatoria gaussiana.

Osservazione 2.7

Siano Z_1, Z_2 variabili aleatorie gaussiane standard ed indipendenti e $\mu_X, \mu_Y; \sigma_X^2, \sigma_Y^2; \rho$ valori assegnati. Fissando opportunamente dei coefficienti $b_{11}, b_{12}, b_{21}, b_{22}, c_1, c_2$, variabili aleatorie X, Y con distribuzione congiunta gaussiana bidimensionale $\mathcal{N}(\mu_X, \mu_Y; \sigma_X^2, \sigma_Y^2; \rho)$ possono essere ottenute tramite una trasformazione della coppia (Z_1, Z_2) ponendo

$$X = b_{11}Z_1 + b_{12}Z_2 + c_1,$$

$$Y = b_{21}Z_1 + b_{22}Z_2 + c_2.$$

In virtù di quanto visto nel paragrafo precedente, possiamo quindi concludere che la coppia (X, Y) può essere costruita a partire da due variabili U_1, U_2 indipendenti e con distribuzione uniforme $R(0, 1)$.

Osservazione 2.8

Estendendo in modo diretto gli argomenti svolti in questo paragrafo si può ottenere la definizione di *distribuzione gaussiana in n dimensioni* ($n > 2$). Basta partire da una n -upla di variabili aleatorie gaussiane standard indipendenti Z_1, \dots, Z_n e costruire un nuovo vettore aleatorio (X_1, \dots, X_n) attraverso la trasformazione

$$\begin{aligned} X_1 &= \mu_1 + \sum_{j=1}^n a_{1,j} \\ &\dots \\ X_n &= \mu_n + \sum_{j=1}^n a_{n,j}, \end{aligned}$$

dove $A \equiv (a_{i,j})$ è una matrice $n \times n$ invertibile e μ_1, \dots, μ_n sono numeri reali. E' evidente che μ_1, \dots, μ_n danno luogo, rispettivamente, ai valori attesi delle variabili X_1, \dots, X_n . Inoltre le varianze delle X_1, \dots, X_n e le covarianze delle coppie (X_h, X_k) si ottengono semplicemente in termini della A . Viceversa, anche A si ottiene dalla conoscenza della matrice di *varianza-covarianza* delle X_1, \dots, X_n . Possiamo cioè affermare che una distribuzione gaussiana n -dimensionale è determinata dal vettore dei valori attesi e dalla matrice di varianza-covarianza.

Una tale distribuzione gode di proprietà analoghe a quelle viste per il caso $n = 2$: tutte le distribuzioni marginali (di qualunque dimensione $1 \leq m < n$) sono gaussiane. Così pure tutte le distribuzioni condizionate di alcune variabili date altre variabili sono gaussiane, e le funzioni di regressione sono funzioni di primo grado nei valori delle variabili condizionanti. Trasformazioni lineari invertibili di vettori aleatori gaussiani, inoltre, danno ancora luogo a vettori aleatori gaussiani.

4.3 Trasformazioni ortogonali di campioni gaussiani e teorema di Cochran

Sia $A \equiv (a_{ij})$ una matrice $n \times n$ e $\mathbf{T} : \mathbb{R}^n \rightarrow \mathbb{R}^n$ la trasformazione lineare ad essa associata: per $\mathbf{x} \in \mathbb{R}^n$

$$\mathbf{T}(\mathbf{x}) = \mathbf{A}\mathbf{x}.$$

Supponiamo ora, in particolare, che \mathbf{A} sia *ortogonale*:

$$\sum_{j=1}^n a_{hj}a_{kj} = 0, \text{ per } 1 \leq h \neq k \leq n$$

$$\sum_{j=1}^n a_{hj}^2 = 1, h = 1, 2, \dots, n.$$

Ne segue

$$A^{-1} = A^*,$$

$$|\det(A)| = |\det(A^*)| = |\det(A^{-1})| = 1,$$

indicando con A^* la matrice trasposta della A . La trasformazione inversa della \mathbf{T} è data da

$$\mathbf{T}^{-1}(\mathbf{y}) = A^*\mathbf{y}.$$

Più esplicitamente, le componenti della \mathbf{T}^{-1} sono

$$T_j^{-1}(y_1, \dots, y_n) = \sum_{i=1}^n a_{ji}y_i.$$

Anche tale trasformazione risulta ortogonale e, per quanto riguarda il modulo del relativo jacobiano, si ha

$$|J_{\mathbf{T}^{-1}}(\mathbf{y})| = |\det(\mathbf{A}^{-1})| = 1, \forall \mathbf{y} \in \mathbb{R}^n. \quad (58)$$

Inoltre si ha che la trasformazione \mathbf{T}^{-1} conserva le distanze; si ha cioè

Lemma 3.1

$$\sum_{j=1}^n [T_j^{-1}(y_1, \dots, y_n)]^2 = \sum_{i=1}^n y_i^2.$$

Pur se questa proprietà risulta ben nota, per comodità del lettore ne forniamo qui di seguito una verifica diretta.

Dimostrazione

$$\begin{aligned} \sum_{j=1}^n [T_j^{-1}(y_1, \dots, y_n)]^2 &= \sum_{j=1}^n \left[\sum_{i=1}^n a_{ji}y_i \right]^2 \\ &= \sum_{j=1}^n \left[\sum_{i=1}^n a_{ji}^2 y_i^2 + \sum_{i_1 \neq i_2} (a_{j i_1} a_{j i_2}) (y_{i_1} y_{i_2}) \right] = \end{aligned}$$

$$\sum_{j=1}^n \left(\sum_{i=1}^n a_{ji}^2 \right) y_i^2 + \sum_{i_1 \neq i_2} \left(\sum_{j=1}^n a_{i_1 j} a_{i_2 j} \right) (y_{i_1} y_{i_2}) = \sum_{i=1}^n y_i^2.$$

Oppure, in forma compatta,

$$\sum_{i=1}^n y_i^2 = \mathbf{y}^* \mathbf{y} = (\mathbf{A}\mathbf{x})^* (\mathbf{A}\mathbf{x}) = \mathbf{x}^* \mathbf{A}^* \mathbf{A} \mathbf{x} = \mathbf{x}^* \mathbf{x} = \sum_{j=1}^n x_j^2 = \sum_{j=1}^n [T_j^{-1}(y_1, \dots, y_n)]^2.$$

Consideriamo ora n variabili aleatorie X_1, \dots, X_n , indipendenti con identica distribuzione $\mathcal{N}(0, \sigma^2)$ e consideriamo le variabili Y_1, \dots, Y_n ottenute applicando la trasformazione \mathbf{T} al vettore $\mathbf{X} \equiv (X_1, \dots, X_n)$:

$$\mathbf{Y} = \mathbf{A}\mathbf{X}$$

Proposizione 3.2

Anche Y_1, \dots, Y_n sono variabili indipendenti con distribuzione $\mathcal{N}(0, \sigma^2)$.

Dimostrazione

Applicando la Proposizione 7.1 del Cap. 1, e tenendo conto della (58), si ottiene

$$f_{\mathbf{Y}}(\mathbf{y}) = f_{\mathbf{X}}(T_1^{-1}(\mathbf{y}), \dots, T_n^{-1}(\mathbf{y})) |J_{\mathbf{T}^{-1}}(\mathbf{y})| = \left(\frac{1}{2\pi\sigma^2} \right)^{\frac{n}{2}} \exp\left\{ -\frac{1}{2\sigma^2} \sum_{j=1}^n [T_j^{-1}(\mathbf{y})]^2 \right\}.$$

Ricordando quindi il precedente Lemma 3.1, otteniamo

$$f_{\mathbf{Y}}(\mathbf{y}) = \left(\frac{1}{2\pi\sigma^2} \right)^{\frac{n}{2}} \exp\left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n y_i^2 \right\}$$

e quest'ultima coincide proprio con la funzione di densità congiunta di n variabili aleatorie Y_1, \dots, Y_n indipendenti con identica distribuzione $\mathcal{N}(0, \sigma^2)$.

Corollario 3.3

Sia \mathbf{A} una matrice ortogonale e X_1, \dots, X_n variabili aleatorie indipendenti e con

$$X_i \sim \mathcal{N}(\mu_i, \sigma^2).$$

Ponendo

$$\mathbf{Y} \equiv (Y_1, \dots, Y_n) = \mathbf{A}\mathbf{X}$$

otteniamo un vettore di variabili aleatorie indipendenti

$$Y_i \sim \mathcal{N}\left(\sum_{j=1}^n a_{ij}\mu_j, \sigma^2\right)$$

Dimostrazione

Ponendo, per $j = 1, \dots, n$,

$$\varepsilon_j = X_j - \mu_j,$$

otteniamo delle variabili aleatorie $\varepsilon_1, \dots, \varepsilon_n$ indipendenti ed identicamente distribuite, con distribuzione $\mathcal{N}(0, \sigma^2)$; quindi anche $A \cdot \varepsilon$ è un vettore di variabili aleatorie indipendenti ed identicamente distribuite, con distribuzione $\mathcal{N}(0, \sigma^2)$.

La conclusione si ottiene osservando che

$$\mathbf{Y} = A \cdot \mathbf{X} = A \cdot \varepsilon + A\mu$$

Come particolare conseguenza della precedente Proposizione 3.2, otteniamo anche il seguente

Corollario 3.4

Una combinazione lineare di variabili aleatorie gaussiane indipendenti segue una distribuzione gaussiana.

Dimostrazione

Siano X_1, \dots, X_n variabili aleatorie indipendenti con $X_j \sim \mathcal{N}(\mu_j, s_j^2)$, e siano c_1, \dots, c_n delle costanti assegnate; poniamo

$$Y = \sum_{j=1}^n c_j X_j.$$

Come sappiamo, possiamo scrivere

$$X_j = \mu_j + s_j Z_j$$

dove Z_1, \dots, Z_n sono variabili aleatorie gaussiane standard, ed anche esse indipendenti. Dunque

$$Y = \sum_{j=1}^n c_j \mu_j + \sum_{j=1}^n c_j s_j Z_j$$

e risulta

$$\mathbb{E}(Y) = \sum_{j=1}^n c_j \mu_j \equiv \mu,$$

$$Var(Y) = \sum_{j=1}^n c_j^2 s_j^2 \equiv \sigma^2.$$

Ponendo

$$\alpha_j \equiv \frac{c_j s_j}{\sigma}, R_j = \sigma (X_j - \mu_j),$$

possiamo scrivere

$$Y - \mu = \sum_{j=1}^n c_j s_j Z_j = \sum_{j=1}^n \alpha_j R_j.$$

Ora R_1, \dots, R_n sono variabili aleatorie i.i.d. con $R_j \sim \mathcal{N}(0, \sigma^2)$, dove $\sum_{j=1}^n \alpha_j^2 = 1$. Dunque $(\alpha_1, \dots, \alpha_n)$ può essere vista come la prima riga di una trasformazione ortogonale e $Y - \mu$ può essere vista come la prima coordinata di un vettore di variabili aleatorie indipendenti identicamente distribuite con distribuzione $\mathcal{N}(0, \sigma^2)$.

Osservazione 3.1

E' utile innanzitutto confrontare il Corollario 3.4 con quanto notato nell'Osservazione 2.6 del precedente paragrafo: lì si trattava il caso $n = 2$, mentre qui abbiamo il caso di un generico $n \geq 2$. D'altra parte qui abbiamo considerato il caso di variabili indipendenti, mentre lì era stato considerato il caso di $n = 2$ variabili non necessariamente indipendenti, purchè con distribuzione congiunta gaussiana bivariata. In entrambe le situazioni, comunque, si ha a che fare con casi particolari di risultati inerenti distribuzioni congiunte gaussiane multivariate. Ricordando l'Osservazione 2.8, vediamo che una variabile aleatoria ottenuta come combinazione lineare di variabili aleatorie gaussiane risulta essa stessa gaussiana se si assume che la loro distribuzione congiunta sia gaussiana.

Veniamo ora al *Teorema di Cochran*. Si tratta di un risultato, di interesse fondamentalmente statistico, che possiamo ottenere come speciale conseguenza degli argomenti fin qui svolti. Un cenno ai motivi per i quali esso può avere interesse nella Statistica Matematica verrà presentato nel paragrafo successivo.

A questo proposito, date n variabili aleatorie X_1, \dots, X_n , poniamo

$$\bar{X} \equiv \frac{\sum_{j=1}^n X_j}{n}, S^2 \equiv \sum_{j=1}^n (X_j - \bar{X})^2.$$

Proposizione 3.5 (Teorema di Cochran)

Siano X_1, \dots, X_n variabili indipendenti con identica distribuzione $\mathcal{N}(\mu, \sigma^2)$. Risulta allora

$$\frac{S^2}{\sigma^2} \sim \text{chi-quadro}(n), \frac{\bar{X} - \mu}{\sqrt{\frac{S^2}{n(n-1)}}} \sim t_n.$$

Dimostrazione

Consideriamo la trasformazione ortogonale definita dalla matrice $A = (a_{ij})$, con elementi definiti come segue:

$$a_{1j} = \frac{\sqrt{n}}{n}, j = 1, 2, \dots, n$$

e, per $i > 1$,

$$a_{ij} = \frac{-1}{\sqrt{n(n-1)}}, j = 1, 2, \dots, i-1$$

$$a_{ii} = \frac{i-1}{\sqrt{n(n-1)}}$$

$$a_{ij} = 0, j = i + 1, \dots, n.$$

Poniamo ora

$$\mathbf{Y} = \mathbf{A}\mathbf{X}$$

ed osserviamo che, in virtù del Corollario 3.3, risulta

$$Y_1 = \sqrt{n}\bar{X} \sim \mathcal{N}(\sqrt{n}\mu, \sigma^2)$$

$$Y_2 = \frac{\sqrt{2}}{2} (X_2 - X_1) \sim \mathcal{N}(0, \sigma^2)$$

$$Y_3 = \frac{1}{\sqrt{6}} (2X_3 - X_1 - X_2) \sim \mathcal{N}(0, \sigma^2)$$

.....

$$Y_n = \frac{(n-1)X_n - X_1 - X_2 - \dots - X_{n-1}}{\sqrt{n(n-1)}} \sim \mathcal{N}(0, \sigma^2).$$

Y_1, Y_2, \dots, Y_n sono stocasticamente indipendenti. Dunque $\frac{Y_2}{\sigma}, \frac{Y_3}{\sigma}, \dots, \frac{Y_n}{\sigma}$ sono variabili stocasticamente indipendenti con distribuzione gaussiana standard. Quindi, in particolare, risulta

$$\frac{\sum_{i=2}^n Y_i^2}{\sigma^2} = \left(\frac{Y_2}{\sigma}\right)^2 + \left(\frac{Y_3}{\sigma}\right)^2 + \dots + \left(\frac{Y_n}{\sigma}\right)^2 \sim \text{chi-quadro}(n-1). \quad (59)$$

Osserviamo ora che, essendo A ortogonale ed in virtù del Lemma 3.1, risulta

$$\sum_{i=1}^n Y_i^2 = \sum_{j=1}^n X_j^2.$$

da cui

$$S^2 = \sum_{j=1}^n (X_j - \bar{X})^2 = \sum_{j=1}^n X_j^2 - n\bar{X}^2 = \sum_{j=1}^n X_j^2 - Y_1^2 = \sum_{i=2}^n Y_i^2 \quad (60)$$

Dalla (59), segue quindi

$$\frac{S^2}{\sigma^2} \sim \text{chi-quadro}(n-1).$$

Dall'indipendenza stocastica fra le coordinate di \mathbf{Y} e dalla (60) segue che le variabili aleatorie \bar{X} ed S^2 sono indipendenti.

La variabile aleatoria $\frac{\bar{X} - \mu}{\sqrt{\frac{S^2}{n(n-1)}}}$ segue quindi una distribuzione t di Student con $(n-1)$ gradi di libertà.

4.4 Un'applicazione statistica del Teorema di Cochran

Dato un valore $\hat{\sigma}^2 > 0$, consideriamo n variabili aleatorie X_1, \dots, X_n , indipendenti con identica distribuzione gaussiana $\mathcal{N}(\mu, \hat{\sigma}^2)$. Stiamo usando qui il simbolo $\hat{\sigma}^2$ per mettere in evidenza che tali variabili hanno una varianza nota. Pensiamo invece che μ sia un parametro incognito e che si ponga quindi il problema statistico di "stimare" tale parametro sulla base dell'osservazione dei valori assunti da X_1, \dots, X_n . Considerazioni di tipo statistico suggeriscono che una buona "stima" di μ , sulla base dell'osservazione di una n -upla di valori x_1, \dots, x_n , sia la loro media aritmetica

$$\bar{x} = \frac{1}{n} \sum_{h=1}^n x_h.$$

Si pone a questo punto il problema di valutare la *precisione* di tale stima. Sappiamo che la variabile aleatoria

$$\bar{X} = \frac{1}{n} \sum_{h=1}^n X_h$$

segue, qualunque sia il valore di μ , una distribuzione $\mathcal{N}(\mu, \frac{\hat{\sigma}^2}{n})$. Dunque la variabile aleatoria standardizzata

$$Z = \frac{\bar{X} - \mu}{\hat{\sigma}} \sqrt{n}$$

segue una distribuzione gaussiana standard. Per ogni $\varepsilon > 0$ fissato, indicando brevemente con il simbolo γ la quantità $2\Phi(\varepsilon) - 1$, possiamo anche scrivere

$$P\{-\varepsilon \leq \frac{\bar{X} - \mu}{\hat{\sigma}} \sqrt{n} \leq \varepsilon\} = \gamma$$

oppure

$$\mathbb{P}\{\bar{X} - \frac{\varepsilon \hat{\sigma}}{\sqrt{n}} \leq \mu \leq \bar{X} + \frac{\varepsilon \hat{\sigma}}{\sqrt{n}}\} = \gamma. \quad (61)$$

Notiamo che γ non dipende dal valore del parametro μ e, per ottenere un valore prefissato per γ , basta porre

$$\varepsilon = \Phi^{-1}\left(\frac{1+\gamma}{2}\right).$$

Diremo allora che l'intervallo aleatorio

$$\left[\bar{X} - \hat{\sigma} \frac{\Phi^{-1}\left(\frac{1+\gamma}{2}\right)}{\sqrt{n}}, \bar{X} + \hat{\sigma} \frac{\Phi^{-1}\left(\frac{1+\gamma}{2}\right)}{\sqrt{n}} \right] \quad (62)$$

costituisce un *intervallo di fiducia a livello* γ per il parametro μ . Cioè, qualunque sia il valore di μ , vi è una probabilità costante γ che le due variabili aleatorie

$\bar{X} - \hat{\sigma} \frac{\Phi^{-1}\left(\frac{1+\gamma}{2}\right)}{\sqrt{n}}$ e $\bar{X} + \hat{\sigma} \frac{\Phi^{-1}\left(\frac{1+\gamma}{2}\right)}{\sqrt{n}}$ cadano rispettivamente a sinistra e a destra del valore μ .

Osservazione 4.1

Notiamo che l’assegnazione di probabilità in (61) riguarda le variabili aleatorie X_1, \dots, X_n . Non si tratta, cioè, di un’assegnazione di probabilità sul parametro μ , che, sebbene *incognito*, non viene qui visto come una variabile aleatoria.

Notiamo inoltre che la costruzione dell’intervallo in (62) di fiducia è basata sulla conoscenza della varianza $\hat{\sigma}^2$.

Ci chiediamo ora come sia possibile costruire un intervallo di fiducia per μ , nel caso in cui le variabili statistiche (osservabili) X_1, \dots, X_n sono indipendenti con identica distribuzione $\mathcal{N}(\mu, \sigma^2)$ con i parametri μ e σ^2 entrambi incogniti. Vorremmo cioè determinare, comunque fissato un *livello di fiducia* γ ($0 < \gamma < 1$) una coppia di funzioni $\alpha_\gamma, \beta_\gamma : \mathbb{R}^n \rightarrow \mathbb{R}$ tali che

$$\alpha_\gamma(x_1, \dots, x_n) \leq \beta_\gamma(x_1, \dots, x_n)$$

e, per ogni valore reale μ ,

$$\mathbb{P}\{\alpha_\gamma(X_1, \dots, X_n) \leq \mu \leq \beta_\gamma(X_1, \dots, X_n)\} = \gamma. \tag{63}$$

Dobbiamo notare che l’esistenza di funzioni $\alpha_\gamma, \beta_\gamma$ che soddisfino la (63) non è a priori scontata in quanto la distribuzione congiunta di (X_1, \dots, X_n) dipende dal valore di μ e quindi anche la distribuzione congiunta di variabili aleatorie $\alpha_\gamma(X_1, \dots, X_n)$ e $\beta_\gamma(X_1, \dots, X_n)$ dipenderà in generale da μ .

La soluzione a tale problema è però fornita dalla precedente Proposizione 3.5. Quest’ultima infatti ci assicura che, ponendo per brevità

$$S^2 := \sum_{h=1}^n (X_h - \bar{X})^2,$$

la variabile aleatoria

$$\hat{Z} := \frac{\bar{X} - \mu}{\sqrt{\frac{S^2}{n-1}}} \sqrt{n}$$

segue una distribuzione del t di Student con $(n - 1)$ gradi di libertà. Indicando con $F_{(n-1)}$ la funzione di ripartizione di tale distribuzione di probabilità e ragionando in modo simile a quanto fatto sopra per il caso di “varianza nota”, possiamo ottenere la seguente conclusione, analoga alla (61).

Proposizione 4.1

Siano X_1, \dots, X_n variabili aleatorie indipendenti con identica distribuzione $\mathcal{N}(\mu, \sigma^2)$. Risulta, indipendentemente dai valori $\mu \in \mathbb{R}$, $\sigma^2 > 0$ e $\gamma \in (0, 1)$,

$$\mathbb{P}\left\{\bar{X} - \sqrt{\frac{S^2}{n-1}} \frac{F_{(n-1)}^{-1}\left(\frac{1+\gamma}{2}\right)}{\sqrt{n}} \leq \mu \leq \bar{X} + \sqrt{\frac{S^2}{n-1}} \frac{F_{(n-1)}^{-1}\left(\frac{1+\gamma}{2}\right)}{\sqrt{n}}\right\} = \gamma.$$

5 Il metodo della Funzione Caratteristica e il Teorema Centrale del Limite per variabili aleatorie i.i.d.

In questo capitolo verrà introdotta la nozione di *Funzione Caratteristica* associata ad una distribuzione di probabilità sulla retta e ne verrà illustrata l'applicazione allo studio del comportamento asintotico delle medie aritmetiche di variabili i.i.d.. Si vedranno in particolare la linea della dimostrazione del Teorema di Lindberg-Lévy ed i motivi di fondo circa il ruolo che, in tale risultato, vi gioca la distribuzione gaussiana standard; tale ruolo è connesso con la *proprietà di stabilità rispetto alla somma*. Il capitolo inizia presentando la definizione di funzione caratteristica e le proprietà che da tale definizione seguono immediatamente. Nella conclusione verranno riportate alcune considerazioni relative alle distribuzioni *con code pesanti* e alla proprietà di stabilità in casi diversi da quello gaussiano.

5.1 Definizione di Funzione Caratteristica e sue proprietà elementari

Iniziamo ricordando innanzitutto gli sviluppi in serie di potenze delle funzioni esponenziale, seno e coseno, nel campo reale: per $x \in \mathbb{R}$, valgono le identità

$$\begin{aligned}e^x &= \sum_{k=0}^{\infty} \frac{x^k}{k!}, \\ \sin x &= \sum_{k=0}^{\infty} \frac{(-1)^k}{(2k+1)!} x^{2k+1} \\ \cos x &= \sum_{k=0}^{\infty} \frac{(-1)^k}{(2k)!} x^{2k}.\end{aligned}$$

Passando al campo complesso, per $y \in \mathbb{R}$ ed essendo i l'unità immaginaria ($i^2 = -1, i^3 = -i, i^4 = 1$), si pone

$$e^{i \cdot x} = \sum_{k=0}^{\infty} \frac{(i \cdot x)^k}{k!}. \quad (64)$$

Ne risulta che e^{ix} è il numero complesso (di modulo 1) dato da

$$e^{iy} = \cos y + i \sin y. \quad (65)$$

Risulta inoltre che, per $y_1, y_2 \in \mathbb{R}$, vale l'identità

$$e^{i(y_1+y_2)} = e^{i \cdot y_1} \cdot e^{i \cdot y_2}. \quad (66)$$

Sia ora X una variabile aleatoria con distribuzione di probabilità descritta dalla funzione di ripartizione $F(x)$

Definizione 1.1

La *funzione caratteristica* di X è la funzione $\varphi_X : \mathbb{R} \rightarrow \mathbb{C}$ definita dalla posizione

$$\varphi_X(t) = \int_{\mathbb{R}} \exp\{i \cdot t \cdot x\} dF(x). \quad (67)$$

Osservazione 1.1

Risulta conveniente, per diversi motivi, parlare di funzione caratteristica associata ad una variabile aleatoria X ed usare il simbolo φ_X . Si deve notare comunque che la φ_X risulta associata alla funzione di ripartizione F piuttosto che alla X : tutte le variabili aleatorie che condividono la stessa distribuzione di probabilità marginale ammettono anche la stessa funzione caratteristica, indipendentemente dallo spazio di probabilità su cui sono definite e da *come* esse siano definite quali funzioni sul relativo spazio.

Osservazione 1.2

Nella formula (67) si è utilizzata la notazione dell'integrale di Stiltjes per sottolineare che la funzione caratteristica può essere definita per una arbitraria funzione di ripartizione F . Ricordando quanto accennato nel Paragrafo 5 del Capitolo 1, notiamo che nel caso discreto e nel caso assolutamente continuo dobbiamo rispettivamente calcolare le funzioni caratteristiche secondo le seguenti formule.

$$\begin{aligned} \varphi_X(t) &= \sum_j \exp\{i \cdot t \cdot x_j\} \cdot p_j, \\ \varphi_X(t) &= \int_{\mathbb{R}} \exp\{i \cdot t \cdot x\} \cdot f(x) dx. \end{aligned} \quad (68)$$

Vedremo in proposito degli esempi notevoli nel successivo paragrafo.

Elenchiamo ora le proprietà che immediatamente seguono dalla definizione (67).

Ricordando la posizione (65) ed il calcolo del valore atteso di una trasformata di una variabile aleatoria reale, notiamo innanzitutto che il valore della φ_X , corrispondente al generico valore $t \in \mathbb{R}$, coincide con il numero complesso

$$\begin{aligned} \varphi_X(t) &= \int_{-\infty}^{+\infty} \cos(t \cdot x) dF_X(x) + i \int_{-\infty}^{+\infty} \sin(t \cdot x) dF_X(x) = \\ &= \mathbb{E}[\cos(t \cdot X)] + i \cdot \mathbb{E}[\sin(t \cdot X)]. \end{aligned} \quad (69)$$

Qualunque sia la distribuzione di probabilità di X , risulta in particolare

$$\varphi_X(0) = 1.$$

Se la distribuzione di probabilità di X è la distribuzione degenerata sul valore $\{0\}$, allora

$$\varphi_X(0) = 1, \quad \forall t \in \mathbb{R}.$$

Se la distribuzione di probabilità di X è la distribuzione degenerata sul valore $\{1\}$, allora

$$\varphi_X(t) = e^{i \cdot t}, \quad \forall t \in \mathbb{R}.$$

Sia X una variabile aleatoria con funzione caratteristica $\varphi_X(t)$ e sia fissata una qualunque costante reale a . Per quanto riguarda la funzione caratteristica della variabile aleatoria $a \cdot X$, risulta

$$\begin{aligned} \varphi_{a \cdot X}(t) &= \mathbb{E}[\cos(t \cdot a \cdot X)] + i \cdot \mathbb{E}[\sin(t \cdot a \cdot X)] = \\ &= \mathbb{E}[\cos(a \cdot t \cdot X)] + i \cdot \mathbb{E}[\sin(a \cdot t \cdot X)] = \varphi_X(a \cdot t). \end{aligned} \quad (70)$$

In particolare si ha

$$\varphi_{-X}(t) = \mathbb{E}[\cos(-t \cdot X)] + i \cdot \mathbb{E}[\sin(-t \cdot X)] = \mathbb{E}[\cos(t \cdot X)] - i \cdot \mathbb{E}[\sin(t \cdot X)] = \overline{\varphi_X(t)}.$$

Otteniamo infine, quale proprietà che immediatamente segue dalla definizione (67), la forma della funzione caratteristica della somma di due variabili aleatorie indipendenti.

Siano X, Y variabili aleatorie indipendenti, con funzioni caratteristiche $\varphi_X(t), \varphi_Y(t)$ e consideriamo il prodotto

$$\begin{aligned} \varphi_X(t) \cdot \varphi_Y(t) &= \\ &= [\mathbb{E}[\cos(t \cdot X)] + i \cdot \mathbb{E}[\sin(t \cdot X)]] \cdot [\mathbb{E}[\cos(t \cdot Y)] + i \cdot \mathbb{E}[\sin(t \cdot Y)]] = \\ &= \mathbb{E}[\cos(t \cdot X)] \cdot \mathbb{E}[\cos(t \cdot Y)] - \mathbb{E}[\sin(t \cdot X)] \cdot \mathbb{E}[\sin(t \cdot Y)] + \\ &+ i \cdot [\mathbb{E}[\cos(t \cdot X)] \cdot \mathbb{E}[\sin(t \cdot X)]] + \mathbb{E}[\sin(t \cdot X)] \cdot \mathbb{E}[\cos(t \cdot Y)]. \end{aligned}$$

Essendo X, Y stocasticamente indipendenti anche le coppie

$$\begin{aligned} &(\cos(t \cdot X), \sin(t \cdot Y)), (\cos(t \cdot X), \cos(t \cdot Y)), \\ &(\sin(t \cdot X), \sin(t \cdot Y)), (\sin(t \cdot X), \cos(t \cdot Y)) \end{aligned}$$

sono costituite da variabili indipendenti. Possiamo quindi scrivere

$$\begin{aligned} \varphi_X(t) \cdot \varphi_Y(t) &= \\ &= \mathbb{E}\{\cos(t \cdot X) \cdot \cos(t \cdot Y) - \sin(t \cdot X) \cdot \sin(t \cdot Y)\} + \\ &+ i \cdot \mathbb{E}\{\cos(t \cdot X) \cdot \sin(t \cdot X) + \sin(t \cdot X) \cdot \cos(t \cdot Y)\} = \\ &= \mathbb{E}[\cos(t \cdot (X + Y))] + i \cdot \mathbb{E}[\sin(t \cdot (X + Y))] \end{aligned}$$

e concludere con la seguente

Proposizione 1.2

Per variabili aleatorie stocasticamente indipendenti X, Y , la funzione caratteristica $\varphi_{(X+Y)}(t)$ è uguale al prodotto delle funzioni caratteristiche $\varphi_X(t), \varphi_Y(t)$.

Tale conclusione ovviamente risulta di importanza fondamentale e motiva l'utilizzo della funzione caratteristica quale strumento per lo studio del comportamento delle medie aritmetiche di variabili aleatorie indipendenti.

Siano in particolare X_1, \dots, X_n n variabili aleatorie indipendenti, identicamente distribuite con comune funzione caratteristica $\varphi_X(t)$, valore atteso nullo e varianza σ^2 . Consideriamone la somma $S_n = \sum_{j=1}^n X_j$, la media aritmetica $\bar{X}_n = \frac{S_n}{n}$, e la variabile standardizzata

$$Z_n = \frac{(\bar{X}_n - \mu) \sqrt{n}}{\sigma}.$$

Combinando la (70) con la Proposizione 1.2 otteniamo che la funzione caratteristica di Z_n è data da

$$\varphi_{Z_n}(t) = \left[\varphi_{S_n} \left(\frac{t\sqrt{n}}{n\sigma} \right) \right] = \left[\varphi_X \left(\frac{t\sqrt{n}}{n\sigma} \right) \right]^n. \quad (71)$$

Nel prossimo Paragrafo 3 verranno richiamati dei risultati basilari, circa le funzioni caratteristiche, che permettono di utilizzare la relazione (71) nella dimostrazione del Teorema di Lindberg-Lévy.

5.2 Funzioni caratteristiche di distribuzioni di probabilità notevoli

Calcoliamo qui di seguito le funzioni caratteristiche di quelle, fra le più notevoli distribuzioni di probabilità sulla retta, che rivestono maggiore importanza nello studio del comportamento asintotico di medie aritmetiche di variabili indipendenti e identicamente distribuite. Iniziamo a considerare dei casi di distribuzioni discrete.

Sia X una variabile aleatoria con distribuzione bernoulliana di parametro p , cioè tale che, per $p \in (0, 1)$,

$$\mathbb{P}(X = 1) = p, \mathbb{P}(X = 0) = 1 - p.$$

Si ha allora

$$\varphi_X(t) = p \cdot e^{i \cdot t} + (1 - p) = 1 + p(e^{i \cdot t} - 1).$$

Consideriamo ora il caso in cui X segue una distribuzione $b(n, p)$ binomiale di parametri n e p .

Per calcolare $\varphi_X(t)$ conviene ricordare che X ha la stessa distribuzione di probabilità della somma di n variabili X_1, \dots, X_n indipendenti, bernoulliane di parametro p . E quindi, in virtù della Proposizione 1.2, possiamo concludere

$$\varphi_X(t) = [1 + p(e^{i \cdot t} - 1)]^n.$$

Per una variabile aleatoria con distribuzione $\mathcal{P}(\lambda)$, Poisson di parametro λ ($\lambda > 0$), si ha invece

$$\varphi_X(t) = e^{-\lambda \cdot t} \sum_{k=0}^{\infty} \left(\frac{\lambda^k e^{i \cdot t \cdot k}}{k!} \right) = \exp\{\lambda(e^{i \cdot t} - 1)\}.$$

Passiamo ora al caso assolutamente continuo, iniziando a considerare la distribuzione esponenziale standard, caratterizzata dalla funzione di densità di probabilità

$$f(x) = e^{-x}.$$

Avremo quindi

$$\varphi_X(t) = \int_0^{+\infty} \cos(t \cdot x) \cdot e^{-x} dx + i \int_0^{+\infty} \sin(t \cdot x) \cdot e^{-x} dx.$$

Si verifica facilmente tramite integrazione per parti che, per $t \in \mathbb{R}$, valgono le relazioni

$$\begin{aligned} \int_0^{+\infty} \cos(t \cdot x) \cdot e^{-x} dx &= \frac{1}{1 + t^2}, \\ \int_0^{+\infty} \sin(t \cdot x) e^{-x} dx &= \frac{t}{1 + t^2}, \end{aligned}$$

da cui otteniamo

$$\varphi_X(t) = \frac{1 + i \cdot t}{1 + t^2} = \frac{1}{1 - i \cdot t}. \quad (72)$$

Osservazione 2.1

Poniamo $\beta(t) = (1 - i \cdot t)$. Attraverso il calcolo di integrali nel campo reale, abbiamo in effetti verificato che, per ogni $t > 0$, vale l'identità

$$\int_0^{+\infty} e^{-\beta(t) \cdot x} dx = \int_0^{+\infty} e^{i \cdot t \cdot x} e^{-x} dx = \frac{1}{1 - i \cdot t} = \frac{1}{\beta(t)}.$$

Cioè vale, per il numero puramente immaginario $\beta(t)$, la stessa identità valida nel caso di valori $\beta(t)$ reali positivi.

E' noto che la distribuzione esponenziale bilatera, caratterizzata dalla funzione di densità di probabilità

$$f(x) = \frac{1}{2}e^{-|x|}, \quad x \in \mathbb{R}$$

si incontra quando consideriamo la variabile "simmetrizzata" $X = X_1 - X_2$, essendo X_1, X_2 variabili indipendenti, esponenziali standard.

Otteniamo allora

$$\varphi_X(t) = \frac{1}{1-i \cdot t} \cdot \frac{1}{1+i \cdot t} = \frac{1}{1+t^2}. \quad (73)$$

Consideriamo invece il caso di una variabile aleatoria T_n con distribuzione di Erlang $G(n, 1)$. Ricordando che tale distribuzione coincide con quella della somma di n variabili aleatorie esponenziali standard indipendenti, dalla (72) otteniamo

$$\varphi_{T_n}(t) = \left(\frac{1}{1-i \cdot t} \right)^n.$$

Passiamo ora al calcolo della funzioni caratteristiche della distribuzione gaussiana standard. Per $Z \sim \mathcal{N}(0, 1)$,

$$\begin{aligned} \varphi_Z(t) &= \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{i \cdot t \cdot x} \exp\left\{-\frac{x^2}{2}\right\} dx. \\ &= \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} \cos(t \cdot z) \cdot \exp\left\{-\frac{z^2}{2}\right\} dz + i \int_{-\infty}^{+\infty} \sin(t \cdot z) \cdot \exp\left\{-\frac{z^2}{2}\right\} dz. \end{aligned}$$

Essendo, per ogni $t \in \mathbb{R}$, $\sin(t \cdot x) \cdot \exp\left\{-\frac{x^2}{2}\right\}$ una funzione dispari dell'argomento x , sommabile, abbiamo che $\varphi_X(t)$ è un numero reale:

$$\begin{aligned} \varphi_X(t) &= \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} \cos(t \cdot x) \cdot \exp\left\{-\frac{x^2}{2}\right\} dx = \\ &= \frac{\sqrt{2}}{\sqrt{\pi}} \int_0^{+\infty} \cos(t \cdot x) \cdot \exp\left\{-\frac{x^2}{2}\right\} dx. \end{aligned}$$

Calcolando tale integrale si dimostra che risulta

$$\varphi_Z(t) = e^{-\frac{1}{2}t^2}.$$

Che debba valere tale identità si può intuire anche attraverso la seguente procedura:

$$\varphi_Z(t) = \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{i \cdot t \cdot z} \exp\left\{-\frac{z^2}{2}\right\} dz =$$

$$\begin{aligned}
&= \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}(z^2 - 2 \cdot i \cdot tz)\right\} dz = \\
&= \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}(z^2 - 2 \cdot i \cdot t \cdot z - t^2)\right\} \cdot \exp\left\{-\frac{t^2}{2}\right\} dz = \\
&\quad \exp\left\{-\frac{t^2}{2}\right\} \cdot \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} \left\{-\frac{1}{2}(z - i \cdot t)^2\right\} dz = \\
&\quad \exp\left\{-\frac{t^2}{2}\right\} \cdot \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} \left\{-\frac{1}{2}(z - m(t))^2\right\} dz,
\end{aligned}$$

avendo posto $m(t) = i \cdot t$.

Se $m(t)$ fosse una funzione a valori reali, avremmo ovviamente (integrale esteso a tutta la retta di una funzione di densità gaussiana di valore atteso $m(t)$ e varianza 1):

$$\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} \left\{-\frac{1}{2}(z - m(t))^2\right\} dz = 1.$$

Analogamente a quanto notato nell'Osservazione 2.1, si ha che tale identità vale anche nel caso puramente immaginario $m(t) = i \cdot t$.

Osservazione 2.2

Quanto precedentemente notato per il calcolo della funzione caratteristica della distribuzione gaussiana standard si estende a tutte le distribuzioni simmetriche: quando la funzione di densità di X è una funzione pari risulta

$$\int_{-\infty}^{+\infty} \sin(t \cdot x) \cdot \exp\left\{-\frac{x^2}{2}\right\} dx = 0$$

e dunque $\varphi_X(t)$ è a valori reali. Se, in particolare, vale relazione

$$X = X_1 - X_2,$$

essendo X_1, X_2 variabili indipendenti identicamente distribuite con funzione caratteristica nota e uguale a $\varphi(t)$, allora

$$\varphi_X(t) = |\varphi(t)|^2,$$

analogamente a quanto prima visto per la distribuzione esponenziale bilatera.

5.3 Proprietà fondamentali delle funzioni caratteristiche e Teorema di Lindberg-Lévy

Elenchiamo qui di seguito i risultati, inerenti proprietà fondamentali delle funzioni caratteristiche, che portano a dimostrare il Teorema di Lindberg-Lévy. Circa tali risultati, ci limiteremo a presentare alcuni aspetti fondamentali e qualche commento, senza riportarne le dimostrazioni. Di queste, che possono

essere trovate su testi classici di Probabilità ¹, risulterebbe piuttosto impegnativo riportare qui tutti i dettagli.

Un primo risultato garantisce l'unicità della distribuzione a cui è associata una funzione caratteristica:

Teorema 3.1

Distribuzioni di probabilità fra loro distinte ammettono funzioni caratteristiche distinte.

A tale proposito possiamo ottenere una *Formula di Inversione* che mostra come, sotto opportuna ipotesi sulla funzione caratteristica, la distribuzione di probabilità associata ammetta densità e questa si ottenga a partire dalla funzione caratteristica stessa. Dobbiamo preliminarmente citare un secondo fondamentale risultato che mostra che la corrispondenza biunivoca che lega fra di esse funzioni di ripartizione e funzioni caratteristiche si rivela continua, rispetto ad opportune nozioni di convergenza.

Per successioni di funzioni di ripartizione una naturale nozione di convergenza è data dalla seguente

Definizione 3.2

Siano date una funzione di ripartizione F ed una successione di funzioni di ripartizione $\{F_n\}_{n=1,2,\dots}$. Si dice che $\{F_n\}_{n=1,2,\dots}$ converge debolmente a F , se $\{F_n(x)\}_{n=1,2,\dots}$ converge a $F(x)$, per ogni x che sia un punto di continuità per F .

La convergenza debole di una successione di distribuzioni di probabilità può essere caratterizzata in termini della successione delle corrispondenti funzioni caratteristiche. Si ha infatti

Teorema 3.3 (*di continuità*)

Siano date una funzione di ripartizione F ed una successione di funzioni di ripartizione $\{F_n\}_{n=1,2,\dots}$, e siano φ e $\{\varphi_n\}_{n=1,2,\dots}$ le corrispondenti funzioni caratteristiche. $\{F_n\}_{n=1,2,\dots}$ converge debolmente a F se e soltanto se $\{\varphi_n\}_{n=1,2,\dots}$ converge puntualmente alla φ su tutta la retta reale.

Esempio 3.1

Un esempio significativo sul ruolo della funzione caratteristica nello studio della convergenza di una successione di distribuzioni di probabilità riguarda l'*approssimazione di Poisson* delle probabilità binomiali.

Fissato $\lambda > 0$, consideriamo al variare di $n \in \mathbb{N}$ la famiglia delle distribuzioni binomiali $b(n, \frac{\lambda}{n})$. E' ben noto che, al divergere di n , si ha una convergenza verso la distribuzione di Poisson $\mathcal{P}(\lambda)$. Guardando alle funzioni caratteristiche,

¹Quali, ad esempio,
W. Feller. An Introduction to Probability and its Applications. Vol II. Second Edition. 1970. J. Wiley & Sons
P. Billingsley. Probability and Measure. 2008. Wiley & Sons
Entrambi i testi presentano un ricchissimo materiale costituito da molti risultati, commenti, ed esempi sull'argomento

vediamo che risulta infatti

$$\lim_{n \rightarrow \infty} \left[1 + \frac{\lambda}{n} (e^{i \cdot t} - 1) \right]^n = \exp\{\lambda (e^{i \cdot t} - 1)\}.$$

Torniamo ora alla questione dell'inversione di una funzione caratteristica. Sia F una funzione di ripartizione e sia la sua funzione caratteristica φ tale che

$$\int_{-\infty}^{+\infty} |\varphi(t)| dt < +\infty. \quad (74)$$

Per semplicità di notazione, consideriamo ora il caso in cui già si sappia che F ammette una funzione di densità $f(x)$. Sia inoltre g una qualunque funzione di densità di probabilità sulla retta e sia γ la funzione caratteristica ad essa associata. Moltiplicando entrambi i membri della (68) per il termine $e^{-i \cdot t \cdot y} \cdot g(t)$ (con $y \in \mathbb{R}$) ed integrando rispetto a t da $-\infty$ a $+\infty$, possiamo scrivere

$$\int_{-\infty}^{+\infty} e^{-i \cdot t \cdot y} \cdot g(t) \cdot \varphi(t) dt = \int_{-\infty}^{+\infty} e^{-i \cdot t \cdot y} \cdot g(t) \cdot \left[\int_{-\infty}^{+\infty} \exp\{i \cdot t \cdot (x - y)\} \cdot f(x) dx \right] dt.$$

Prima scambiando l'ordine di integrazione a secondo membro e poi ricordando che si è indicata con γ la funzione caratteristica associata alla densità g otteniamo

$$\int_{-\infty}^{+\infty} e^{-i \cdot t \cdot y} \cdot g(t) \cdot \varphi(t) dt = \int_{-\infty}^{+\infty} f(x) \cdot \gamma(x - y) dx. \quad (75)$$

Consideriamo ora il particolare caso in cui g e γ siano rispettivamente la funzione di densità e la funzione caratteristica della distribuzione gaussiana di valore atteso 0 e varianza $\frac{1}{a^2}$:

$$g(t) = \frac{a}{\sqrt{2\pi}} \exp\left\{-\frac{a^2}{2} t^2\right\}, \gamma(v) = \exp\left\{-\frac{v^2}{2a^2}\right\}$$

Otteniamo dunque

$$\int_{-\infty}^{+\infty} e^{-i \cdot t \cdot y} \cdot \frac{a}{\sqrt{2\pi}} \exp\left\{-\frac{a^2}{2} t^2\right\} \cdot \varphi(t) dt = \int_{-\infty}^{+\infty} f(x) \cdot \exp\left\{-\frac{(x - y)^2}{2a^2}\right\} dx.$$

che riscriviamo nella forma

$$\frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{-i \cdot t \cdot y} \cdot \exp\left\{-\frac{a^2}{2} t^2\right\} \cdot \varphi(t) dt = \int_{-\infty}^{+\infty} f(x) \cdot \frac{1}{a\sqrt{2\pi}} \exp\left\{-\frac{(y - x)^2}{2a^2}\right\} dx. \quad (76)$$

Osserviamo che a secondo membro compare il prodotto di convoluzione fra la funzione di densità f e funzione di densità di una distribuzione gaussiana di

valore atteso 0 e varianza a^2 . Indicando rispettivamente con X e Z due variabili aleatorie indipendenti con distribuzione F e $\mathcal{N}(0, 1)$ rispettivamente vediamo che l'espressione a secondo membro coincide quindi con $f_{Y_a}(y)$, il valore calcolato in $y \in \mathbb{R}$ della funzione di densità della variabile aleatoria

$$Y_a = X + aZ.$$

In virtù dell'identità (76) e dell'arbitrarietà di $a > 0$, ci si può convincere euristicamente che la distribuzione di probabilità di X è individuata univocamente dalla conoscenza della $\varphi(t)$. Al tendere a 0 di a , cioè della varianza a^2 dell'addendo aZ , $f_{Y_a}(y)$ convergerà a $f(y)$. D'altra parte l'espressione a primo membro convergerà a

$$\frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{-i \cdot t \cdot y} \cdot \varphi(t) dt.$$

Sotto la condizione (74), tale passaggio al limite può essere reso rigoroso e si ottiene quindi l'identità

$$f(y) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{-i \cdot t \cdot y} \cdot \varphi(t) dt. \quad (77)$$

Ne segue anche che $f(y)$ è necessariamente continua e limitata. Come già citato, abbiamo inizialmente assunto che comunque la F ammetta una funzione di densità. Per giungere alle precedenti conclusioni, tale assunzione non sarebbe effettivamente necessaria.

Un'ulteriore aspetto fondamentale del metodo della funzione caratteristica risiede nel legame che sussiste fra le derivate successive, calcolate nell'origine, della $\varphi(t)$ e i momenti di diverso ordine della distribuzione $F(x)$.

Per i nostri scopi, consideriamo quanto segue.

Le derivate prima e seconda di $\varphi(t)$ sono rispettivamente definite dalle relazioni

$$\varphi'(t) = \frac{d}{dt} \int_{-\infty}^{+\infty} \cos(t \cdot x) dF(x) + i \frac{d}{dt} \int_{-\infty}^{+\infty} \sin(t \cdot x) dF(x),$$

$$\varphi''(t) = \frac{d^2}{dt^2} \int_{-\infty}^{+\infty} \cos(t \cdot x) dF(x) + i \frac{d^2}{dt^2} \int_{-\infty}^{+\infty} \sin(t \cdot x) dF(x).$$

Cioè

$$\varphi'(t) = - \int_{-\infty}^{+\infty} x \sin(t \cdot x) dF(x) + i \int_{-\infty}^{+\infty} x \cos(t \cdot x) dF(x),$$

$$\varphi''(t) = - \int_{-\infty}^{+\infty} x^2 \cos(t \cdot x) dF(x) - i \int_{-\infty}^{+\infty} x^2 \sin(t \cdot x) dF(x).$$

Vediamo dunque che, in $t = 0$, tale derivate esistono finite se e solo se rispettivamente risulta

$$\mathbb{E}(|X|) < +\infty, \mathbb{E}(X^2) < +\infty,$$

ed, in tal caso, si ha

$$\varphi'(0) = i\mathbb{E}(X), \varphi''(0) = -\mathbb{E}(X^2). \quad (78)$$

Consideriamo ora il caso di una variabile aleatoria X con valore atteso nullo e varianza finita σ^2 . In base alla formula di Taylor, e tenendo conto di (78) potremo scrivere:

$$\varphi_X(t) = 1 - \frac{\sigma^2}{2}t^2 + \psi(t) \quad (79)$$

dove

$$\lim_{t \rightarrow 0} \frac{\psi(t)}{t^2} = 0. \quad (80)$$

Raccogliendo tutti questi risultati, segue direttamente la seguente linea di dimostrazione del Teorema di Lindberg-Lévy, già enunciato nel Paragrafo 5 del Capitolo 3.

Consideriamo allora n variabili aleatorie X_1, \dots, X_n indipendenti, identicamente distribuite con comune funzione caratteristica $\varphi_X(t)$, valore atteso μ e varianza σ^2 e guardiamo alle variabili standardizzate $Z_n = \frac{(\bar{X}_n - \mu)\sqrt{n}}{\sigma}$.

Dobbiamo dimostrare che la distribuzione di probabilità di Z_n converge debolmente alla distribuzione gaussiana standard.

Senza alcuna perdita di generalità, possiamo assumere che sia $\mu = 0$. Se così non fosse, infatti, basterebbe considerare le nuove variabili centrate X'_1, X'_2, \dots definite da $X'_j = X_j - \mu$, applicare a queste il risultato e poi tornare alle X_j .

Combinando la (79) con la (71) del primo paragrafo, otteniamo

$$\begin{aligned} \varphi_{Z_n}(t) &= \left[1 - \frac{\sigma^2}{2} \left(\frac{t\sqrt{n}}{n\sigma} \right)^2 + \psi\left(\frac{t\sqrt{n}}{n\sigma}\right) \right]^n = \\ &= \left[1 - \frac{1}{n} \left(\frac{t^2}{2} + n \cdot \psi\left(\frac{t\sqrt{n}}{n\sigma}\right) \right) \right]^n. \end{aligned}$$

In virtù del Teorema 3.3, e ricordando che la funzione caratteristica della distribuzione gaussiana standard è data da $\exp\{-\frac{1}{2}t^2\}$, la dimostrazione del teorema si riduce alla verifica che la successione $\{\varphi_{Z_n}(t)\}$ converge puntualmente alla funzione $\exp\{-\frac{1}{2}t^2\}$, per ogni $t \in \mathbb{R}$.

Ciò si può dimostrare rigorosamente in virtù del fatto che, per la (80), si ha

$$\lim_{n \rightarrow \infty} n \cdot \psi\left(\frac{t\sqrt{n}}{n\sigma}\right) = 0.$$

5.4 Proprietà di stabilità rispetto alla somma

Ci si può domandare quale speciale caratteristica della distribuzione gaussiana standard possa spiegarne il particolare ruolo nella teoria delle somme di variabili aleatorie indipendenti, identicamente distribuite. A tale domanda, si può dare risposta cominciando ad osservare quanto segue.

Siano Z', Z'' due variabili aleatorie indipendenti con distribuzione $\mathcal{N}(0, \sigma^2)$. Sappiamo che per esse vale la condizione $Z' + Z'' \sim \mathcal{N}(0, 2\sigma^2)$, cioè

$$Z' + Z'' \stackrel{d}{=} Z' \sqrt{2}.$$

Come già visto, questa relazione si traduce, in termini della funzione caratteristica, nella forma

$$[\varphi_Z(t)]^2 = \varphi_Z(t\sqrt{2}).$$

Sia ora F una qualunque distribuzione di probabilità di valore atteso 0 e varianza finita σ^2 , e sia X_1, X_2, \dots una successione di variabili i.i.d con distribuzione F . Ponendo

$$Y_n = \frac{\bar{X}_n}{\sigma} \sqrt{n}$$

$$\bar{X}_{2n} = \frac{1}{n} \sum_{j=n+1}^{2n} X_j, Y_{2n} = \frac{\bar{X}_{2n}}{\sigma} \sqrt{n},$$

possiamo scrivere

$$\sqrt{2} \cdot Y_{2n} = (Y_n + Y_{2n}^n). \quad (81)$$

Supponiamo ora che G sia una distribuzione di probabilità (di valore atteso 0) tale che la distribuzione di probabilità della variabile Y_n converga a G al divergere di n . Indichiamo con γ la funzione caratteristica di G e con ϑ', ϑ'' due variabili indipendenti con distribuzione G .

Facendo divergere n vediamo quindi dalla (81) che, siccome le distribuzioni di Y_{2n}, Y_n e Y_{2n}^n convergono tutte a quella di ϑ' e ϑ'' , si deve avere

$$\vartheta' + \vartheta'' \stackrel{d}{=} \vartheta' \sqrt{2},$$

o equivalentemente, in termini della funzione caratteristica γ ,

$$[\gamma(t)]^2 = \gamma(t\sqrt{2}). \quad (82)$$

Si può dimostrare che le distribuzioni gaussiane $\mathcal{N}(0, \sigma^2)$ sono le uniche distribuzioni di probabilità per cui possano valere tali proprietà; in particolare $\gamma(t) = \exp\{-\frac{\sigma^2}{2}t^2\}$ sono le uniche funzioni caratteristiche per cui possa valere la relazione (82). E quindi la distribuzione gaussiana standard è l'unica distribuzione che può presentarsi come limite delle distribuzioni delle medie aritmetiche standardizzate di variabili i.i.d.

Il Teorema di Lindberg-Lévy di fatto dimostra che una tale convergenza effettivamente sussiste, qualunque sia la distribuzione F delle variabili di partenza,

purchè dotata di varianza finita. Notiamo a questo proposito che la condizione di varianza finita è essenziale per poter parlare di variabili standardizzate.

Per $a > 0$, possiamo, più in generale, considerare una condizione del tipo

$$\vartheta' + \vartheta'' \stackrel{d}{=} a \cdot \vartheta', \quad (83)$$

essendo ancora ϑ', ϑ'' due variabili indipendenti con uguale distribuzione G (di valore atteso nullo). In termini della funzione caratteristica (83) equivale a

$$[\varphi(t)]^2 = a \cdot \varphi(t). \quad (84)$$

Ci possiamo domandare

i) esistono altre distribuzioni, oltre alle $\mathcal{N}(0, \sigma^2)$, per cui possa valere la relazione (83)?

ii) in caso affermativo, quali possono essere i valori della costante a ?

Le distribuzioni di probabilità che soddisfano la relazione (83) vengono dette *distribuzioni stabili (rispetto alla somma)*.

La risposta alla prima domanda, cioè se possano esistere altre distribuzioni stabili oltre alle gaussiane di media nulla, è effettivamente affermativa e lo dimostreremo subito con un esempio.

Anche alla seconda domanda si può dare una risposta chiara, che successivamente citeremo. Notiamo comunque che deve necessariamente valere $a = \sqrt{2}$, se si richiede che le variabili in gioco abbiano varianza finita. Possiamo concludere affermando che le distribuzioni $\mathcal{N}(0, \sigma^2)$ sono le uniche stabili con varianza finita.

Tutte le altre distribuzioni stabili hanno necessariamente varianza infinita. Quella mostrata nel prossimo esempio neanche ammette valore atteso.

Esempio 4.1

Nel precedente Paragrafo 2 si era visto, relativamente alla funzione caratteristica della distribuzione esponenziale bilatera, che risulta, per ogni $t \in \mathbb{R}$,

$$\frac{1}{2} \int_{\mathbb{R}} \exp\{i \cdot t \cdot x\} \cdot \exp\{-|x|\} dx = \frac{1}{1+t^2}.$$

e quindi anche, ricordando la formula (77),

$$\frac{1}{2} \exp\{-|y|\} = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{-i \cdot t \cdot y} \cdot \frac{1}{1+t^2} dt.$$

Eseguendo il cambiamento di variabile $v = -t$ dentro l'integrale nel membro a sinistra e tenuto conto che $\frac{1}{1+t^2}$ è una funzione pari, tale identità può essere riscritta nella forma

$$\int_{-\infty}^{+\infty} e^{i \cdot v \cdot y} \cdot \frac{1}{\pi} \cdot \frac{1}{1+v^2} dv = \exp\{-|y|\}, v \in \mathbb{R}.$$

Ricordiamo ora che la funzione $\frac{1}{\pi} \cdot \frac{1}{1+v^2}$ altro non è che la densità di probabilità della distribuzione di Cauchy. Ed abbiamo così verificato che la funzione caratteristica di tale distribuzione è data da

$$\varphi(t) = \exp\{-|t|\}.$$

La distribuzione di Cauchy è dunque una distribuzione stabile, in quanto la (84) risulta verificata con $a = 2$. Più in generale vediamo che, se X_1, \dots, X_n sono variabili i.i.d. con distribuzione di Cauchy, allora anche la loro media aritmetica \bar{X}_n segue la stessa distribuzione!

Tale conclusione ci può apparire paradossale, e comunque anti-intuitiva. In particolare non potrà valere, per successioni di variabili di Cauchy indipendenti, una *legge dei grandi numeri*, in quanto non si ha alcun tipo di convergenza delle medie aritmetiche verso una variabile degenere. Ciò è coerente con la seguente circostanza:

come si è detto- e come messo in evidenza dal fatto che la funzione $\exp\{-|t|\}$ non è derivabile in $t = 0$ - la distribuzione di Cauchy non ammette valore atteso (e la sua varianza è uguale a $+\infty$).

Che ciò possa apparire anti-intuitivo è dovuto al fatto che possiamo essere abituati a tarare il nostro ragionamento probabilistico sui casi in cui esistono i valori attesi e si abbiano varianze finite.

Le distribuzioni in cui ciò non accade sono dette *a coda pesante* e sono tali che, all'aumentare del numero delle osservazioni, aumenta la probabilità che si presenti un'osservazione di valore così grande (in modulo) da assumere un peso predominante rispetto alla media aritmetica di tutte le osservazioni precedenti.

Tali distribuzioni non vanno considerate come patologiche; esse hanno acquisito, negli ultimi due/tre decenni, un ruolo sempre più rilevante nello studio di nuovi fenomeni quali l'analisi di dati nei mercati finanziari e il traffico di files su reti complesse.

Possiamo ora rispondere anche alla domanda ii) che c'eravamo precedentemente posti: si dimostra (si vedano ad esempio i testi indicati nella precedente nota) che esistono ∞^1 distribuzioni stabili. Ognuna di esse è caratterizzata da un parametro α , detto *esponente caratteristico*, con $0 < \alpha \leq 2$.

Il caso $\alpha = 2$ corrisponde alle distribuzioni gaussiane (di valore atteso nullo), il caso $\alpha = 1$ corrisponde alla distribuzione di Cauchy. Il caso generale è caratterizzato dalla condizione che la funzione caratteristica soddisfa la (83) con $\alpha = 2^{\frac{1}{\alpha}}$.

Nessuna di tali distribuzioni, escludendo il caso gaussiano, può avere varianza finita e la caratteristica di code pesanti aumenta al diminuire di α .

6 Processi di Poisson e alcune loro proprietà

In questo capitolo verranno studiate le nozioni basilari circa i *Processi di Poisson omogenei*; ne vedremo la definizione, la *legge del processo*, la *proprietà di statistica ordinata*. Da questi argomenti, emergeranno in modo naturale i legami con le distribuzioni di Poisson, esponenziali, gamma, e beta. Faremo anche un brevissimo accenno ai processi di Poisson *non omogenei*. Il lettore avrà già una certa familiarità con processi *aleatori a tempo discreto*, quali le *catene di Markov a tempo discreto*. Il Processo di Poisson fornisce invece un primo esempio di *processo aleatorio a tempo continuo*, anzi, più in particolare, di *catena di Markov a tempo continuo*. Tale nozione è, inoltre, di fondamentale interesse in quanto costituisce un caso particolare e notevolissimo di *processo di conteggio* e di *processo di rinnovo*. I processi di Poisson mostrano una struttura molto semplice e godono al tempo stesso di proprietà così notevoli da dar luogo al modello idealizzato di processo di conteggio e di processo di rinnovo. Altre loro proprietà, che verranno qui tralasciate ma che sono comunque assai interessanti, fanno dei processi di Poisson una base fondamentale per lo studio di diversi aspetti di carattere sia teorico che applicativo nei campi della probabilità e dei processi aleatori.

6.1 Definizione e proprietà fondamentali di Processo di Poisson omogeneo

Un *processo di Poisson* è innanzitutto un particolare *processo di conteggio*. Per poter definire che cosa sia un processo di Poisson è dunque necessario tener presente che cosa si intenda per processo di conteggio. Limitandoci ad una descrizione in termini schematici e semplificativi, si può richiamare quanto segue. Pensiamo ad un fenomeno che si presenti ripetutamente al trascorrere del tempo con ripetizioni che avvengono ad istanti aleatori ed imprevedibili (cioè non precedute da segnali premonitori). Immediati e naturali esempi di tale schema sono i seguenti: arrivo di telefonate al centralino telefonico di un grande albergo, emissione di particelle da parte di un materiale radioattivo, denunce di sinistri ad una compagnia di assicurazione, etc...

Parleremo in particolare di processi di conteggio *semplici*, cioè tali da escludere, in pratica, la possibilità di due arrivi contemporaneamente.

Sulla retta reale fissiamo ora un *unità di misura*, intesa come l'appropriata unità di tempo per l'analisi del fenomeno in questione (un minuto, un millisecondo, un giorno lavorativo, etc) e fissiamo un *origine temporale*, a partire dalla quale si cominciano a contare le ripetizioni del fenomeno. Per ogni $t > 0$, indichiamo con N_t la variabile aleatoria che indica il *numero degli arrivi* (cioè delle volte in cui si è ripetuto il fenomeno) nell'intervallo di tempo $[0, t]$. La collezione ordinata delle variabili aleatorie $\{N_t\}_{t \geq 0}$ è detta processo di conteggio.

Un processo di conteggio costituisce un semplice esempio di *processo aleatorio*. In termini ancora semplificativi, possiamo dire che ci troviamo di fronte ad un processo aleatorio ogni volta che consideriamo una funzione aleatoria di una variabile temporale. Nel caso di processi di conteggio la funzione del tempo

$\{N_t\}_{t \geq 0}$ è una funzione a gradini, con salti unitari che si presentano ad istanti aleatori; ciascun salto si ha in corrispondenza all'istante in cui avviene una ripetizione, o un *arrivo*, del fenomeno studiato. Gli istanti di arrivo sono dunque aleatori.

Nello studio di un processo di conteggio è quindi naturale considerare due successioni di variabili aleatorie $\{T_1, T_2, \dots\}$ e $\{X_1, X_2, \dots\}$: T_1, T_2, \dots indicano gli *istanti di arrivo*, mentre X_1, X_2, \dots indicano i cosiddetti *intertempi* o *tempi di interarrivo*, cioè le lunghezze degli intervalli temporali fra un arrivo ed il successivo. Potremo scrivere, più precisamente, le seguenti relazioni fra tali variabili aleatorie:

$$N_t = \sup\{n \in \mathbb{N} | T_n \leq t\},$$

oppure

$$N_t = \sum_{n=1}^{\infty} \mathbf{1}_{\{T_n \leq t\}}, \quad (85)$$

$$X_1 = T_1, X_2 = T_2 - T_1, \dots, X_n = T_n - T_{n-1}, \dots$$

$$T_1 = X_1, T_2 = X_1 + X_2, \dots, T_n = \sum_{k=1}^n X_k. \quad (86)$$

Notiamo inoltre che risulta

$$\{N_t < n\} \equiv \{T_n > t\} \quad (87)$$

e che T_{N_t} indica l'istante in cui è occorso l'ultimo arrivo precedente all'istante t .

La *legge* di un processo di conteggio semplice è determinata dalla specificazione delle distribuzioni congiunte dei vettori aleatori

$$(T_1, T_2, \dots, T_n), n = 2, 3, \dots$$

oppure

$$(X_1, X_2, \dots, X_n), n = 2, 3, \dots$$

Speciali casi di processi di conteggio sono i *processi di rinnovo*.

Questi sono definiti dalla condizione che $X_1, X_2, \dots, X_n, \dots$ sia una successione di variabili aleatorie indipendenti, identicamente distribuite, con distribuzione assegnata F .

I processi di Poisson, a loro volta, sono dei processi di rinnovo assai speciali: gli intertempi $X_1, X_2, \dots, X_n, \dots$ seguono una distribuzione esponenziale. Vedremo qui di seguito che tale condizione implica delle proprietà notevolissime.

Consideriamo dunque un processo di conteggio $\{N_t\}_{t \geq 0}$ in cui gli intertempi $X_1, X_2, \dots, X_n, \dots$ sono variabili aleatorie indipendenti, identicamente distribuite, con distribuzione $\mathcal{E}(\mu)$.

Diremo che $\{N_t\}_{t \geq 0}$ è un processo di Poisson (omogeneo) di intensità μ .

Innanzitutto vogliamo ricavare, per $t > 0$ fissato, la distribuzione di probabilità della variabile aleatoria discreta N_t .

A tale scopo risulterà utile calcolare la probabilità condizionata

$$P\{N_{t+\Delta t} = n + 1 | N_t = n, T_1 = t_1, \dots, T_n = t_n\}.$$

Dobbiamo notare che, nonostante l'evento $\{N_t = n, T_1 = t_1, \dots, T_n = t_n\}$, rispetto a cui si sta condizionando, sia un evento di probabilità nulla, possiamo ugualmente dare a tale probabilità condizionata un significato rigoroso, in quanto possiamo anche scrivere

$$P\{N_{t+\Delta t} = n + 1 | N_t = n, T_1 = t_1, \dots, T_n = t_n\} :=$$

$$P\{T_{n+1} \leq t + \Delta t, T_{n+2} > t + \Delta t | T_{n+1} > t, T_1 = t_1, \dots, T_n = t_n\}, \quad (88)$$

e a quest'ultima espressione può essere attribuito un significato basato sul fatto che le variabili aleatorie T_1, T_2, \dots, T_n ammettono una funzione di densità congiunta: per ogni $n = 1, 2, \dots$,

$$P\{T_{n+1} \leq t + \Delta t, T_{n+2} > t + \Delta t | T_{n+1} > t, T_1 = t_1, \dots, T_n = t_n\} =$$

$$\frac{\int_{t+\Delta t}^{+\infty} \int_t^{t+\Delta t} f_{T_1 \dots T_{n+2}}(t_1, \dots, t_n, \xi, \eta) d\xi d\eta}{\int_t^{+\infty} f_{T_1 \dots T_{n+1}}(t_1, \dots, t_n, \xi) d\xi}.$$

Tenendo conto della speciale condizione $X_n \sim \mathcal{E}(\mu)$ ed in virtù della proprietà di mancanza di memoria delle distribuzioni esponenziali, si ha la seguente

Proposizione 5.1

Per $0 < t_1 < t_2 < \dots < t_n < t$, risulta

$$\lim_{\Delta t \rightarrow 0} \frac{P\{N_{t+\Delta t} = n + 1 | N_t = n, T_1 = t_1, \dots, T_n = t_n\}}{\Delta t} = \lim_{\Delta t \rightarrow 0} \frac{P\{N_{t+\Delta t} = n + 1 | N_t = n\}}{\Delta t} = \mu. \quad (89)$$

inoltre

$$\lim_{\Delta t \rightarrow 0} \frac{P\{N_{t+\Delta t} > n + 1 | T_1 = t_1, \dots, T_n = t_n\}}{\Delta t} = \lim_{\Delta t \rightarrow 0} \frac{P\{N_{t+\Delta t} > n + 1 | T_n = t_n\}}{\Delta t} = 0. \quad (90)$$

Dimostrazione

Notiamo che, in termini delle variabili $X_1, X_2, \dots, X_n, \dots$, la probabilità condizionata

$$P\{T_{n+1} \leq t + \Delta t, T_{n+2} > t + \Delta t | T_{n+1} > t, T_1 = t_1, \dots, T_n = t_n\},$$

in (88) può essere riscritta come

$$\mathbb{P}\{X_{n+1} \leq t - t_n + \Delta t, X_{n+2} > \Delta t | X_1 = t_1, \dots, X_n = t_n - t_{n-1}, X_{n+1} > t - t_n\}.$$

Tenendo conto dell'ipotesi che X_1, X_2, \dots, X_n sono i.i.d. $\sim \mathcal{E}(\mu)$, tale probabilità condizionata è uguale a

$$(1 - \exp\{-\mu\Delta t\}) \exp\{-\mu\Delta t\}.$$

Inoltre

$$\mathbb{P}\{N_{t+\Delta t} > n + 1 | T_1 = t_1, \dots, T_n = t_n\} =$$

$$= \mathbb{P}\{X_{n+1} \leq t - t_n + \Delta t, X_{n+2} \leq \Delta t | X_1 = t_1, \dots, X_n = t_n - t_{n-1}, X_{n+1} > t - t_n\} =$$

$$(1 - \exp\{-\mu\Delta t\})^2.$$

Dalle quali uguaglianze segue la tesi.

Poniamo ora

$$p_k(t) = P\{N_t = k\}, k = 1, 2, \dots$$

A partire dalle (90) e (89), ricaveremo per le $p_k(t)$ un sistema di equazioni differenziali (ordinarie, lineari, del primo ordine) la cui soluzione permetterà di ottenere la distribuzione di N_t per ogni $t > 0$ fissato.

Notiamo inizialmente che, essendo

$$T_1 = X_1; X_1 \sim \mathcal{E}(\mu),$$

possiamo scrivere, ricordando l'osservazione in (87),

$$p_0(t) = \mathbb{P}\{T_1 > t\} = \exp\{-\mu t\}, \forall t > 0. \quad (91)$$

Per un fissato $n > 0$, scriveremo invece, utilizzando la formula delle probabilità totali,

$$p_n(t + \Delta t) = \sum_{k=0}^{\infty} p_k(t) \mathbb{P}\{N_{t+\Delta t} = n | N_t = k\} =$$

$$p_{n-1}(t) \mu \Delta t + p_n(t) (1 - \mu \Delta t) + o(\Delta t).$$

Possiamo quindi ottenere l'equazione differenziale

$$\frac{d}{dt} p_n(t) = \mu p_{n-1}(t) - \mu p_n(t).$$

Imponendo la condizione iniziale

$$p_n(0) = 0,$$

otteniamo un sistema che sappiamo ammettere una ed una sola soluzione.

Risolvendo ricorsivamente per $n = 1, 2, \dots$, si dimostra facilmente che tale soluzione è data da

$$p_n(t) = \exp\{-\mu t\} \frac{(\mu t)^n}{n!}. \quad (92)$$

Possiamo concludere dunque che

$$N_t \sim \mathcal{P}(\mu t).$$

Vogliamo ora ottenere la *legge congiunta del processo a più istanti*: per ogni scelta di un naturale m , di m istanti

$$0 < t_1 < \dots < t_m$$

e di m valori

$$k_1 \leq \dots < k_m,$$

vogliamo ricavare la probabilità congiunta

$$P\{N_{t_1} = k_1, \dots, N_{t_m} = k_m\};$$

in altre parole, vogliamo ricavare la distribuzione congiunta delle variabili aleatorie

$$N_{t_1}, \dots, N_{t_m}.$$

A tale proposito osserviamo che, come conseguenza della definizione di processo di Poisson e della proprietà di mancanza di memoria delle distribuzioni esponenziali, si ha che il comportamento probabilistico del processo $\{N_t\}_{t \geq 0}$ in un intervallo $[s, s+t]$ non dipende dal comportamento del processo nel precedente intervallo $[0, s]$ ed è uguale al comportamento del processo nell'intervallo $[0, t]$; più formalmente possiamo dire

$$N_s \perp\!\!\!\perp (N_{s+t} - N_s); \quad (93)$$

$$(N_{s+t} - N_s) \sim \mathcal{P}(\mu t). \quad (94)$$

Possiamo scrivere ora, per la legge delle probabilità composte,

$$\begin{aligned} & P\{N_{t_1} = k_1, \dots, N_{t_m} = k_m\} = \\ & = P\{N_{t_1} = k_1\} P\{N_{t_2} = k_2 | N_{t_1} = k_1\} \cdot \dots \cdot P\{N_{t_m} = k_m | N_{t_{m-1}} = k_{m-1}\} = \\ & = P\{N_{t_1} = k_1\} P\{N_{t_2} - N_{t_1} = k_2 - k_1\} \cdot \dots \cdot P\{N_{t_m} - N_{t_{m-1}} = k_m - k_{m-1}\} = \\ & \exp\{-\mu t_1\} \frac{[\mu t_1]^{k_1}}{k_1!} \exp\{-\mu(t_2 - t_1)\} \frac{[\mu(t_2 - t_1)]^{k_2 - k_1}}{(k_2 - k_1)!} \cdot \dots \cdot \exp\{-\mu(t_m - t_{m-1})\} \frac{[\mu(t_m - t_{m-1})]^{k_m - k_{m-1}}}{(k_m - k_{m-1})!} \\ & \exp\{-\mu t_m\} \mu^{k_m} \frac{t_1^{k_1} (t_2 - t_1)^{k_2 - k_1} \cdot \dots \cdot (t_m - t_{m-1})^{k_m - k_{m-1}}}{k_1! (k_2 - k_1)! \cdot \dots \cdot (k_m - k_{m-1})!}. \end{aligned}$$

Per quanto riguarda i tempi di arrivo possiamo immediatamente ottenere, in base al Corollario 4.2 del Capitolo 3, che la variabile aleatoria T_n segue una distribuzione $G(n, \mu)$.

Possiamo riottenere dunque l'espressione (41) per la funzione di distribuzione di $G(n, \mu)$ combinando insieme le formule (85), (91), (92).

6.2 Statistiche ordinate di variabili i.i.d. e partizioni casuali di intervalli

Nel corso del secondo capitolo avevamo osservato che la distribuzione congiunta di n variabili aleatorie è uniforme sul cubo unitario

$$\Gamma^{(n)} \equiv \{\mathbf{x} = (x_1, \dots, x_n) \mid 0 \leq x_j \leq 1\}$$

se e soltanto se esse sono indipendenti e identicamente distribuite con distribuzione uniforme su $R[0, 1]$.

Più in generale poniamo, per $t > 0$,

$$\Gamma_t^{(n)} = \{\mathbf{x} \equiv (x_1, \dots, x_n) \mid 0 \leq x_j \leq t\}$$

e osserviamo che la distribuzione congiunta di n variabili aleatorie è uniforme su $\Gamma_t^{(n)}$ se e soltanto se esse sono indipendenti e identicamente distribuite con distribuzione uniforme $R[0, t]$.

Indichiamo ora con $\Lambda_t^{(n)}$ e $\Delta_t^{(n)}$ le regioni

$$\Lambda_t^{(n)} \equiv \{\mathbf{x} \in \Gamma_t^{(n)} \mid x_1 \leq x_2 \leq \dots \leq x_n\},$$

$$\Delta_t^{(n)} \equiv \{\mathbf{x} \in \Gamma_t^{(n)} \mid \sum_{j=1}^n x_j \leq t\}.$$

Osservazione 2.1

La trasformazione $\tau : \Lambda_t^{(n)} \rightarrow \Delta_t^{(n)}$ definita dalle relazioni

$$\begin{aligned} x_1 &= t_1 \\ x_2 &= t_2 - t_1 \\ \dots &\dots \\ x_n &= t_n - t_{n-1} \end{aligned} \tag{95}$$

stabilisce una corrispondenza biunivoca fra i punti delle regioni $\Lambda_t^{(n)}$ e $\Delta_t^{(n)}$ e risulta

$$(t_1, \dots, t_n) \in \Lambda_t^{(n)} \Leftrightarrow (x_1, \dots, x_n) \in \Delta_t^{(n)}. \tag{96}$$

Osservazione 2.2

Per una generica permutazione $\pi = (\pi(1), \dots, \pi(n))$ di $\{1, \dots, n\}$, indichiamo con $\phi_\pi : \Gamma_t^{(n)} \rightarrow \Gamma_t^{(n)}$ la trasformazione definita dalla relazione

$$\phi_\pi(\mathbf{x}) \rightarrow (x_{\pi(1)}, \dots, x_{\pi(n)}).$$

Per brevità, in quanto segue, omettiamo gli apici (n) fintanto che si possa escludere la possibilità di equivoci. Poniamo inoltre

$$\Lambda_{t,\pi} = \phi_\pi(\Lambda_t)$$

ed indichiamo con $\lambda_n(B)$ il volume (misura di Peano-Jordan o di Lebesgue) nello spazio n -dimensionale di un dominio (misurabile) $B \subset \mathbb{R}^n$. ϕ_π è una trasformazione biunivoca ed ovviamente risulta

$$\lambda_n(\Gamma_t) = t^n, \Gamma_t = \bigcup_{\pi} \Lambda_{t,\pi},$$

$$\lambda_n(\Lambda_{t,\pi'}) = \lambda_n(\Lambda_{t,\pi''}), \lambda_n(\Lambda_{t,\pi'} \cap \Lambda_{t,\pi''}) = 0, \text{ per } \pi' \neq \pi''.$$

Ne segue

$$\lambda_n(\Lambda_t) = \lambda_n(\Lambda_{t,\pi}) = \frac{t^n}{n!}, \forall \pi.$$

Inoltre, essendo uguale ad 1 il determinante jacobiano della trasformazione τ in (95), vediamo anche che

$$\lambda_n(\Delta_t) = \lambda_n(\Lambda_t) = \frac{t^n}{n!}.$$

In quanto segue ci proponiamo di illustrare il significato di n -uple di variabili aleatorie aventi distribuzioni congiunte $\mathcal{U}_n(\Lambda_t)$ e $\mathcal{U}_n(\Delta_t)$, uniformi su Λ_t e Δ_t rispettivamente.

Sia \mathbf{Y} un vettore di variabili aleatorie reali Y_1, \dots, Y_n . Con il termine *statistiche d'ordine* (o *ordinate*) si indicano le n variabili

$$Y_{(1)}, \dots, Y_{(n)}$$

ottenute riordinando Y_1, \dots, Y_n in senso crescente; perciò, in particolare

$$Y_{(1)} = \min_{1 \leq i \leq n} Y_i, Y_{(n)} = \max_{1 \leq i \leq n} Y_i.$$

Proposizione 2.1 Se Y_1, \dots, Y_n sono indipendenti identicamente distribuite con distribuzione uniforme $R[0, t]$, allora

$$(Y_{(1)}, \dots, Y_{(n)}) \sim \mathcal{U}_n(\Lambda_t).$$

Dimostrazione

Osserviamo che, siccome $\mathbb{P}\{(Y_1, \dots, Y_n) \in \Gamma_t\} = 1$, risulta

$$\mathbb{P}\{(Y_{(1)}, \dots, Y_{(n)}) \in \Lambda_t\} = 1.$$

Inoltre, per ogni $B \subset \Lambda_t$ misurabile,

$$\mathbb{P}\{(Y_{(1)}, \dots, Y_{(n)}) \in B\} = \mathbb{P}\{(Y_1, \dots, Y_n) \in \bigcup_{\pi} \phi_\pi(B)\}.$$

D'altra parte risulta

$$\lambda_n \left(\bigcup_{\pi} \phi_{\pi}(B) \right) = n! \lambda_n(B)$$

e dunque, essendo

$$(Y_1, \dots, Y_n) \sim \mathcal{U}_n(\Gamma_t) \text{ e } \lambda_n(\Gamma_t) = t^n,$$

$$P\{(Y_1, \dots, Y_n) \in \bigcup_{\pi} \phi_{\pi}(B)\} = \frac{n!}{t^n} \lambda_n(B).$$

Si può allora concludere

$$\mathbb{P}\{(Y_{(1)}, \dots, Y_{(n)}) \in B\} = n! \lambda_n(B) = \frac{\lambda_n(B)}{\lambda_n(\Lambda_t)}.$$

Diremo che $Y_{(1)}, \dots, Y_{(n)}$ costituiscono una *partizione casuale* dell'intervallo $[0, t]$.

Prima di proseguire vogliamo ricavare, per $h = 1, 2, \dots, n$, la funzione di densità marginale unidimensionale $f_{Y_{(h)}}$ della variabile aleatoria $Y_{(h)}$. Per ottenere tale funzione di densità potremmo ricorrere alla applicazione della formula che in generale esprime una densità marginale in termini della congiunta, che in tal caso risulta dunque data da

$$f_{\mathbf{Y}_{(n)}}(y_1, \dots, y_n) = \frac{n!}{t^n} \mathbf{1}_{\Lambda_t}(y_1, \dots, y_n).$$

E' utile però qui ragionare in modo sintetico, tenendo presente il significato probabilistico di $f_{Y_{(h)}}$:

$$f_{Y_{(h)}}(y) = \lim_{\Delta y \rightarrow 0} \frac{P\{y \leq Y_{(h)} \leq y + \Delta y\}}{\Delta y}. \quad (97)$$

Tenendo conto del fatto che le variabili originarie Y_1, \dots, Y_n sono stocasticamente indipendenti (e uniformemente distribuite su $[0, t]$) e che quindi, per $1 \leq i \neq j \leq n$,

$$\mathbb{P}\{(y \leq Y_i \leq y + \Delta x) \cap (y \leq Y_j \leq y + \Delta y)\} = \frac{1}{t^2} (\Delta y)^2,$$

possiamo scrivere, per $0 \leq y \leq t$,

$$\mathbb{P}\{y \leq Y_{(h)} \leq y + \Delta y\} =$$

$$\mathbb{P}\{(Y_i \leq y, \text{ per } (h-1) \text{ indici } i) \cap (y \leq Y_r \leq y + \Delta y \text{ per un indice } r) \cap (Y_j > y + \Delta y \text{ per } (n-h) \text{ indici } j)\} + o(\Delta y) =$$

$$\frac{1}{t^n} n \binom{n-1}{h-1} y^{h-1} \Delta y (t-y)^{n-h} + o(\Delta y).$$

In base alla (97), possiamo dunque concludere

$$f_{Y_{(h)}}(y) = \frac{1}{t^n} \frac{n!}{(h-1)!(n-h)!} y^{h-1} (t-y)^{n-h}.$$

Oppure, ricordando la definizione di densità $beta(\alpha, \beta)$ vista nel Capitolo 3, **Proposizione 2.2**

$$\frac{Y_{(h)}}{t} \sim beta(h, n-h+1).$$

La funzione di ripartizione di $Y_{(h)}$ si può, a sua volta, ottenere quale primitiva della funzione di densità $f_{Y_{(h)}}$. Ma, anche in questo caso, conviene ottenere il risultato in modo sintetico, guardando direttamente al significato probabilistico: per $0 \leq y \leq t$,

$$F_{Y_{(h)}}(y) = P\{Y_{(h)} \leq y\} = P\{Y_i \leq y, \text{ per almeno } h \text{ indici } i\} =$$

$$\frac{1}{t^n} \sum_{k=h}^n \binom{n}{k} y^k (t-y)^{n-k}.$$

Passiamo ora a considerare il vettore delle $(n+1)$ variabili V_1, \dots, V_n definite dalla trasformazione

$$\mathbf{V} = \tau(\mathbf{Y}_{(\cdot)}),$$

cioè:

$$V_1 = Y_{(1)}, V_2 = Y_{(2)} - Y_{(1)}, \dots, V_n = Y_{(n)} - Y_{(n-1)}.$$

In virtù della relazione (96) e tenendo ancora conto che il determinante jacobiano della trasformazione τ in (95), è uguale ad 1, otteniamo

Proposizione 2.3

$$(V_1, \dots, V_n) \sim U_n(\Delta_t).$$

In altre parole, possiamo concludere che un vettore di variabili aleatorie (V_1, \dots, V_n) ammette una distribuzione congiunta uniforme sul dominio $\Delta_t^{(n)}$ se e soltanto se le variabili aleatorie

$$V_1, V_1 + V_2, \dots, V_1 + \dots + V_n$$

danno luogo ad una partizione casuale dell'intervallo $[0, t]$, cioè hanno la stessa distribuzione congiunta del vettore delle statistiche ordinate di variabile aleatorie i.i.d condistribuzione uniforme nell'intervallo $[0, t]$.

6.3 Processi di Poisson omogenei e partizioni casuali di intervalli

Sia X_1, X_2, \dots una successione di variabili aleatorie non negative e siano T_1, T_2, \dots le relative somme parziali:

$$T_1 = X_1, T_2 = X_1 + X_2, \dots$$

Il seguente risultato riguarda, per n fissato, la distribuzione congiunta di X_1, X_2, \dots, X_n condizionatamente alla conoscenza del valore assunto da T_{n+1} , nel caso in cui X_1, X_2, \dots, X_n vengano inizialmente assunte come i.i.d. esponenziali.

Proposizione 3.1

Siano X_1, X_2, \dots indipendenti, identicamente distribuite con $X_i \sim \mathcal{E}(\mu), \mu > 0$. Condizionatamente a $(T_{n+1} = t)$, la distribuzione condizionata di X_1, X_2, \dots, X_n è uniforme sull'insieme $\Delta_t^{(n)}$.

Dimostrazione

Dobbiamo ricavare innanzitutto la densità congiunta di $(X_1, X_2, \dots, X_n, T_{n+1})$. Consideriamo allora la trasformazione $\mathcal{V} : \mathbb{R}_+^{n+1} \rightarrow \mathbb{R}_+^{n+1}$ definita da

$$X_1 = x_1, X_2 = x_2, \dots, X_n = x_n, X_{n+1} = T_{n+1} - \sum_{j=1}^n x_j.$$

Tenendo conto che il determinante jacobiano di tale trasformazione è costante, ed uguale ad 1, e che risulta, per ipotesi,

$$f_{X_1, \dots, X_{n+1}}(x_1, \dots, x_{n+1}) = \mu^{n+1} \exp\{-\mu \sum_{j=1}^n x_j\},$$

otteniamo, per $\mathbf{x} \equiv (x_1, \dots, x_n) \in \Delta_t$

$$f_{X_1, \dots, X_n, T_{n+1}}(x_1, \dots, x_n, t) = \mu^{n+1} \exp\{-\mu t\}.$$

Ricordando inoltre che la distribuzione di T_{n+1} è $G(n+1, \mu)$, cioè che per $t > 0$

$$f_{T_{n+1}}(t) = \frac{\mu^{n+1}}{n!} t^n \exp\{-\mu t\},$$

possiamo concludere, per $\mathbf{x} \equiv (x_1, \dots, x_n) \in \Delta_t$,

$$\begin{aligned} f_{X_1, \dots, X_n}(x_1, \dots, x_n | T_{n+1} = t) &= \frac{f_{X_1, \dots, X_n, T_{n+1}}(x_1, \dots, x_n, t)}{f_{T_{n+1}}(t)} = \\ &= \frac{n!}{t^n}. \end{aligned}$$

Consideriamo ora un processo di Poisson $\{N_t\}_{t \geq 0}$, di intensità μ , e ricordiamo che i relativi istanti di arrivo possono essere visti come le somme parziali

$$T_1 = X_1, T_2 = X_1 + X_2, \dots, T_{n+1} = T_n + X_{n+1} = \sum_{j=1}^{n+1} X_j \dots$$

di variabili aleatorie indipendenti (gli “intertempi”)

$$X_1, X_2, \dots, X_n, X_{n+1}, \dots$$

identicamente distribuite con $X_i \sim \mathcal{E}(\mu)$. Dalla precedente proposizione possiamo allora ricavare il seguente

Corollario 3.2

La distribuzione condizionata di (T_1, \dots, T_n) dato $\{T_{n+1} = t\}$, è uniforme sull'insieme $\Lambda_t^{(n)}$.

Notiamo che, in un processo di conteggio, gli eventi

$$\{N_t = n\}, \{T_n \leq t < T_{n+1}\}$$

sono fra di loro equivalenti, e sono ovviamente implicati dall'evento $\{T_n = t\}$.

Fra le varie proprietà notevoli dei processi di Poisson, si ha però che la distribuzione condizionata di (T_1, \dots, T_n) dato $\{N_t = n\}$ coincide con la distribuzione condizionata di (T_1, \dots, T_n) dato $\{T_{n+1} = t\}$. Si ha infatti la seguente

Proposizione 3.3

La distribuzione condizionata di (T_1, \dots, T_n) dato $\{N_t = n\}$, è uniforme sull'insieme $\Lambda_t^{(n)}$.

Dimostrazione

Si tratta di dimostrare che, per ogni sottoinsieme misurabile $B \subset \Lambda_t^{(n)}$, risulta

$$\mathbb{P}\{(T_1, \dots, T_n) \in B | N_t = n\} = \frac{n!}{t^n} \lambda_n(B).$$

A tale fine osserviamo innanzitutto che, per $B \subset \Lambda_t^{(n)}$, vale l'implicazione

$$\{(T_1, \dots, T_n) \in B\} \Rightarrow \{T_n \leq t\},$$

da cui

$$\mathbb{P}\{((T_1, \dots, T_n) \in B) \cap (T_n \leq t < T_{n+1})\} = \mathbb{P}\{((T_1, \dots, T_n) \in B) \cap (T_{n+1} > t)\}$$

e dunque

$$\mathbb{P}\{(T_1, \dots, T_n) \in B | N_t = n\} = \frac{\mathbb{P}\{((T_1, \dots, T_n) \in B) \cap (T_{n+1} > t)\}}{\mathbb{P}\{N_t = n\}}.$$

Teniamo ora conto della forma della densità congiunta di $(T_1, \dots, T_n, T_{n+1})$:

$$f_{T_1, \dots, T_n, T_{n+1}}(t_1, \dots, t_n, \eta) = \mu^{n+1} \exp\{-\mu\eta\} 1_{\Lambda_\eta^{(n)}}.$$

Possiamo allora scrivere

$$\mathbb{P}\{(T_1, \dots, T_n) \in B | N_t = n\} =$$

$$\begin{aligned}
& \frac{\int_t^\infty \int \dots \int_B f_{T_1, \dots, T_n, T_{n+1}}(t_1, \dots, t_n, \eta) dt_1, \dots, dt_n d\eta}{\exp\{-\mu t\} \mu^n \frac{t^n}{n!}} = \\
& = \frac{n! \int_t^\infty \int \dots \int_B \mu^{n+1} \exp\{-\mu \eta\} dt_1, \dots, dt_n d\eta}{\exp\{-\mu t\} \mu^n} = \\
& = \frac{n! \mu^n \int_t^\infty \mu \exp\{-\mu \eta\} [\int \dots \int_B dt_1, \dots, dt_n] d\eta}{\exp\{-\mu t\} \mu^n} = \\
& = \frac{n!}{t^n} \lambda_n(B) \frac{\int_t^\infty \mu \exp\{-\mu \eta\} d\eta}{\exp\{-\mu t\}} = \frac{n!}{t^n} \lambda_n(B).
\end{aligned}$$

In virtù del precedente Corollario 3.2 la Proposizione 3.3 mostra quindi che, dato $\{N_t = n\}$, la distribuzione condizionata del vettore (T_1, \dots, T_n) coincide con la distribuzione congiunta delle statistiche ordinate di n variabili aleatorie i.i.d. con distribuzione uniforme sull'intervallo $[0, t]$. Tale proprietà è anche conosciuta con il nome di *proprietà di statistica ordinata* dei processi di Poisson.

Come immediata conseguenza delle Proposizioni 3.3 e 2.2 abbiamo

Corollario 3.4

Consideriamo ora, dato l'evento $(N_t = n)$, la distribuzione di probabilità (unidimensionale) condizionata della variabile N_s , per $0 < s < t$; vogliamo cioè, per $k = 0, 1, \dots, n$, calcolare

$$\mathbb{P}\{N_s = k | N_t = n\} = \mathbb{P}\{T_k \leq s < T_{k+1} | N_t = n\}.$$

Utilizzando la Formula di Bayes, si ottiene

$$\begin{aligned}
\mathbb{P}\{N_s = k | N_t = n\} &= \mathbb{P}\{N_s = k\} \frac{\mathbb{P}\{N_t = n | N_s = k\}}{\mathbb{P}\{N_t = n\}} = \\
&= \exp\{-\mu s\} \frac{s^k}{k!} \frac{n!}{\exp\{-\mu t\} t^n} \frac{\exp\{-\mu(t-s)\} (t-s)^{n-k}}{(n-k)!} = \\
&= \binom{n}{k} \left(\frac{s}{t}\right)^k \left(1 - \frac{s}{t}\right)^{n-k}.
\end{aligned}$$

Possiamo dunque concludere che la distribuzione condizionata di N_s dato $(N_t = n)$ è una distribuzione binomiale di parametri n e $p = \frac{s}{t}$. Notiamo però che allo stesso risultato si può anche giungere direttamente tramite un semplice ragionamento basato sulla proprietà di statistica ordinata del processo di Poisson.

Consideriamo ancora un processo di conteggio semplice $\{N_t\}_{t \geq 0}$. Lasciamo cadere l'ipotesi che sia un processo di Poisson omogeneo, cioè tale che

$$\lim_{\Delta t \rightarrow 0} \frac{P\{N_{t+\Delta t} = n+1 | N_t = n\}}{\Delta t} = \mu.$$

Ma, mantenendo l'ipotesi che tale limite non dipenda da n , ammettiamo la possibilità che esso risulti una funzione della variabile t : per $n = 0, 1, \dots$,

$$\lim_{\Delta t \rightarrow 0} \frac{P\{N_{t+\Delta t} = n + 1 | N_t = n\}}{\Delta t} = \mu(t).$$

Parleremo allora di processo di Poisson *non omogeneo*. Si può dimostrare ancora che la distribuzione condizionata di (T_1, T_2, \dots, T_n) dato $\{T_{n+1} = t\}$ coincide con la distribuzione congiunta delle statistiche ordinate di n variabili aleatorie i.i.d. con distribuzione concentrata sull'intervallo $[0, t]$, ma non più uniforme su tale intervallo.

Ad una diversa generalizzazione della nozione di processo di Poisson si giunge nel seguente caso: su uno stesso spazio di probabilità viene definita una successione di intertempi X_1, X_2, \dots ed un'ulteriore variabile aleatoria M , con una sua distribuzione marginale F_M ; si suppone che, *condizionatamente* a ciascuna ipotesi $\{M = \mu\}$, X_1, X_2, \dots siano indipendenti, identicamente distribuite, esponenziali con $X_i \sim \mathcal{E}(\mu)$, per ogni $\mu > 0$. Consideriamo poi il processo di conteggio $\{N_t\}_{t \geq 0}$ avente come intertempi X_1, X_2, \dots . Processi di conteggio di tale tipo vengono detti *Mixed Poisson*.

A meno che F_M non risulti essere una distribuzione degenera, X_1, X_2, \dots non sono stocasticamente indipendenti, e $\{N_t\}_{t \geq 0}$ non può risultare un processo di Poisson. Tale affermazione si spiega pensando, ad esempio, a due intervalli temporali $I \equiv [0, t]$ e $I' \equiv [t, t + \delta]$: il numero aleatorio N_t e il numero aleatorio $N_{t+\delta} - N_t$ -che contano quanti arrivi vi siano in I e I' rispettivamente- non possono essere fra loro indipendenti. Ciononostante, anche per il processo $\{N_t\}_{t \geq 0}$ si può sostanzialmente estendere la dimostrazione della Proposizione 3.3 e, come conseguenza diretta, si ha che i processi Mixed Poisson condividono con i processi di Poisson omogenei e non omogenei e la *proprietà di Markovianità*. Possiamo qui formulare tale proprietà come segue: la distribuzione condizionata della variabile $N_{t+\delta}$ dato $(N_t = n)$ coincide con la distribuzione condizionata di $N_{t+\delta}$ data un'informazione del tipo

$$(N_t = n; T_1 = t_1, \dots, T_n = t_n),$$

cioè data l'informazione completa circa l'evoluzione del processo fino al tempo t . Dopo aver osservato il processo dal tempo 0 al tempo t , è dunque soltanto il valore N_t quello che va registrato nella previsione dell'evoluzione futura, da t in poi.

Processi di conteggio di Poisson omogenei, non omogenei e Mixed Poisson intervengono ricorrentemente in numerose problematiche della probabilità applicata.