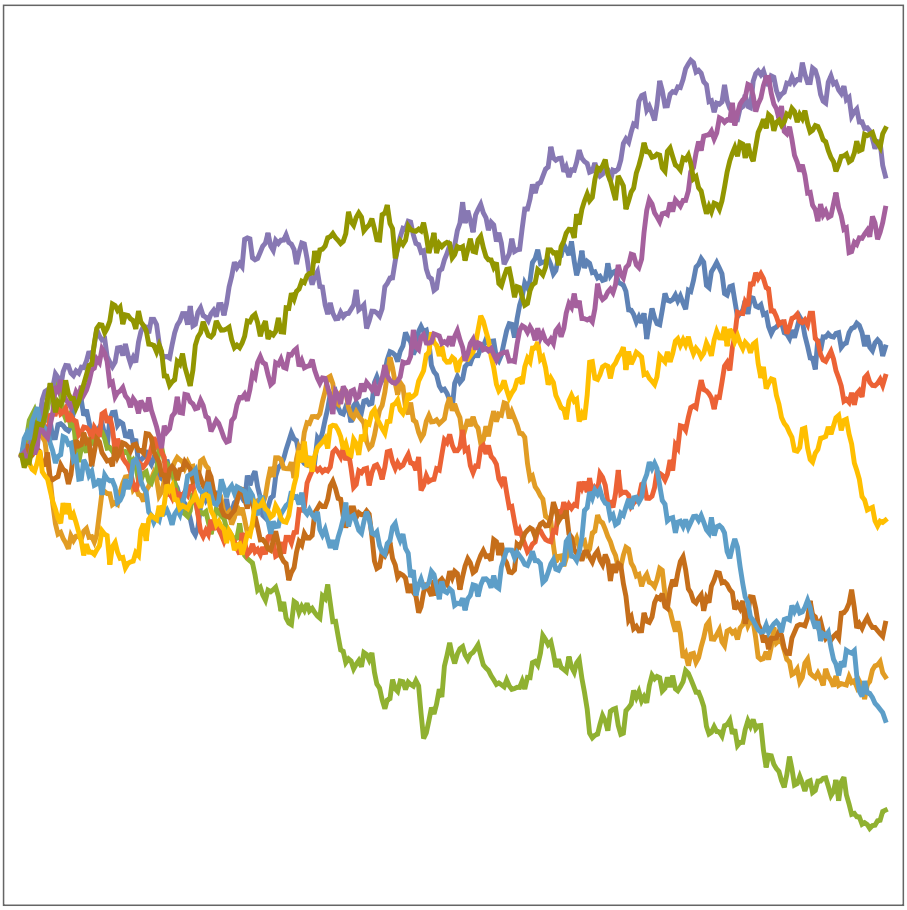
Fondamenti di Probabilità

Lucio Barabesi



Pagina intenzionalmente vuota

Capitolo 1

Eventi e classi di eventi

1.1. Nozioni preliminari

Si consideri un *esperimento* (o un *fenomeno*) nei confronti del quale ci si trova in condizioni di *incertezza*. L'incertezza è determinata dal fatto che l'esperimento (o il fenomeno) in questione sia suscettibile di verificarsi secondo una pluralità di risultati, tali che uno (e uno solo) di essi si deve necessariamente realizzare. In questo caso, l'esperimento (o il fenomeno) è detto *aleatorio*.

L'insieme Ω i cui elementi rappresentano i possibili risultati dell'esperimento aleatorio, è detto *spazio fondamentale* (o *insieme delle eventualità*) e il generico elemento ω di Ω è anche detto *eventualità*. Ogni parte E di Ω identifica un *evento* e il risultato ω realizza E se $\omega \in E$. Inoltre, un evento $\{\omega\}$ composto da un solo risultato è detto *evento elementare*. Infine, l'evento Ω viene anche detto *evento certo*, in quanto si realizza sempre, mentre l'insieme vuoto $\emptyset$ è detto *evento impossibile*.

Per quanto riguarda la *cardinalità* di Ω, vi sono tre principali situazioni di interesse. Nel primo caso, Ω è finito con $\text{card}(\Omega) = n$, ovvero contiene un numero finito n di risultati e dunque si ha $\Omega = \{\omega_1, \dots, \omega_n\}$. Nel secondo caso, Ω è numerabile, ovvero contiene una infinità numerabile di risultati e dunque $\Omega = \{\omega_1, \omega_2, \dots\}$. Infine, nel terzo caso, Ω è non numerabile, ovvero contiene un'infinità non numerabile di risultati.

• **Esempio 1.1.1.** (Testa e croce) Si consideri una moneta con le facce contrassegnate dai simboli "testa" e "croce", che viene comunemente adottata nel più semplice gioco d'azzardo, detto appunto *testa e croce*. Questa terminologia deriva dalle sagome che erano coniate su molte monete italiane dei secoli passati, in quanto una faccia usualmente raffigurava il volto del re, mentre l'altra riportava il simbolo cristiano della croce. Simili terminologie sono comuni a molte culture, con un nome generalmente derivante dalle raffigurazioni sulle facce dei conii. Ad esempio, nell'antica Roma i due simboli venivano denominati "navis" e "caput", dal momento che su alcune monete romane era rappresentata una nave su una faccia e la testa dell'imperatore sull'altra. Analogamente, l'uso nel mondo anglosassone dei termini "head" e "tail" per i simboli sulle due facce deriva probabilmente dalla moneta da dieci centesimi di sterlina, su cui erano rappresentate la faccia del monarca regnante, ed un leone araldico con una coda in evidenza.

Figura 1.1.1. Una lira del 1863 raffigurante i simboli "testa" e croce".

Com'è noto, il gioco consiste semplicemente nello scommettere su una delle due facce e nel lanciare in aria la moneta. La faccia vincente è quella mostrata dalla moneta dopo la caduta. Dunque, il gioco è in effetti un esperimento aleatorio. Se si indicano rispettivamente con ω_1 e ω_2 i risultati relativi al verificarsi delle facce contrassegnate dalla testa e dalla croce, lo spazio fondamentale è dato da

$\Omega = \{\omega_1, \omega_2\}$ e quindi Ω è finito con $\text{card}(\Omega) = 2$. Si osservi che esiste una certa arbitrarietà nella scelta della codifica dei risultati. Ad esempio, molti autori preferiscono indicare in modo più esplicito i due risultati solamente con i simboli t e c. In questo caso, lo spazio fondamentale viene scritto semplicemente come $\Omega = \{t, c\}$. □

• **Esempio 1.1.2.** (Gioco dei dadi) I dadi sono stati adottati a scopo di gioco d'azzardo fino dall'antichità. La loro origine sembra risalire a circa 5000 anni fa nella regione dell'attuale India. Riferimenti al gioco dei dadi sono contenuti nella Bibbia, e il gioco d'azzardo con tre dadi era molto popolare nell'antica Grecia e soprattutto nell'antica Roma. Nel Medioevo il gioco con i dadi era uno dei passatempi preferiti dei cavalieri. Molti problemi classici di calcolo delle probabilità si sono originati dal gioco dei dadi (si veda Gordon, 1997, Hald, 1990). Addirittura, in quello che può essere considerato il primo trattamento della Teoria della Probabilità, ovvero nel *Liber de Ludo Aleae* dell'italiano Gerolamo Cardano (1501-1576), il gioco dei dadi è centrale nell'esposizione. Cardano ebbe una vita avventurosa e molto travagliata, e fu anche un giocatore d'azzardo oltre che matematico, tanto che nel *Liber de Ludo Aleae* (scritto verso il 1560, ma pubblicato postumo solo nel 1663) esiste addirittura una sezione dedicata ai metodi per barare efficacemente (per una biografia di Cardano si veda Ore, 1953).

Figura 1.1.2. Gerolamo Cardano (1501-1576).

I dadi comunemente adottati nel gioco d'azzardo sono cubi con le facce marcate da k punti, dove $k = 1, \ldots, 6$, e le somme dei punti su due facce contrapposte sono pari a sette (si veda la Figura 1.1.3). Dadi di questo tipo sono stati adoperati anche dagli antichi etruschi, non solo per gioco, ma si suppone anche per divinazione.

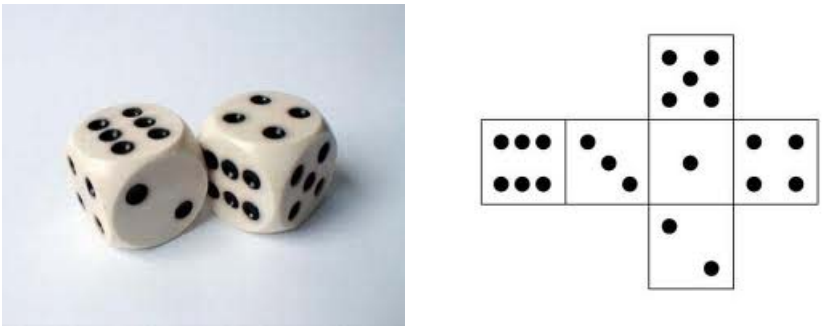


Figura 1.1.3. Comuni dadi da gioco e "sviluppo" di un singolo dado.

Se si considera un singolo dado, il gioco consiste nel far rotolare il dado su una superficie piana e convenzionalmente viene preso come esito del lancio il valore che si viene a trovare sulla faccia rivolta verso l'alto quando il dado termina il proprio movimento. Anche in questo caso il gioco è dunque un esperimento aleatorio. Se si indica con ω_k il risultato relativo al verificarsi della faccia contrassegnata da k punti, si ha $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4, \omega_5, \omega_6\}$ e quindi Ω è finito con $\text{card}(\Omega) = 6$.

Tenendo presente la discussione nell'Esempio 1.1.1, la codifica dello spazio fondamentale può essere data da $\Omega = \{1, 2, 3, 4, 5, 6\}$, interpretando tuttavia le cifre come simboli piuttosto che come numeri naturali. In questo caso, l'evento che si verifichi un numero pari di punti è dato da $E = \{2, 4, 6\}$, ovvero E è la parte di Ω costituita da tutti i risultati relativi al verificarsi di una faccia contrassegnata da un numero pari di punti. Inoltre, l'evento che si verifichi un numero di punti maggiore o uguale a 6 è dato da $E = \{6\}$, che è effettivamente un evento elementare e che non deve essere ingenuamente confuso con il risultato 6. Infine, l'evento che si verifichi un numero di punti maggiore di 8 è dato da $E = \emptyset$, ovvero E è un evento impossibile. □

• **Esempio 1.1.3.** Si consideri l'esperimento aleatorio che consiste nell'inserire in modo casuale m palline in M celle. In questo caso, gli elementi di Ω sono in corrispondenza biunivoca con tutte le possibili configurazioni di palline nelle celle. Se le palline sono distinguibili, allora Ω è finito con $\text{card}(\Omega) = M^m$, mentre se le palline sono indistinguibili, allora $\text{card}(\Omega) = \binom{M+m-1}{m}$. Al fine di rappresentare le varie configurazioni, almeno per valori modesti di m e M, si adottano di solito M spazi fra $(M + 1)$ barre per indicare le celle, mentre le palline vengono indicate come simboli differenti all'interno delle barre se sono distinguibili e con un asterisco se sono indistinguibili. Ad esempio, supponendo che $M = 3$ e $m = 2$, nel caso di palline distinguibili (che vengono indicate di seguito con i simboli a e b), si ha $\text{card}(\Omega) = 3^2 = 9$. Dunque, in modo leggermente informale anche se efficace, i possibili risultati dell'esperimento aleatorio possono essere rappresentati come

$$\begin{aligned} &\omega_1 = \,|\, ab \,|\, |\, |\,, \ \omega_2 = \,|\, |\, ab \,|\, |\,, \ \omega_3 = \,|\, |\, |\, ab \,|\,, \\ &\omega_4 = \,|\, a \,|\, b \,|\, |\,, \ \omega_5 = \,|\, a \,|\, |\, b \,|\,, \ \omega_6 = \,|\, |\, a \,|\, b \,|\,, \\ &\omega_7 = \,|\, b \,|\, a \,|\, |\,, \ \omega_8 = \,|\, b \,|\, |\, a \,|\,, \ \omega_9 = \,|\, |\, b \,|\, a \,|\,. \end{aligned}$$

Nel caso di palline indistinguibili si ha $\text{card}(\Omega) = \binom{4}{2} = 6$, e in questo caso si ottiene

$$\begin{aligned} &\omega_1 = \,|\, ** \,|\, |\, |\,, \ \omega_2 = \,|\, |\, ** \,|\, |\,, \ \omega_3 = \,|\, |\, |\, ** \,|\,, \\ &\omega_4 = \,|\, * \,|\, * \,|\, |\,, \ \omega_5 = \,|\, * \,|\, |\, * \,|\,, \ \omega_6 = \,|\, |\, * \,|\, * \,|\,. \end{aligned}$$

Si supponga che sia possibile inserire al più una pallina per cella. Assumendo che $m \leq M$, se le palline sono distinguibili si ha $\text{card}(\Omega) = \frac{M!}{(M-m)!}$, mentre se le palline sono indistinguibili allora $\text{card}(\Omega) = \binom{M}{m}$. Se pone $M = 3$ e $m = 2$, nel caso di palline distinguibili risulta $\text{card}(\Omega) = \frac{3!}{1!} = 6$, e i possibili risultati dell'esperimento aleatorio possono essere rappresentati come

$$\begin{aligned} &\omega_1 = \,|\, a \,|\, b \,|\, |\,, \ \omega_2 = \,|\, a \,|\, |\, b \,|\,, \ \omega_3 = \,|\, |\, a \,|\, b \,|\,, \\ &\omega_4 = \,|\, b \,|\, a \,|\, |\,, \ \omega_5 = \,|\, b \,|\, |\, a \,|\,, \ \omega_6 = \,|\, |\, b \,|\, a \,|\,. \end{aligned}$$

Al contrario, nel caso di palline indistinguibili si ha $\text{card}(\Omega) = \binom{3}{2} = 3$, e si ottiene

$$\omega_1 = \,|\, * \,|\, * \,|\, |\,, \ \omega_2 = \,|\, * \,|\, |\, * \,|\,, \ \omega_3 = \,|\, |\, * \,|\, * \,|\,.$$

Per i dettagli relativi al calcolo combinatorio relativo alle configurazioni di palline nelle celle del presente esempio e per le corrispondenti applicazioni alla meccanica quantistica, si veda Feller (1968, p.9) e gli Esercizi 1.1.1, 1.1.2, 1.1.3. e 1.1.4. □

• **Esempio 1.1.4.** Si consideri una moneta con le facce contrassegnate dai simboli "testa" e "croce" e l'esperimento aleatorio che consiste nel lanciare ripetutamente la moneta fino a quando non si presenta il simbolo "testa". Se si indica con ω_n il risultato relativo al verificarsi del primo simbolo "testa" all'n-esimo lancio, lo spazio fondamentale è dato da $\Omega = \{\omega_1, \omega_2, \dots\}$ e quindi Ω è numerabile. Tenendo presente la discussione fatta nell'Esempio 1.1.1, una codifica più immediata dello spazio fondamentale è data da $\Omega = \{1, 2, \dots\}$. Considerando una situazione specifica, l'evento E tale che si verifichi il primo simbolo testa ad un numero pari di lanci è dato da $E = \{2, 4, \dots\}$. Dunque, anche l'evento E è numerabile. □

• **Esempio 1.1.5.** Si consideri il tempo di attesa per un evento sismico in una determinata regione. Ogni risultato relativo a questo fenomeno aleatorio è del tipo "il tempo di attesa è pari a t unità temporali" con $t \geq 0$. Dunque, una formalizzazione dello spazio fondamentale è data da $\Omega = [0, \infty[$, per cui Ω non è numerabile. Considerando una situazione specifica, l'evento tale che il tempo di attesa sia compreso tra 1 e 3 unità temporali è $E = [1, 3]$ e quindi anche l'evento E non è numerabile. □

1.2. Classi di eventi e operazioni sugli eventi

Un insieme $\mathcal{E}$ costituito da eventi di Ω è detto *classe di eventi*. La classe costituita da tutti gli eventi di Ω, e dagli eventi $\emptyset$ e Ω, è detto *insieme delle parti* di Ω ed è indicato con la notazione $\mathcal{P}(\Omega)$. Una classe di eventi verrà talvolta indicata con $(E_k)_{k\in I}$, dove I è un opportuno insieme di indici, eventualmente non numerabile. Se $I = \{1, \dots, n\}$, per indicare la classe si adotta la notazione $(E_k)_{k=1}^n$. Se $I = \{1, 2, \dots\}$, la classe è detta *successione di eventi* e viene indicata con $(E_n)_{n\geq 1}$. Infine, se $I = \{0, 1, \dots\}$, per indicare la successione di eventi si adotta anche la notazione $(E_n)_{n\geq 0}$.

Dal momento che in effetti gli eventi sono insiemi, le usuali operazioni insiemistiche possono essere immediatamente considerate sugli eventi di Ω. Tuttavia, è opportuno dare a queste operazioni una lettura nel presente ambito con adeguate terminologie e definizioni. Dati due eventi E_1 e E_2, si dice che E_1 *implica* E_2 e si adotta la notazione $E_1 \subset E_2$, quando il verificarsi di E_1 comporta necessariamente il verificarsi di E_2, ovvero se e solo se un risultato che realizza E_1 è anche un risultato che realizza E_2. Inoltre, due eventi E_1 e E_2 sono detti *uguali* e si adotta la notazione $E_1 = E_2$, quando il verificarsi dell'uno implica il verificarsi dell'altro e *viceversa*, ovvero se si ha contemporaneamente $E_1 \subset E_2$ e $E_2 \subset E_1$. Infine, l'*evento opposto* (o *evento contrario*) di E è dato da

$$E^c = \{\omega \in \Omega : \omega \notin E\} ,$$

ovvero l'evento E^c si verifica quando non si verifica E. In particolare, si ha $\Omega^c = \emptyset$.

Data una classe di eventi $(E_k)_{k\in I}$ di Ω, l'*evento unione* è dato da

$$\bigcup_{k\in I} E_k = \{\omega \in \Omega : \exists k \in I, \omega \in E_k\} ,$$

ovvero l'evento $\bigcup_{k\in I} E_k$ si verifica quando si verifica almeno uno degli eventi della classe. Nel caso in cui la classe è costituita da n eventi $(E_k)_{k=1}^n$ di Ω, l'evento unione viene usualmente scritto come $\bigcup_{k=1}^n E_k$, e se $n = 2$ l'evento unione viene indicato semplicemente con la notazione $E_1 \cup E_2$. Infine, quando si considera una successione di eventi $(E_n)_{n\geq 1}$, l'evento unione viene indicato con $\bigcup_{n=1}^\infty E_n$.

Data una classe di eventi $(E_k)_{k\in I}$ di Ω, l'*evento intersezione* è dato da

$$\bigcap_{k\in I} E_k = \{\omega \in \Omega : \forall k \in I, \omega \in E_k\} ,$$

ovvero l'evento $\bigcap_{k\in I} E_k$ si verifica quando si verificano tutti gli eventi della classe. In modo simile a quanto fatto per l'evento unione, nel caso in cui la classe sia costituita da n eventi, l'evento intersezione viene scritto come $\bigcap_{k=1}^n E_k$, e se $n = 2$ si adotta la notazione $E_1 \cap E_2$. Infine, nel caso di una successione di eventi, l'evento intersezione viene indicato con $\bigcap_{n=1}^\infty E_n$.

Esistono due utili relazioni, che prendono nome dal matematico inglese Augustus De Morgan (1806-1871), autore fra l'altro del testo *An Essay on Probabilities* (1838), e che permettono di esprimere un evento intersezione come un evento unione (e *viceversa*), ovvero

$$\left(\bigcap_{k\in I} E_k\right)^c = \bigcup_{k\in I} E_k^c$$

e

$$\left(\bigcup_{k\in I} E_k\right)^c = \bigcap_{k\in I} E_k^c \, .$$

Dati due eventi E_1 e E_2, l'*evento differenza* (o *evento complementare* di E_2 rispetto ad E_1) è dato da

$$E_1 \setminus E_2 = \{\omega \in \Omega : \omega \in E_1, \omega \notin E_2\} \, ,$$

ovvero l'evento $E_1 \setminus E_2$ si verifica quando si verifica E_1 e non si verifica E_2. In particolare, si ha $E^c = \Omega \setminus E$, mentre risulta evidente la relazione

$$E_1 \setminus E_2 = E_1 \cap E_2^c \, .$$

Si noti che se $E_1 \subset E_2$, allora si ha $E_1 \setminus E_2 = \emptyset$, in quanto non è possibile che si verifichi E_1 e non si verifichi E_2 quando il verificarsi di E_1 implica il verificarsi di E_2. Inoltre, è ovvio che $E \setminus E = \emptyset$, in quanto è impossibile l'evento che si verifica quando contemporaneamente si verifica e non si verifica E. Infine, se $E_2 \subset E_1$, allora l'evento $E_1 \setminus E_2$ è detto *differenza propria*.

• **Esempio 1.2.1.** Si consideri l'esperimento aleatorio che consiste nel lancio di un dado. Dati gli eventi $E_1 = \{1, 2\}$ e $E_2 = \{1, 2, 3, 4\}$, si ha $E_1 \subset E_2$. Inoltre, risulta $E_1^c = \{3, 4, 5, 6\}$ e $E_2^c = \{5, 6\}$, mentre si ha $E_1 \cup E_2 = \{1, 2, 3, 4\}$ e $E_1 \cap E_2 = \{1, 2\}$. Dato l'ulteriore evento $E_3 = \{1, 5\}$, risulta $\bigcup_{k=1}^3 E_k = \{1, 2, 3, 4, 5\}$ e $\bigcap_{k=1}^3 E_k = \{1\}$. Infine, si ha $E_1 \setminus E_2 = \emptyset$, mentre $E_2 \setminus E_1 = \{3, 4\}$. □

Gli eventi della classe $(E_k)_{k=1}^n$ sono detti *incompatibili* quando il verificarsi di uno di essi esclude il verificarsi di tutti gli altri, ovvero quando $E_k \cap E_j = \emptyset$ per ogni $k \neq j = 1, \ldots, n$. Ovviamente, due eventi E_1 e E_2 sono incompatibili se $E_1 \cap E_2 = \emptyset$. In particolare, E e E^c sono eventi incompatibili, così come ciascun evento elementare è incompatibile con tutti gli altri eventi elementari di Ω. Inoltre, l'evento impossibile è incompatibile con tutti gli altri eventi. Infine, se E_1 e E_2 sono incompatibili, allora $E_1 \setminus E_2 = E_1$.

Gli eventi della classe $(E_k)_{k=1}^n$ sono detti *esaustivi* quando uno di essi si deve necessariamente verificare, ovvero quando $\bigcup_{k=1}^n E_k = \Omega$. Si dice che gli eventi della classe $(E_k)_{k=1}^n$ costituiscono una *partizione finita* di Ω quando sono incompatibili e esaustivi. In modo analogo, gli eventi della successione $(E_n)_{n\geq 1}$, sono detti *incompatibili* quando $E_n \cap E_k = \emptyset$ per ogni $n \neq k = 1, 2, \ldots$, mentre sono detti *esaustivi* quando $\bigcup_{n=1}^\infty E_n = \Omega$. Gli eventi della successione $(E_n)_{n\geq 1}$ costituiscono una *partizione numerabile* di Ω quando sono incompatibili e esaustivi.

• **Esempio 1.2.2.** Si consideri il lancio di un dado. Gli eventi $E_1 = \{1, 3\}$, $E_2 = \{2, 4, 5\}$ e $E_3 = \{6\}$ costituiscono una partizione finita di Ω. Gli eventi $E_1 = \{1, 3, 4\}$, $E_2 = \{2, 4, 5\}$ e $E_3 = \{4, 6\}$ sono esaustivi, anche se non costituiscono una partizione finita di Ω. □

Data una successione $(E_n)_{n\geq 1}$, si dice *limite inferiore* della successione l'evento

$$\liminf_n E_n = \bigcup_{n=1}^\infty \bigcap_{k=n}^\infty E_k \, ,$$

ovvero l'evento $\liminf_n E_n$ appartiene a tutti gli eventi della successione stessa ad esclusione al più di un numero finito di questi. Inoltre, si dice *limite superiore* della successione l'evento

$$\limsup_n E_n = \bigcap_{n=1}^\infty \bigcup_{k=n}^\infty E_k \, ,$$

ovvero l'evento $\limsup_n E_n$ appartiene ad una infinità numerabile di eventi della successione stessa. Se i due eventi $\liminf_n E_n$ e $\limsup_n E_n$ coincidono, allora si dice *limite* della successione l'evento

$$\lim_n E_n = \liminf_n E_n = \limsup_n E_n \ .$$

Inoltre, se $E_n \subset E_{n+1}$ per ogni $n = 1, 2, \ldots$, la successione è detta *crescente*, mentre se $E_{n+1} \subset E_n$ per ogni $n = 1, 2, \ldots$, la successione è detta *decrescente*. Successioni crescenti o decrescenti sono dette *monotone*. Dalle precedenti definizioni risulta facile verificare che una successione crescente ammette sempre limite, dato da

$$\lim_n E_n = \bigcup_{n=1}^{\infty} E_n \ ,$$

così come una successione decrescente ammette sempre limite, dato da

$$\lim_n E_n = \bigcap_{n=1}^{\infty} E_n \ .$$

• **Esempio 1.2.3.** Si supponga di considerare un esperimento aleatorio che consiste nello scegliere casualmente un punto su un segmento di lunghezza unitaria. In questo caso, lo spazio fondamentale associato all'esperimento aleatorio è dato da $\Omega = [0, 1]$. Si consideri la successione $(E_n)_{n\geq 1}$ il cui generico evento E_n risulta $E_n = [0, \frac{1}{4}(1 - (-1)^n)]$. Dunque, si ha $E_n = [0, \frac{1}{2}]$ se n è dispari, mentre $E_n = \{0\}$ se n è pari, da cui

$$\liminf_n E_n = \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} E_k = \bigcup_{n=1}^{\infty} \{0\} = \{0\}$$

e

$$\limsup_n E_n = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} E_k = \bigcap_{n=1}^{\infty} [0, \tfrac{1}{2}] = [0, \tfrac{1}{2}] \ .$$

Non esiste dunque il limite della successione in quanto $\liminf_n E_n$ e $\limsup_n E_n$ non coincidono. Se si considera invece la successione decrescente $(E_n)_{n\geq 1}$ il cui generico evento risulta $E_n = [0, n^{-1}]$, il suo limite è dato da

$$\lim_n E_n = \bigcap_{n=1}^{\infty} E_n = \{0\} \ ,$$

in quanto, verificandosi l'evento elementare $\{0\}$, si verifica ciascun evento E_n della successione. Se si considera invece la successione crescente $(E_n)_{n\geq 1}$ il cui generico evento è $E_n = [0, \frac{1}{2}(1 - n^{-1})]$, il suo limite risulta

$$\lim_n E_n = \bigcup_{n=1}^{\infty} E_n = [0, \tfrac{1}{2}] \ ,$$

in quanto, se si verifica l'evento $[0, \frac{1}{2}]$, si verifica almeno un evento E_n della successione. □

Si noti che le operazioni definite per gli eventi possono essere estese in modo generale alle classi di eventi, dal momento che in effetti classi di eventi costituiscono insiemi di insiemi. In particolare, date due classi di eventi $\mathcal{E}_1$ e $\mathcal{E}_2$, si dice che $\mathcal{E}_1$ è *contenuta* in $\mathcal{E}_2$ e si adotta la notazione $\mathcal{E}_1 \subset \mathcal{E}_2$, se e solo se per ogni evento $E \in \mathcal{E}_1$ si ha $E \in \mathcal{E}_2$. Due classi di eventi $\mathcal{E}_1$ e $\mathcal{E}_2$ sono dette *uguali* e si adotta la

notazione $\mathcal{E}_1 = \mathcal{E}_2$, quando si ha contemporaneamente $\mathcal{E}_1 \subset \mathcal{E}_2$ e $\mathcal{E}_2 \subset \mathcal{E}_1$. Se $(\mathcal{E}_k)_{k\in I}$ rappresenta una classe di classi di eventi, dove I è un insieme eventualmente non numerabile, la rispettiva *unione* è definita come

$$\bigcup_{k\in I} \mathcal{E}_k = \{E \in \mathcal{P}(\Omega) : \exists k \in I, E \in \mathcal{E}_k\} ,$$

mentre la rispettiva *intersezione* è definita come

$$\bigcap_{k\in I} \mathcal{E}_k = \{E \in \mathcal{P}(\Omega) : \forall k \in I, E \in \mathcal{E}_k\} .$$

Se $I = \{1, 2, \dots\}$, in modo simile a quanto fatto per le successioni di eventi, si adotta la notazione $(\mathcal{E}_n)_{n\geq 1}$, mentre l'unione e l'intersezione vengono indicate rispettivamente con $\bigcup_{n=1}^{\infty}\mathcal{E}_n$ e $\bigcap_{n=1}^{\infty}\mathcal{E}_n$.

1.3. σ-algebre

Considerato un esperimento aleatorio, anche se ogni parte di Ω può essere interpretata come un evento, certe parti potrebbero non essere interessanti ai fini di un determinato problema, oppure potrebbero essere troppo complicate per essere analizzate. In ogni caso specifico, è conveniente dunque scegliere una classe di eventi non vuota, in modo che possieda caratteristiche di stabilità rispetto alle principali operazioni insiemistiche. Una importante classe di eventi che gode di questa proprietà è la cosiddetta σ-algebra.

Definizione 1.3.1. Si dice *σ-algebra* di eventi una classe $\mathcal{F}$ non vuota di eventi di Ω tale che:
i) se $E \in \mathcal{F}$, allora $E^c \in \mathcal{F}$;
ii) se $(E_n)_{n\geq 1}$ è una successione di eventi di $\mathcal{F}$, allora $\bigcup_{n=1}^{\infty} E_n \in \mathcal{F}$. □

In pratica la σ-algebra è una classe di eventi "chiusa" rispetto alle operazioni di complementazione e unione numerabile. Dunque, esempi immediati di σ-algebre sono la classe $\mathcal{P}(\Omega)$, oppure la classe $\{\emptyset, \Omega\}$.

• **Esempio 1.3.1.** Si consideri lo spazio fondamentale $\Omega = \{\omega_1, \omega_2, \omega_3\}$. Le possibili σ-algebre che si possono costruire da Ω sono date da

$$\mathcal{F}_1 = \{\emptyset, \Omega\} , \mathcal{F}_2 = \{\emptyset, \{\omega_1\}, \{\omega_2, \omega_3\}, \Omega\} , \mathcal{F}_3 = \{\emptyset, \{\omega_2\}, \{\omega_1, \omega_3\}, \Omega\} ,$$

$$\mathcal{F}_4 = \{\emptyset, \{\omega_3\}, \{\omega_1, \omega_2\}, \Omega\} , \mathcal{F}_5 = \{\emptyset, \{\omega_1\}, \{\omega_2\}, \{\omega_3\}, \{\omega_1, \omega_2\}, \{\omega_1, \omega_3\}, \{\omega_2, \omega_3\}, \Omega\} .$$

Si noti che $\mathcal{F}_1$ è contenuta in tutte le altre σ-algebre, mentre $\mathcal{F}_5 = \mathcal{P}(\Omega)$ contiene tutte le altre. □

Proposizione 1.3.2. *Ogni σ-algebra $\mathcal{F}$ contiene la σ-algebra $\{\emptyset, \Omega\}$.*
Dimostrazione. Dal momento che ogni σ-algebra è non vuota, esiste almeno un evento $E \in \mathcal{F}$. Allora, dalla Definizione 1.3.1, si ha anche $E^c \in \mathcal{F}$ e $E \cup E^c = \Omega \in \mathcal{F}$. Dunque, $\Omega^c = \emptyset \in \mathcal{F}$. □

Tenendo presente la precedente Proposizione, la σ-algebra $\{\emptyset, \Omega\}$ è detta la più piccola σ-algebra su Ω. Al contrario, dal momento che per ogni σ-algebra $\mathcal{F}$ su Ω si ha $\mathcal{F} \subset \mathcal{P}(\Omega)$, allora $\mathcal{P}(\Omega)$ è detta la più grande σ-algebra su Ω.

Proposizione 1.3.3. *Se $(E_n)_{n\geq 1}$ è una successione di eventi di $\mathcal{F}$, allora $\bigcap_{n=1}^{\infty} E_n \in \mathcal{F}$.*

Dimostrazione. Dal momento che $(E_n^c)_{n\geq 1}$ è una successione di eventi di $\mathcal{F}$, si ha $\bigcup_{n=1}^{\infty} E_n^c \in \mathcal{F}$ e di conseguenza $(\bigcup_{n=1}^{\infty} E_n^c)^c \in \mathcal{F}$. Dalla relazione di De Morgan si ha $\bigcap_{n=1}^{\infty} E_n = (\bigcup_{n=1}^{\infty} E_n^c)^c$, da cui segue la tesi.

□

Proposizione 1.3.4. *Se $E_1, E_2 \in \mathcal{F}$, allora $E_1 \setminus E_2 \in \mathcal{F}$.*

Dimostrazione. Dal momento che $E_1, E_2^c \in \mathcal{F}$, anche $E_1 \cap E_2^c \in \mathcal{F}$. La tesi segue dalla relazione $E_1 \setminus E_2 = E_1 \cap E_2^c$. □

Attraverso le precedenti Proposizioni si deduce che la σ-algebra è una classe di eventi "chiusa" rispetto alle principali operazioni insiemistiche considerate nella Sezione 1.2, ovvero rispetto alla complementazione, all'unione e all'intersezione numerabile, e alla differenza. Inoltre, se $(E_n)_{n\geq 1}$ è una successione di eventi di $\mathcal{F}$ anche gli eventi $\liminf_n E_n$, $\limsup_n E_n$ e $\lim_n E_n$ appartengono alla σ-algebra, in quanto sono definiti tramite unioni e intersezioni numerabili di eventi.

Proposizione 1.3.5. *Se $(\mathcal{F}_k)_{k\in I}$ è una classe di σ-algebre, allora l'intersezione $\mathcal{F} = \bigcap_{k\in I} \mathcal{F}_k$ è una σ-algebra.*

Dimostrazione. L'intersezione è non vuota, dal momento che contiene almeno la σ-algebra $\{\emptyset, \Omega\}$. Se $E \in \mathcal{F}$, allora $E \in \mathcal{F}_k$ per ogni $k \in I$. Dunque, $E^c \in \mathcal{F}_k$ per ogni $k \in I$ e quindi $E^c \in \mathcal{F}$. In modo simile, se $(E_n)_{n\geq 1} \in \mathcal{F}$, allora $E_n \in \mathcal{F}_k$ per ogni $n \geq 1$ e $k \in I$. Dunque, $\bigcup_{n=1}^{\infty} E_n \in \mathcal{F}_k$ per ogni $k \in I$ e quindi $\bigcup_{n=1}^{\infty} E_n \in \mathcal{F}$. □

Anche se la precedente Proposizione assicura che l'intersezione di σ-algebre è ancora una σ-algebra, si noti che l'unione $\bigcup_{k\in I} \mathcal{F}_k$ non è in generale una σ-algebra.

Nella Teoria della Probabilità risulta fondamentale considerare inizialmente una classe di eventi con una struttura semplice con proprietà minimali e successivamente considerare una σ-algebra che la contiene. Più esattamente, data una classe di eventi $\mathcal{E}$ su Ω, si dice *σ-algebra generata* da $\mathcal{E}$, e si indica con $\sigma(\mathcal{E})$, la più piccola σ-algebra che contiene $\mathcal{E}$. Dunque, se $(\mathcal{F}_k)_{k\in I}$ è la classe di σ-algebre tali che $\mathcal{E} \subset \mathcal{F}_k$, allora

$$\sigma(\mathcal{E}) = \bigcap_{k\in I} \mathcal{F}_k \, .$$

Evidentemente $\sigma(\mathcal{E})$ contiene, oltre a tutti gli eventi di $\mathcal{E}$, anche tutti gli eventi che si possono ottenere al più da un insieme numerabile di operazioni effettuate sugli eventi di $\mathcal{E}$.

• **Esempio 1.3.2.** Si consideri di nuovo l'Esempio 1.3.1. Data la classe di eventi $\mathcal{E} = \{\emptyset, \{\omega_2, \omega_3\}\}$, allora $\mathcal{E} \subset \mathcal{F}_2$ e $\mathcal{E} \subset \mathcal{F}_5$. Quindi si ha

$$\sigma(\mathcal{E}) = \bigcap_{k\in\{2,5\}} \mathcal{F}_k = \mathcal{F}_2 \, .$$

Si noti inoltre che

$$\bigcup_{k\in\{2,3\}} \mathcal{F}_k = \{\emptyset, \{\omega_1\}, \{\omega_2\}, \{\omega_1, \omega_3\}, \{\omega_2, \omega_3\}, \Omega\}$$

non è una σ-algebra. Ad esempio, la precedente classe non contiene l'evento

$$E = \{\omega_1, \omega_3\} \cap \{\omega_2, \omega_3\} = \{\omega_3\} \, .$$

□

1.4. Spazi probabilizzabili

Se $\mathcal{F}$ è una σ-algebra su Ω, la coppia $(\Omega, \mathcal{F})$ è detta *spazio probabilizzabile*. Uno spazio probabilizzabile può essere costruito a partire da una classe iniziale $\mathcal{E}$ di eventi di Ω, in modo tale da considerare la σ-algebra $\mathcal{F}$ generata da $\mathcal{E}$, ovvero $\mathcal{F} = \sigma(\mathcal{E})$. Se la classe $\mathcal{E}$ è scelta in maniera appropriata, $\sigma(\mathcal{E})$ soddisfa qualsiasi esigenza di carattere pratico, nel senso che contiene tutti gli eventi sui quali normalmente si opera. In ogni caso, sussiste una certa arbitrarietà nella scelta della classe iniziale, in quanto dipende dal criterio con il quale si considerano alcuni eventi interessanti e altri eventi non interessanti.

Se $\Omega = \{\omega_1, \dots, \omega_n\}$ è finito, si può scegliere la classe degli eventi elementari $\mathcal{E} = (\{\omega_k\})_{k=1}^n$ come classe iniziale. In questo caso, $\mathcal{F} = \sigma(\mathcal{E}) = \mathcal{P}(\Omega)$. Risulta facile verificare che $\mathcal{F}$ contiene 2^n eventi, ovvero $\text{card}(\mathcal{F}) = 2^n$. In modo analogo, se $\mathcal{E} = (E_k)_{k=1}^n$ è una partizione finita di Ω, allora $\mathcal{F} = \sigma(\mathcal{E})$ e si ha di nuovo $\text{card}(\mathcal{F}) = 2^n$.

• **Esempio 1.4.1.** Se si considera il lancio di un dado, allora si può scegliere $\mathcal{E} = (\{k\})_{k=1}^6$, in modo tale che si ha $\mathcal{F} = \mathcal{P}(\Omega)$ con $\text{card}(\mathcal{F}) = 2^6 = 64$. Tuttavia, se l'interesse dell'esperimento aleatorio è focalizzato solamente sul verificarsi di una faccia pari o dispari, si potrebbero considerare solamente gli eventi $E_1 = \{1, 3, 5\}$ e $E_2 = \{2, 4, 6\}$, in modo tale che $\mathcal{E} = \{E_1, E_2\}$ costituisce una partizione finita di Ω. In questo caso, si ha

$$\mathcal{F} = \sigma(\mathcal{E}) = \{\emptyset, E_1, E_2, \Omega\} ,$$

con $\text{card}(\mathcal{F}) = 2^2 = 4$. □

• **Esempio 1.4.2.** Un esperimento aleatorio consiste nello spezzare casualmente un segmento di lunghezza unitaria in due parti. Anche se in questo caso lo spazio fondamentale $\Omega = [0, 1]$ contiene una infinità non numerabile di eventi elementari, si può considerare gli eventi $E_1 = [0, \frac{1}{3}[$, $E_2 = [\frac{1}{3}, \frac{2}{3}[$ e $E_3 = [\frac{2}{3}, 1]$, che rappresentano rispettivamente l'evento che il segmento sia stato spezzato nella parte inferiore, centrale o superiore. Evidentemente, $\mathcal{E} = \{E_1, E_2, E_3\}$ costituisce una partizione finita di Ω, per cui

$$\mathcal{F} = \sigma(\mathcal{E}) = \{\emptyset, E_1, E_2, E_3, E_1 \cup E_2, E_1 \cup E_3, E_2 \cup E_3, \Omega\} .$$

Tuttavia, la ricchezza di eventi contenuta in Ω non viene effettivamente considerata a causa della estrema semplicità della classe iniziale. □

Se Ω è numerabile, ovvero $\Omega = \{\omega_1, \omega_2, \dots\}$, si può scegliere di nuovo la classe degli eventi elementari $\mathcal{E} = (\{\omega_n\})_{n \geq 1}$ come classe iniziale. Dunque, si ha ancora $\mathcal{F} = \sigma(\mathcal{E}) = \mathcal{P}(\Omega)$.

Se Ω è non numerabile, allora non è opportuno considerare l'insieme delle parti $\mathcal{P}(\Omega)$, in quanto questa classe risulta troppo ampia da probabilizzare (si veda il Capitolo 3). Quindi, ci si deve limitare a scegliere una adeguata classe iniziale, in modo tale che $\sigma(\mathcal{E})$ contenga gli eventi di interesse per l'analisi dell'esperimento aleatorio.

• **Esempio 1.4.3.** Si consideri lo spazio fondamentale $\Omega = \mathbb{R}$. In questo caso, la classe iniziale $\mathcal{E}$ può essere scelta come la classe degli eventi del tipo $]a, b]$. Se si considera dunque la σ-algebra $\mathcal{F}$ generata da $\mathcal{E}$, ovvero $\mathcal{F} = \sigma(\mathcal{E})$, è opportuno analizzare alcune categorie di eventi che appartengono a $\mathcal{F}$. La successione di eventi $(E_n)_{n \geq 1}$, dove $E_n =]a - n^{-1}, b]$ appartiene a $\mathcal{F}$, ed essendo questa successione decrescente, il suo limite è dato da

$$\lim_n E_n = \bigcap_{n=1}^{\infty}]a - n^{-1}, b] = [a, b] .$$

Quindi, $\mathcal{F}$ contiene gli eventi del tipo $[a, b]$. Dal momento che

$$\bigcap_{n=1}^{\infty}]a - n^{-1}, a] = \{a\} ,$$

la σ-algebra $\mathcal{F}$ contiene anche tutti gli eventi elementari. Inoltre, la successione di eventi $(E_n)_{n\geq 1}$, dove $E_n =]a - n, b]$ appartiene a $\mathcal{F}$, ed essendo questa successione crescente, il suo limite è dato da

$$\lim_n E_n = \bigcup_{n=1}^{\infty}]a - n, b] =]-\infty, b] .$$

Quindi, $\mathcal{F}$ contiene eventi del tipo $]-\infty, b]$. In generale, si può verificare che la σ-algebra $\mathcal{F}$ generata da $\mathcal{E}$ contiene tutti gli eventi del tipo $]a, b[$, $]a, b]$, $[a, b[$ e $[a, b]$, gli eventi del tipo $]-\infty, b]$, $]-\infty, b[$, $[a, \infty[$ e $]a, \infty[$, oltre a tutti gli eventi elementari. Inoltre, $\mathcal{F}$ contiene tutti gli eventi ottenibili da un insieme numerabile di operazioni insiemistiche su questi eventi. □

Nel linguaggio della Teoria della Misura, la σ-algebra discussa nell'Esempio 1.4.3 è detta *σ-algebra di Borel*, che prende nome dal matematico francese Émile Borel (1871-1956), uno dei pionieri della Teoria della Misura e delle sue applicazioni alla Teoria della Probabilità. Generalmente, la σ-algebra di Borel è indicata con $\mathcal{B}(\mathbb{R})$ e i suoi elementi sono detti *Boreliani*. Inoltre, la coppia $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ è uno *spazio misurabile*. Si può dimostrare che la costruzione della σ-algebra di Borel può anche essere fatta a partire dalla classe degli insiemi aperti di $\mathbb{R}$, dal momento che ogni insieme aperto di $\mathbb{R}$ può essere espresso come unione numerabile di insiemi del tipo $]a, b[$ (si veda Dudley, 2004, p.98). Questa costruzione è equivalente a quelle fatte partendo da una qualsiasi classe di insiemi del tipo $]a, b[$, $]a, b]$, $[a, b[$ e $[a, b]$, o da una qualsiasi classe di insiemi del tipo $]-\infty, b]$, $]-\infty, b[$, $[a, \infty[$ e $]a, \infty[$. Se $\mathbb{T} \subset \mathbb{R}$, si adotta anche la notazione $\mathcal{B}(\mathbb{T})$ per indicare la σ-algebra i cui elementi sono del tipo $B \cap \mathbb{T}$ con $B \in \mathcal{B}(\mathbb{R})$. Si noti infine che la costruzione di insiemi che non sono Boreliani è molto laboriosa (si veda Dudley, 2004, p.496) e dunque $\mathcal{B}(\mathbb{R})$ soddisfa pienamente alle esigenze di carattere pratico.

Figura 1.4.1. Émile Borel (1871-1956).

1.5. Prodotto cartesiano di spazi fondamentali

Si considerino n esperimenti aleatori, a ciascuno dei quali risulta associato lo spazio fondamentale Ω_k, dove $k = 1, \ldots, n$, e si combinino in modo da ottenere un unico esperimento aleatorio. In questo caso, lo *spazio fondamentale prodotto* risulta essere il prodotto cartesiano

$$\Omega = \Omega_1 \times \cdots \times \Omega_n ,$$

ovvero gli elementi di Ω sono tutte le possibili n-ple $(\omega_1, \dots, \omega_n)$ tali che $\omega_1 \in \Omega_1, \dots, \omega_n \in \Omega_n$. In questo caso, se $E_1 \subset \Omega_1, \dots, E_n \subset \Omega_n$, si dice *evento rettangolare* un evento E dello spazio fondamentale prodotto Ω i cui elementi risultano essere tutte le possibili n-ple $(\omega_1, \dots, \omega_n)$ con $\omega_1 \in E_1, \dots, \omega_n \in E_n$, ovvero

$$E = E_1 \times \cdots \times E_n \, .$$

Inoltre, l'evento $E_k \subset \Omega_k$ è detto *evento proiezione* dell'evento rettangolare E su Ω_k. Infine, se J è una scelta di indici di $I = \{1, \dots, n\}$, un evento $E \subset \Omega$ è detto *evento cilindrico* se si verificano gli eventi $E_k \subset \Omega_k$ con $k \in J$, ovvero

$$E = F_1 \times \cdots \times F_n \, ,$$

dove $F_k = E_k$ se $k \in J$, mentre $F_k = \Omega_k$ se $k \in I \setminus J$.

Se $(\Omega_1, \mathcal{F}_1), \dots, (\Omega_n, \mathcal{F}_n)$ sono n spazi probabilizzabili corrispondenti agli n esperimenti aleatori, la *σ-algebra prodotto* $\mathcal{F}$ relativa allo spazio fondamentale prodotto Ω è la σ-algebra generata dalla classe dagli eventi rettangolari di Ω e si scrive

$$\mathcal{F} = \mathcal{F}_1 \otimes \cdots \otimes \mathcal{F}_n \, .$$

La coppia $(\Omega, \mathcal{F})$ è detta *spazio probabilizzabile prodotto*. Inoltre, per quanto riguarda la costruzione dello spazio probabilizzabile $(\Omega, \mathcal{F})$, se Ω è finito o numerabile, allora si può scegliere $\mathcal{F} = \mathcal{P}(\Omega)$. Se invece almeno uno degli spazi fondamentali Ω_k è non numerabile, allora anche Ω è non numerabile. In questo caso, $\mathcal{F}$ può essere scelta come la σ-algebra generata dagli eventi rettangolari di Ω.

Nella Teoria della Misura si considera usualmente lo spazio prodotto $\mathbb{R}^n = \mathbb{R} \times \cdots \times \mathbb{R}$. In questo caso, la σ-algebra di Borel su $\mathbb{R}^n$ è data da $\mathcal{B}(\mathbb{R}^n) = \mathcal{B}(\mathbb{R}) \otimes \cdots \otimes \mathcal{B}(\mathbb{R})$ e può essere costruita a partire dagli insiemi rettangolari di $\mathbb{R}^n$ del tipo $]a_1, b_1] \times \cdots \times]a_n, b_n]$ (o in modo simile a quanto visto nella Sezione 1.4, dal prodotto cartesiano di ogni tipo di intervallo o semiretta). Nel linguaggio della Teoria della Misura, la coppia $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ è uno *spazio misurabile prodotto*. Infine, se $\mathbb{T} \subset \mathbb{R}^n$, si adotta la notazione $\mathcal{B}(\mathbb{T})$ per indicare la σ-algebra i cui elementi sono del tipo $B \cap \mathbb{T}$ con $B \in \mathcal{B}(\mathbb{R}^n)$.

In generale, se si considera una classe di spazi fondamentali $(\Omega_k)_{k \in I}$, dove I è un insieme eventualmente non numerabile, allora lo spazio fondamentale prodotto è dato da

$$\Omega = \prod_{k \in I} \Omega_k \, ,$$

ovvero in questo caso gli elementi di Ω risultano essere tutte le famiglie del tipo $(\omega_k)_{k \in I}$ tali che $\omega_k \in \Omega_k$ per ogni $k \in I$. Inoltre, si dice *evento rettangolare* un evento E dello spazio fondamentale prodotto Ω i cui elementi risultano essere tutte le possibili famiglie $(\omega_k)_{k \in I}$ tali che $\omega_k \in E_k$ per ogni $k \in I$, ovvero

$$E = \prod_{k \in I} E_k \, .$$

Infine, se J è una scelta di un numero finito di indici di I, un evento $E \subset \Omega$ è detto *evento cilindrico* se si verificano gli eventi $E_k \subset \Omega_k$ con $k \in J$, ovvero

$$E = \prod_{k \in I} F_k \, ,$$

dove $F_k = E_k$ se $k \in J$, mentre $F_k = \Omega_k$ se $k \in I \setminus J$.

Se $((\Omega_k, \mathcal{F}_k))_{k \in I}$ è la classe di spazi probabilizzabili corrispondente alla classe di esperimenti aleatori, la *σ-algebra prodotto* $\mathcal{F}$ sullo spazio fondamentale prodotto Ω è la σ-algebra generata dalla classe degli eventi cilindrici di Ω e viene scritta come

$$\mathcal{F} = \bigotimes_{k \in I} \mathcal{F}_k \, .$$

Anche in questo caso, la coppia $(\Omega, \mathcal{F})$ è detta *spazio probabilizzabile prodotto*.

• **Esempio 1.5.1.** Si considerino due lanci di una moneta. Dal momento che ogni lancio è un esperimento aleatorio che ha come spazio fondamentale $\Omega_k = \{t, c\}$, con $k = 1, 2$ (dove al solito t e c rappresentano gli esiti relativi al verificarsi delle due facce della moneta), lo spazio fondamentale prodotto relativo all'esperimento aleatorio combinato risulta

$$\Omega = \Omega_1 \times \Omega_2 = \{(t,t), (t,c), (c,t), (c,c)\} \, .$$

La σ-algebra prodotto è data da $\mathcal{F} = \mathcal{F}_1 \otimes \mathcal{F}_2 = \mathcal{P}(\Omega)$, dove $\mathcal{F}_1 = \mathcal{P}(\Omega_1)$ e $\mathcal{F}_2 = \mathcal{P}(\Omega_2)$, con $\text{card}(\mathcal{F}) = 2^4 = 16$. Dunque, risulta

$$\begin{aligned} \mathcal{F} = \{&\emptyset, \{(t,t)\}, \{(t,c)\}, \{(c,t)\}, \{(c,c)\}, \{(t,t),(t,c)\}, \{(t,t),(c,t)\}, \{(t,t),(c,c)\}, \\ &\{(t,c),(c,t)\}, \{(t,c),(c,c)\}, \{(c,t),(c,c)\}, \{(t,t),(t,c),(c,t)\}, \\ &\{(t,t),(t,c),(c,c)\}, \{(t,t),(c,t),(c,c)\}, \{(t,c),(c,t),(c,c)\}, \Omega\} \, . \end{aligned}$$

L'evento $E = \{(t,t),(c,t)\}$ è rettangolare, dal momento che può essere espresso come $E = E_1 \times E_2$, dove $E_1 = \{t, c\}$ è un evento di Ω_1, mentre $E_2 = \{t\}$ è un evento di Ω_2. L'evento E è anche cilindrico, dal momento che $E_1 = \Omega_1$, e può essere rappresentato dunque come $E = \Omega_1 \times E_2$. Al contrario, l'evento $E = \{(t,t),(t,c),(c,t)\}$ non è rettangolare, in quanto non può essere espresso come prodotto cartesiano di eventi di Ω_1 e Ω_2. □

• **Esempio 1.5.2.** (Schema di Bernoulli) Si analizza la generalizzazione dell'Esempio 1.5.1. Si consideri un esperimento aleatorio con esito dicotomico, ovvero suscettibile di assumere due soli risultati ω_1 e ω_2 del tipo "successo" e "insuccesso". Uno schema che sta alla base di molti modelli probabilistici è appunto la ripetizione di questo esperimento aleatorio per n volte. Lo spazio fondamentale prodotto $\Omega = \Omega_1 \times \cdots \times \Omega_n$ relativo alla combinazione degli n esperimenti aleatori risulta allora composto da 2^n elementi del tipo $(\omega_{j_1}, \ldots, \omega_{j_n})$, dove $j_k = 1, 2$ e $k = 1, \ldots, n$. In effetti, $(\omega_{j_1}, \ldots, \omega_{j_n})$ è il risultato dell'esperimento combinato tale che "ω_{j_1} si è verificato nel primo esperimento, … , ω_{j_n} si è verificato nell'n-esimo esperimento". In pratica Ω è costituito da tutte le 2^n possibili n-ple composte dai risultati ω_1 e ω_2. Quindi, la σ-algebra prodotto è data da $\mathcal{F} = \mathcal{F}_1 \otimes \cdots \otimes \mathcal{F}_n = \mathcal{P}(\Omega)$ ed è pertanto costituita da 2^{2^n} eventi. Lo schema appena considerato può essere ulteriormente generalizzato considerando un'infinità numerabile di ripetizioni dell'esperimento aleatorio. In questo caso, lo spazio fondamentale prodotto $\Omega = \prod_{n \geq 1} \Omega_n$ relativo alla combinazione degli esperimenti aleatori risulta allora composto dalle successioni del tipo $(\omega_{j_n})_{n \geq 1}$, dove $j_n = 1, 2$. Inoltre, la σ-algebra prodotto $\mathcal{F} = \bigotimes_{n \geq 1} \mathcal{F}_n$ può essere costruita a partire dalla classe degli eventi cilindrici. Risulta interessante notare che se $\Omega_n = \{0, 1\}$, allora ogni elemento di Ω è in effetti una successione di simboli 0 e 1. Tenendo presente la rappresentazione binaria dei numeri in $[0, 1]$, ogni evento elementare di Ω può essere quindi messo in corrispondenza con un numero in $[0, 1]$. Lo schema considerato è quindi equivalente all'esperimento aleatorio che consiste nello scegliere in modo casuale un punto su un segmento di lunghezza unitaria (per un approfondimento di questi argomenti si consideri Billingsley, 1995, p.1). Si deve infine sottolineare che il cosiddetto "schema delle prove ripetute" ora descritto è anche detto schema di Bernoulli, dal momento che fu introdotto dal matematico svizzero Jakob Bernoulli (1654-1705), un membro della famosa famiglia di matematici e fisici che include fra l'altro il fratello Johann Bernoulli (1667-1748) e il nipote Daniel Bernoulli (1700-1782). Jakob Bernoulli è l'autore di *Ars Conjectandi* (pubblicato postumo nel 1713), l'opera che ha posto le basi del calcolo combinatorio e introdotto alcuni risultati fondamentali della Teoria della Probabilità. □


JACOBI BERNOULLI,
ARS CONJECTANDI,
TRACTATUS
DE SERIEBUS INFINITIS,
DE LUDO PILÆ
RETICULARIS.
BASILEÆ,

Figura 1.5.1. Jakob Bernoulli (1654-1705) e frontespizio di *Ars Conjectandi* (1713).

• **Esempio 1.5.3.** Si consideri il tempo di attesa per un evento sismico in due determinate regioni. Ogni risultato di questo fenomeno aleatorio è del tipo "i tempi di attesa sono rispettivamente pari a t_1 e t_2 unità temporali", dove $t_1, t_2 \geq 0$. Quindi, dal momento che lo spazio fondamentale relativo a ciascun fenomeno aleatorio è del tipo $\Omega_k = [0, \infty[$ con $k = 1, 2$, lo spazio fondamentale prodotto risulta

$$\Omega = \Omega_1 \times \Omega_2 = [0, \infty[\times [0, \infty[\, .$$

L'evento

$$E = \{(t_1, t_2) \in \Omega : t_1 \in [1, 3], t_2 \in [4, \infty[\}$$

è rettangolare dal momento che $E = E_1 \times E_2$, dove $E_1 \subset \Omega_1$ e $E_2 \subset \Omega_2$ sono dati da $E_1 = [1, 3]$ e $E_2 = [4, \infty[$. Al contrario l'evento

$$E = \{(t_1, t_2) \in \Omega : t_1 + t_2 \in [0, 3]\} \, ,$$

ovvero "la somma dei tempi di attesa non supera le 3 unità temporali", non è dato dal prodotto cartesiano di eventi di Ω_1 e Ω_2. Infine l'evento

$$E = \{(t_1, t_2) \in \Omega : t_1 \in [0, 3]\} \, ,$$

ovvero "il tempo di attesa nella prima regione non supera le 3 unità temporali", risulta esprimibile come prodotto cartesiano $E = E_1 \times \Omega_2$, dove $E_1 \subset \Omega_1$ è dato da $E_1 = [0, 3]$. Quindi, E rappresenta l'evento cilindrico che si verifichi E_1 nel primo esperimento aleatorio. Per quanto riguarda infine la σ-algebra prodotto, $\mathcal{F}$ è in questo caso la σ-algebra generata dagli eventi rettangolari del tipo $E_1 \times E_2$, dove $E_1 \in \mathcal{F}_1$ e $E_2 \in \mathcal{F}_2$, mentre $\mathcal{F}_1$ e $\mathcal{F}_2$ sono σ-algebre di eventi rispettivamente costruite come nella Sezione 1.4 nel caso di un solo esperimento aleatorio con spazio fondamentale di cardinalità non numerabile. Se si considera infine il tempo di attesa per un evento sismico per regioni situate lungo un intero meridiano, allora lo spazio fondamentale prodotto risulta $\Omega = \prod_{k \in I} \Omega_k$, dove I rappresenta l'insieme di posizioni sul meridiano. In questo caso, I è un insieme che contiene un'infinità non numerabile di indici e $\mathcal{F}$ può essere scelta come la σ-algebra generata dalla classe degli eventi cilindrici di Ω. □

1.6. Riferimenti bibliografici

Per quanto riguarda la storia della Teoria della Probabilità, si suggerisce di consultare i testi di Hald (1998, 2003). L'approccio alla probabilità basato sulla Teoria della Misura è estesamente considerato nei testi avanzati di Ash e Doléans-Dade (2000), Billingsley (1995), Borovkov (2013), Foata e Fuchs

(1998), Gut (2005), Rao e Swift (2006), Resnick (2014) e Walsh (2012) e nei testi introduttivi di Jacod e Protter (2004), Letta (1993), Rosenthal (2006) e Venkatesh (2012). Basandosi sul medesimo approccio, i testi di Dudley (2004) e Kallenberg (2002) forniscono risultati e dimostrazioni in modo dettagliato e sono opportuni per un lettore esigente. Per un approccio classico alla probabilità, si dovrebbe consultare il fondamentale testo di Feller (1968). Per quanto riguarda le tecniche di enumerazione e il calcolo combinatorio si veda Graham *et al.* (1994).

1.7. Esercizi svolti

Sezione 1.1

• **Esercizio 1.1.1.** (Modello di Maxwell-Boltzmann) Si consideri l'esperimento aleatorio che consiste nell'inserire in modo casuale m palline distinguibili in M celle. Si determini la cardinalità dello spazio fondamentale Ω relativo all'esperimento.

Soluzione. Al fine di enumerare le configurazioni di palline nelle celle, si consideri il principio di moltiplicazione. Questo principio afferma che in una sequenza di m scelte, per cui vi sono M_1 alternative al primo livello, ..., M_m alternative all'm-esimo livello, il numero di possibili modi in cui possono essere fatte le scelte è dato da $\prod_{k=1}^{m} M_k$. Nel caso specifico, si deve scegliere dove inserire la prima pallina nelle M celle, ..., l'm-esima pallina nelle M celle. Dunque, $\text{card}(\Omega) = M^m$. Nel modello di Maxwell-Boltzmann, inizialmente considerato nella fisica teorica, questa cardinalità è in effetti quella delle modalità in cui si possono disporre m particelle in M regioni spaziali. □

• **Esercizio 1.1.2.** Si consideri l'esperimento aleatorio che consiste nell'inserire in modo casuale m palline distinguibili in M celle in modo tale da avere al massimo una pallina per cella, assumendo che $m \leq M$. Si determini la cardinalità dello spazio fondamentale Ω relativo all'esperimento.

Soluzione. Per il principio di moltiplicazione considerato nell'Esercizio 1.1.1, si deve scegliere dove inserire la prima pallina nelle M celle, la seconda pallina nelle rimanenti $(M-1)$ celle disponibili, ..., l'm-esima pallina nelle ultime $(M-m+1)$ celle disponibili. Dunque, si ha

$$\text{card}(\Omega) = M(M-1) \times \cdots \times (M-m+1) = \frac{M!}{(M-m)!} .$$

Si osservi che se $m = M$, allora risulta $\text{card}(\Omega) = M!$. □

• **Esercizio 1.1.3.** (Modello di Bose-Einstein) Si consideri l'esperimento aleatorio che consiste nell'inserire in modo casuale m palline indistinguibili in M celle. Si determini la cardinalità dello spazio fondamentale Ω relativo all'esperimento.

Soluzione. Per ogni configurazione di palline nelle celle si può adottare una rappresentazione per mezzo di due simboli, ovvero le barre e gli asterischi. Più esattamente, si considerano M spazi fra $(M+1)$ barre per indicare le celle, mentre le palline indistinguibili vengono indicate con m asterischi disposti fra le barre. Si tenga presente che il simbolo iniziale e finale in ogni configurazione deve essere necessariamente la barra. Al fine di ottenere una configurazione, si deve dunque scegliere dove inserire la prima pallina fra le $(M-1)$ barre interne, ovvero si hanno M possibili scelte. Si ottiene in questo modo una sequenza di M simboli, ovvero $(M-1)$ barre e un asterisco. La seconda pallina viene inserita fra gli M simboli precedenti, ovvero si hanno $(M+1)$ possibili scelte. Si ottiene quindi una sequenza di $(M+1)$ simboli, ovvero $(M-1)$ barre e due asterischi. Iterando il procedimento, l'm-esima pallina viene inserita fra gli $(M+m-2)$ simboli precedenti, ovvero si hanno $(M+m-1)$ possibili scelte, ottenendo infine una sequenza di $(M+m-1)$ simboli, ovvero $(M-1)$ barre e m asterischi. Le palline sono indistinguibili e quindi l'ordine in cui vengono inserite è ininfluente. Dunque, dal momento che esistono $m!$ possibili ordini di inserimento e per il principio di moltiplicazione, si ha

$$\text{card}(\Omega) = \frac{M(M+1) \times \cdots \times (M+m-1)}{m!} = \binom{M+m-1}{m}.$$

Nel modello di Bose-Einstein considerato nella meccanica statistica, questa cardinalità è in effetti quella delle modalità in cui si possono disporre m particelle indistinguibili in M regioni spaziali. Il modello si applica ad esempio ai fotoni ed atomi che contengono un numero pari di particelle elementari.

□

• **Esercizio 1.1.4.** (Modello di Fermi-Dirac) Si consideri l'esperimento aleatorio che consiste nell'inserire in modo casuale m palline indistinguibili in M celle in modo da avere al massimo una pallina per cella, assumendo che $m \leq M$. Si determini la cardinalità dello spazio fondamentale Ω relativo all'esperimento.

Soluzione. Sulla base dell'Esercizio 1.1.2, dal momento che esistono $m!$ possibili ordini di inserimento di palline indistinguibili, si ha

$$\text{card}(\Omega) = \frac{M(M-1) \times \cdots \times (M-m+1)}{m!} = \binom{M}{m}.$$

Nel modello di Fermi-Dirac, questa cardinalità è in effetti quella delle modalità in cui si possono disporre m particelle indistinguibili in M regioni spaziali, in modo che ogni regione contenga al massimo una particella. Il modello si applica ad esempio ad elettroni, protoni e neutroni. □

Sezione 1.2

• **Esercizio 1.2.1.** Considerato lo spazio fondamentale Ω e la successione di eventi $(E_n)_{n\geq 1}$, si verifichi che $\liminf_n E_n \subseteq \limsup_n E_n$ e che

$$\limsup_n E_n = (\liminf_n E_n^c)^c.$$

Soluzione. Si assuma che $\omega \in \liminf_n E_n$. Tenendo presente la definizione di limite inferiore di una successione di eventi, per qualche n si ha $\omega \in \bigcap_{k=n}^{\infty} E_k$ e quindi risulta $\omega \in E_k$ per ogni $k \geq n$. Dunque, si ha che $\omega \in \bigcup_{k=n}^{\infty} E_k$ per ogni n, ovvero $\omega \in \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} E_k = \limsup_n E_n$. Quindi, $\liminf_n E_n \subseteq \limsup_n E_n$. Inoltre, tenendo presente la relazione di De Morgan, si ha

$$\limsup_n E_n = \bigcap_{n=1}^{\infty} \left(\bigcap_{k=n}^{\infty} E_k^c \right)^c = \left(\bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} E_k^c \right)^c = (\liminf_n E_n^c)^c. \qquad \square$$

• **Esercizio 1.2.2.** Si consideri lo spazio fondamentale $\Omega = [0,2]$ e la successione di eventi $(E_n)_{n\geq 1}$ tale che $E_n = [0, 1 + (-1)^{n+1} n^{-1}]$. Si determini $\liminf_n E_n$ e $\limsup_n E_n$.

Soluzione. Si ha

$$\bigcap_{k=n}^{\infty} E_k = \begin{cases} [0, 1-n^{-1}] & n \text{ pari} \\ [0, 1-(n+1)^{-1}] & n \text{ dispari} \end{cases}$$

e dunque risulta

$$\liminf_n E_n = \bigcup_{n=1}^{\infty} [0, 1-(n+1)^{-1}] = [0,1[\,.$$

Analogamente, si ha

$$\bigcup_{k=n}^{\infty} E_k = \begin{cases} [0, 1 + (n+1)^{-1}] & n \text{ pari} \\ [0, 1 + n^{-1}] & n \text{ dispari} \end{cases}$$

e dunque risulta

$$\limsup_n E_n = \bigcap_{n=1}^{\infty} [0, 1 + n^{-1}] = [0, 1] .$$

Non esiste il limite della successione in quanto $\liminf_n E_n$ e $\limsup_n E_n$ non coincidono. □

• **Esercizio 1.2.3.** Si consideri uno spazio fondamentale Ω e la successione di eventi $(E_n)_{n\geq 1}$ tale che $E_n = E$ se n è pari, mentre $E_n = F$ se n è dispari. Si determini $\liminf_n E_n$ e $\limsup_n E_n$.

Soluzione. Si ha

$$\bigcap_{k=n}^{\infty} E_k = \bigcap_{k=n}^{\infty} (E_k \cap E_{k+1}) = \bigcap_{k=n}^{\infty} (E \cap F) = E \cap F$$

e dunque risulta

$$\liminf_n E_n = \bigcup_{n=1}^{\infty} (E \cap F) = E \cap F .$$

Analogamente, si ha

$$\bigcup_{k=n}^{\infty} E_k = \bigcup_{k=n}^{\infty} (E_k \cup E_{k+1}) = \bigcup_{k=n}^{\infty} (E \cup F) = E \cup F$$

e dunque risulta

$$\limsup_n E_n = \bigcap_{n=1}^{\infty} (E \cup F) = E \cup F .$$

Non esiste il limite della successione in quanto $\liminf_n E_n$ e $\limsup_n E_n$ non coincidono. □

Sezione 1.3

• **Esercizio 1.3.1.** Si consideri lo spazio fondamentale relativo al lancio di un dado, ovvero $\Omega = \{1, 2, 3, 4, 5, 6\}$, e la classe di eventi $\mathcal{E} = \{E_1, E_2\}$, dove $E_1 = \{1, 2\}$ e $E_2 = \{2\}$. Si descriva la σ-algebra $\sigma(\mathcal{E})$ generata da $\mathcal{E}$.

Soluzione. La σ-algebra $\mathcal{F} = \sigma(\mathcal{E})$ deve contenere gli eventi $E_1 \cap E_2 = \{2\}$, $E_1 \cap E_2^c = \{1\}$ e $E_1^c \cap E_2^c = \{3, 4, 5, 6\}$ che costituiscono una partizione di Ω, oltre agli eventi $\emptyset$ e Ω. Dunque, la σ-algebra richiesta è data da

$$\mathcal{F} = \{\emptyset, \{1\}, \{2\}, \{1, 2\}, \{3, 4, 5, 6\}, \{1, 3, 4, 5, 6\}, \{2, 3, 4, 5, 6\}, \Omega\} . \qquad \square$$

• **Esercizio 1.3.2.** Si consideri spazio fondamentale Ω e la classe di eventi $\mathcal{E} = \{E_1, E_2\}$. Si descriva la σ-algebra $\sigma(\mathcal{E})$ generata da $\mathcal{E}$.

Soluzione. La σ-algebra $\mathcal{F} = \sigma(\mathcal{E})$ può essere costruita dalla partizione di Ω data dai 4 eventi $E_1 \cap E_2$, $E_1 \cap E_2^c$, $E_1^c \cap E_2$ e $E_1^c \cap E_2^c$. La σ-algebra $\mathcal{F}$ è tale che $\text{card}(\mathcal{F}) = 2^4$ e contiene tutti gli eventi che si possono ottenere mediante le opportune unioni di eventi della partizione. In generale, se si considera la classe di eventi $\mathcal{E} = \{E_1, \ldots, E_n\}$, la σ-algebra $\mathcal{F} = \sigma(\mathcal{E})$ può essere costruita dalla

partizione di Ω data dai 2^n eventi del tipo $\bigcap_{k=1}^n F_k$, dove $F_k = E_k$ se $k \in J$ e $F_k = E_k^c$ se $k \in I \setminus J$, mentre J è una scelta di indici di $I = \{1, \ldots, n\}$. In questo caso risulta $\text{card}(\mathcal{F}) = 2^{2^n}$. □

Sezione 1.4

• **Esercizio 1.4.1.** Si consideri lo spazio fondamentale $\Omega =]0, 1]$ e la classe di eventi $\mathcal{F}$ di Ω tale che se $E \in \mathcal{F}$, allora E è numerabile o E^c è numerabile. Si verifichi che $\mathcal{F}$ è una σ-algebra.

Soluzione. La classe $\mathcal{F}$ è non vuota dal momento che $E =]0, 1] \cap \mathbb{Q} \in \mathcal{F}$ è numerabile. Dalla definizione di $\mathcal{F}$ si verifica immediatamente che se $E \in \mathcal{F}$ allora $E^c \in \mathcal{F}$. Inoltre, sia $(E_n)_{n\geq 1}$ una successione di eventi di $\mathcal{F}$. Tenendo presente che un'unione numerabile di eventi numerabili è ancora numerabile, se tutti gli eventi della successione sono numerabili si ha $\bigcup_{n=1}^\infty E_n \in \mathcal{F}$. Al contrario, si supponga che almeno un evento E_k della successione sia tale che E_k^c è numerabile. Dunque, si ha $(\bigcup_{n=1}^\infty E_n)^c = \bigcap_{n=1}^\infty E_n^c \subset E_k^c \in \mathcal{F}$ e quindi anche in questo caso risulta $\bigcup_{n=1}^\infty E_n \in \mathcal{F}$. Si deve quindi concludere che $\mathcal{F}$ è una σ-algebra. □

Sezione 1.5

• **Esercizio 1.5.1.** Si consideri lo spazio probabilizzabile prodotto $(\Omega, \mathcal{F})$ con $\Omega = \Omega_1 \times \Omega_2$ e la classe di eventi $\mathcal{F}_1$ tale che se $E \in \Omega_1$, allora $E \times \Omega_2 \in \mathcal{F}$. Si verifichi che $\mathcal{F}_1$ è una σ-algebra sullo spazio fondamentale Ω_1.

Soluzione. Si osservi che $\mathcal{F}_1$ è una classe di eventi cilindrici. La classe è non vuota dal momento che $\mathcal{F}$ è una σ-algebra. Inoltre, se $E \in \mathcal{F}_1$ si ha

$$E^c \times \Omega_2 = (E \times \Omega_2)^c \in \mathcal{F}$$

e quindi $E^c \in \mathcal{F}_1$. Infine, se $(E_n)_{n\geq 1}$ è una successione di eventi di $\mathcal{F}_1$ si ha

$$\bigcup_{n=1}^\infty (E_n \times \Omega_2) = \left(\bigcup_{n=1}^\infty E_n \right) \times \Omega_2 \in \mathcal{F} \,,$$

per cui $\bigcup_{n=1}^\infty E_n \in \mathcal{F}_1$. Dunque, $\mathcal{F}_1$ è una σ-algebra. □

Pagina intenzionalmente vuota

Capitolo 2

Misure di probabilità

2.1. Definizione assiomatica di probabilità

Il trattamento rigoroso della Teoria della Probabilità è stato introdotto intorno al 1930 dal matematico russo Andrej Nikolaevic Kolmogorov (1903-1987). Kolmogorov sviluppò il concetto di probabilità assumendo la Teoria della Misura come metalinguaggio per la Teoria della Probabilità, in modo da superare il dibattito fra quanti consideravano la probabilità come limite di frequenze relative, ovvero la cosiddetta impostazione frequentista, e quanti cercavano un fondamento logico della stessa (per un'analisi approfondita dei vari concetti di probabilità, si veda von Plato, 1994). Al contrario, l'approccio proposto da Kolmogorov semplicemente definisce la probabilità in modo assiomatico, postulando in effetti che la probabilità sia una misura normalizzata (Kolmogorov, 1933). Risulta quindi opportuno definire inizialmente il concetto di misura prima di introdurre quello di probabilità.

Figura 2.1.1. Andrej Nikolaevic Kolmogorov (1903-1987).

Definizione 2.1.1. Una applicazione $\mu : \mathcal{F} \to [0, \infty[$ è una *misura* su $\mathcal{F}$ se per ogni successione di eventi incompatibili $(E_n)_{n\geq 1} \in \mathcal{F}$ si ha

$$\mu\left(\bigcup_{n=1}^{\infty} E_n\right) = \sum_{n=1}^{\infty} \mu(E_n) \, . \qquad \square$$

La precedente definizione stabilisce che una misura è una funzione non negativa su $\mathcal{F}$ con la cosiddetta proprietà della σ-*additività* (o *additività numerabile*). Una misura è detta invece *additiva* se dati n eventi incompatibili $(E_k)_{k=1}^{n} \in \mathcal{F}$, allora $\mu(\bigcup_{k=1}^{n} E_k) = \sum_{k=1}^{n} \mu(E_k)$. Inoltre, se $\mu(\Omega) = 1$ la misura è detta *normalizzata*. Nel linguaggio della Teoria della Misura, la terna $(\Omega, \mathcal{F}, \mu)$ è detta *spazio misurato*.

Data l'importanza del concetto di probabilità, in modo leggermente ridondante, viene data di seguito anche la classica definizione assiomatica di probabilità attraverso tre assiomi (che evidentemente definiscono di nuovo la probabilità come misura normalizzata).

Definizione 2.1.2. Una applicazione $P : \mathcal{F} \to [0, \infty[$ è una *misura di probabilità* (o semplicemente *probabilità*) se:
i) $P(E) \geq 0$ per ogni $E \in \mathcal{F}$;
ii) $P(\Omega) = 1$;
iii) per ogni successione di eventi incompatibili $(E_n)_{n\geq 1} \in \mathcal{F}$ si ha

$$P\left(\bigcup_{n=1}^{\infty} E_n\right) = \sum_{n=1}^{\infty} P(E_n) \, . \qquad \square$$

In effetti, gli assiomi *i*) e *iii*) della Definizione 2.1.2 stabiliscono che la probabilità è una misura σ-additiva, mentre l'assioma *ii*) stabilisce che questa misura è normalizzata a 1. La terna $(\Omega, \mathcal{F}, P)$ è detta *spazio probabilizzato*. Si noti che l'approccio assiomatico non fornisce il modo con cui selezionare la misura di probabilità P. In pratica, l'individuo che analizza l'esperimento o il fenomeno aleatorio esprime un "grado di fiducia" nei confronti degli eventi di $\mathcal{F}$, che lo portano a definire una data misura di probabilità. Ovviamente, un differente individuo potrebbe essere incline a specificare un diverso "grado di fiducia" nei confronti degli stessi eventi di $\mathcal{F}$, in modo tale da definire una differente misura di probabilità rispetto al primo individuo. Quindi, la scelta di P, ovvero la costruzione di uno spazio probabilizzato, ha per sua natura un carattere arbitrario e sarà discussa in dettaglio nella Sezione 2.3.

2.2. Alcune proprietà della probabilità

Come conseguenza della definizione assiomatica di probabilità derivano numerose proprietà. Alcune di queste proprietà sono enunciate di seguito.

Proposizione 2.2.1. *Se $(\Omega, \mathcal{F}, P)$ è uno spazio probabilizzato, si ha $P(\emptyset) = 0$.*
Dimostrazione. Dal momento che $\Omega = \Omega \cup \emptyset$, essendo Ω e $\emptyset$ incompatibili, per l'assioma *iii*) della Definizione 2.1.2 si ha $P(\Omega) = P(\Omega) + P(\emptyset)$ e quindi risulta $P(\emptyset) = 0$ tenendo presente l'assioma *ii*) della Definizione 2.1.2. $\square$

Proposizione 2.2.2. *Sia $(\Omega, \mathcal{F}, P)$ uno spazio probabilizzato. Se $E \in \mathcal{F}$, allora*

$$P(E^c) = 1 - P(E) \, .$$

Dimostrazione. Dal momento che $\Omega = E \cup E^c$, essendo E e E^c incompatibili, per l'assioma *iii*) della Definizione 2.1.2 si ottiene $P(\Omega) = P(E) + P(E^c)$, da cui si ha la tesi tenendo presente l'assioma *ii*) della Definizione 2.1.2. $\square$

• **Esempio 2.2.1.** (Problema dei compleanni) In modo simile all'Esempio 1.1.3, si considerino m palline distinguibili inserite in modo casuale in M celle. Si supponga che ogni configurazione di palline nelle celle sia ugualmente probabile, ovvero che per ogni evento elementare di Ω si abbia $P(\{\omega_k\}) = M^{-m}$. Si vuole determinare la probabilità di avere una configurazione con almeno due palline in una cella. In questo caso, se E rappresenta l'evento di ottenere una configurazione con al più una pallina per cella, dal momento che esistono $\frac{M!}{(M-m)!}$ di tali configurazioni se $m \leq M$, si ha

$$P(E) = \frac{M!}{(M-m)!}\,\frac{1}{M^m}\,,$$

mentre $P(E) = 0$ se $m > M$. Dunque, la probabilità richiesta è data da

$$P(E^c) = 1 - \frac{M!}{(M-m)!}\,\frac{1}{M^m}$$

se $m \leq M$, mentre $P(E^c) = 1$ se $m > M$. Risulta inoltre facile verificare che $P(E^c)$ è una funzione crescente di m per un dato M. Questi risultati possono essere utilizzati nella soluzione del cosiddetto problema dei compleanni. Supponendo che m persone siano riunite in una stanza, questo problema consiste nell'ottenere la probabilità di avere almeno due persone con il compleanno nello stesso giorno dell'anno. Se si assume che tutte le configurazioni di compleanni siano ugualmente probabili e di avere un anno pari a $M = 365$ giorni, la probabilità richiesta si ottiene immediatamente dalla precedente espressione. In modo piuttosto controintuitivo, è sufficiente che $m = 23$ persone siano presenti nella stanza per ottenere che questa probabilità sia maggiore di $\frac{1}{2}$. In effetti in questo caso si ha $P(E^c) \simeq 0.507$. Si noti che $P(E^c)$ cresce rapidamente al crescere di m. In effetti se $m = 50$ persone sono presenti nella stanza, si ha $P(E^c) \simeq 0.970$. Per curiosità relative al problema dei compleanni si veda Olofsson (2015, p.61). □

Proposizione 2.2.3. *Sia $(\Omega, \mathcal{F}, P)$ uno spazio probabilizzato. Se $E_1, E_2 \in \mathcal{F}$ e $E_1 \subset E_2$, allora*

$$P(E_2 \setminus E_1) = P(E_2) - P(E_1)\,.$$

Inoltre, si ha $P(E_1) \leq P(E_2)$.

Dimostrazione. Dal momento che $E_1 \subset E_2$, E_2 può essere scritto come

$$E_2 = E_1 \cup (E_2 \setminus E_1)\,,$$

dove E_1 e $E_2 \setminus E_1$ sono tra loro incompatibili. Dunque, per l'assioma *iii*) della Definizione 2.1.2 si ha

$$P(E_2) = P(E_1) + P(E_2 \setminus E_1)\,,$$

da cui segue immediatamente la prima parte. Essendo $P(E_2 \setminus E_1) \geq 0$ per l'assioma *i*) della Definizione 2.1.2, si ha la seconda parte. □

Come conseguenza della Proposizione 2.2.3, dal momento che per ogni $E \in \mathcal{F}$ si ha $E \subset \Omega$, allora $P(E) \leq P(\Omega) = 1$. Tenendo presente l'assioma *i*) della Definizione 2.1.2, si può quindi concludere che $0 \leq P(E) \leq 1$ per ogni $E \in \mathcal{F}$.

Proposizione 2.2.4. *Sia $(\Omega, \mathcal{F}, P)$ uno spazio probabilizzato. Se $E_1, E_2 \in \mathcal{F}$, allora*

$$P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1 \cap E_2)\,.$$

Dimostrazione. Tenendo presente le relazioni

$$E_1 \cup E_2 = E_1 \cup (E_2 \setminus E_1)$$

e

$$E_2 = (E_1 \cap E_2) \cup (E_2 \setminus E_1)\,,$$

dal momento che i secondi membri delle due uguaglianze sono dati dall'unione di eventi incompatibili, per l'assioma *iii*) della Definizione 2.1.2 risulta

$$P(E_1 \cup E_2) = P(E_1) + P(E_2 \setminus E_1)$$

e

$$P(E_2) = P(E_1 \cap E_2) + P(E_2 \setminus E_1) \, .$$

Sottraendo membro a membro queste relazioni si ottiene la tesi. □

• **Esempio 2.2.2.** (Gioco del bridge) Il gioco del bridge è stato frequentemente analizzato in termini teorici dai matematici. Émile Borel è stato addirittura l'autore, insieme al giocatore di scacchi e di bridge André Chéron (1895-1980), di un manuale specifico, ovvero *Théorie Mathématique du Bridge à la Portée de Tous* (1940). Nel gioco del bridge viene assegnata ad ogni giocatore una mano composta da 13 carte scelte casualmente da un mazzo di 52 carte. Quindi, se si considera la mano di carte assegnata ad un singolo giocatore, i possibili risultati sono dati da tutte le scelte di 13 carte dal mazzo, ovvero $\text{card}(\Omega) = \binom{52}{13}$. Se si assume che nessun baro sia presente al tavolo di gioco, si può supporre che ogni mano di carte sia ugualmente probabile, ovvero che per ogni evento elementare si abbia $P(\{\omega_k\}) = \binom{52}{13}^{-1}$. Si consideri dunque la probabilità di ottenere una mano che contiene l'asso, il re, la regina, il fante e il dieci di almeno un seme di cuori o di quadri. Se si indica con E_1 e E_2 gli eventi tali che le cinque carte descritte siano presenti nella mano rispettivamente per il seme di cuori e di quadri, allora si ha

$$P(E_1) = P(E_2) = \binom{47}{8}\binom{52}{13}^{-1} \, .$$

In effetti vi sono $\binom{47}{8}$ mani equiprobabili che contengono le cinque carte d'interesse per il seme di cuori e altrettante mani per il seme di quadri. Inoltre, si ha

$$P(E_1 \cap E_2) = \binom{42}{3}\binom{52}{13}^{-1} \, ,$$

dal momento che vi sono $\binom{42}{3}$ mani equiprobabili che contengono contemporaneamente le cinque carte d'interesse del seme di cuori e del seme di quadri. Dunque la probabilità richiesta è data da

$$P(E_1 \cup E_2) = 2\binom{47}{8}\binom{52}{13}^{-1} - \binom{42}{3}\binom{52}{13}^{-1} \simeq 0.001 \, .$$

Questo esempio richiede la conoscenza elementare dei metodi di enumerazione (per una introduzione all'argomento si veda Graham *et al.*, 1994). Ulteriori problemi relativi al gioco del bridge sono considerati da Ash (2008) e da Feller (1968). □

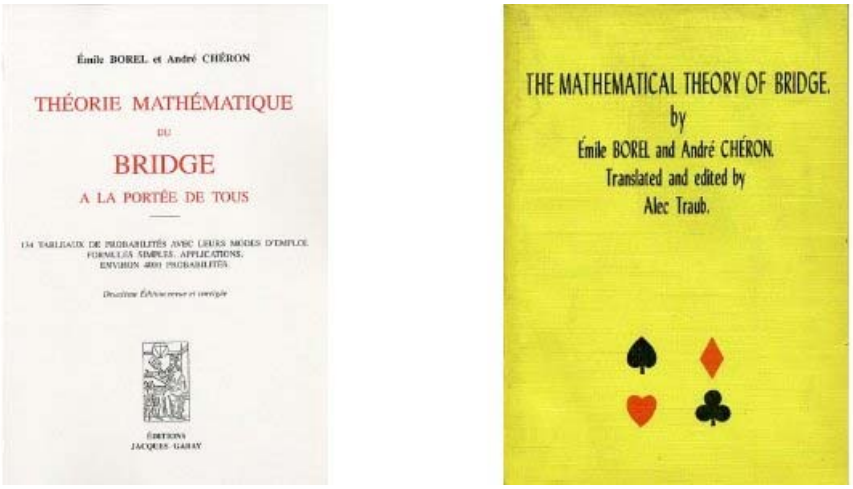


Figura 2.2.1. Frontespizio dell'edizione originale e dell'edizione inglese di *Théorie Mathématique du Bridge à la Portée de Tous* (1940).

La Proposizione 2.2.4 può essere utilizzato in maniera ricorsiva per la determinazione della probabilità dell'unione di n eventi. Ad esempio, dati tre eventi E_1, E_2 e E_3, tenendo presente le proprietà associativa e distributiva dell'unione di eventi e applicando tre volte la Proposizione 2.2.4, si ottiene

$$\begin{aligned} P(E_1 \cup E_2 \cup E_3) &= P(E_1 \cup E_2) + P(E_3) - P((E_1 \cup E_2) \cap E_3) \\ &= P(E_1) + P(E_2) - P(E_1 \cap E_2) + P(E_3) - P((E_1 \cap E_3) \cup (E_2 \cap E_3)) \\ &= P(E_1) + P(E_2) + P(E_3) - P(E_1 \cap E_2) - P(E_1 \cap E_3) - P(E_2 \cap E_3) + P(E_1 \cap E_2 \cap E_3) \,. \end{aligned}$$

In generale, si può ottenere la probabilità dell'unione di n eventi mediante la *Formula di Inclusione ed Esclusione* considerata nel seguente Teorema. La formula è stata introdotta da Abraham de Moivre (1667-1754), uno dei padri della Teoria della Probabilità, anche se è comunemente legata al nome del matematico francese Jules Henri Poincaré (1854-1912). Abraham de Moivre nacque in Francia anche se passò la maggior parte della sua vita in esilio in Inghilterra, dove scrisse quello che può essere considerato il primo manuale di Teoria della Probabilità, ovvero *The Doctrine of Chances* pubblicato nel 1718.

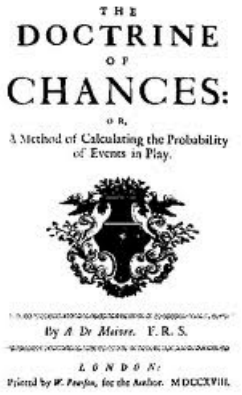

THE
DOCTRINE
OF
CHANCES:
OR,
A Method of Calculating the Probability
of Events in Play.
By A. De Moivre. F. R. S.
LONDON:

Figura 2.2.2. Abraham de Moivre (1667-1754) e frontespizio di *The Doctrine of Chances* (1718).

Teorema 2.2.5. (Formula di Inclusione ed Esclusione) *Sia* $(\Omega, \mathcal{F}, P)$ *uno spazio probabilizzato. Dati* n *eventi* $(E_k)_{k=1}^n \in \mathcal{F}$*, si ha*

$$P\left(\bigcup_{k=1}^{n} E_k\right) = \sum_{k=1}^{n} (-1)^{k+1} S_{k,n} \,,$$

dove

$$S_{k,n} = \sum_{1 \le j_1 < \ldots < j_k \le n} P(E_{j_1} \cap \ldots \cap E_{j_k}) \,.$$

Dimostrazione. La relazione può essere verificata in modo laborioso per induzione. Per una semplice dimostrazione basata sul valore atteso, si veda l'Esercizio 4.2.5. □

• **Esempio 2.2.3.** (Problema delle concordanze) Si considerino n celle numerate in cui vengono inserite in modo casuale n palline a loro volta numerate, in modo tale che una cella contenga una sola pallina. I possibili risultati sono dati da tutte le configurazioni di palline nelle celle e quindi risulta $\text{card}(\Omega) = n!$, ovvero gli elementi di Ω sono in corrispondenza biunivoca con le possibili permutazioni dei primi n interi. Supponendo che ogni configurazione sia ugualmente probabile, ovvero che per ogni evento elementare $\omega_k \in \Omega$ si abbia $P(\{\omega_k\}) = (n!)^{-1}$, si vuole determinare la probabilità che almeno una pallina sia in una cella con un numero identico. Se E_k rappresenta l'evento

per cui la configurazione è tale che la pallina numerata con k è nella cella con il medesimo numero, allora la probabilità in questione risulta $P(\bigcup_{k=1}^{n} E_k)$. Dal momento che esistono $(n-1)!$ configurazioni per cui la pallina numerata con j è nella cella numerata con j, allora si ha

$$P(E_j) = \frac{(n-1)!}{n!} .$$

Inoltre, esistono $(n-2)!$ configurazioni per cui la pallina numerata con j_1 è nella cella numerata con j_1, mentre la pallina numerata con j_2 è nella cella numerata con j_2. Dunque, si ha

$$P(E_{j_1} \cap E_{j_2}) = \frac{(n-2)!}{n!} .$$

In generale, per $k = 1, \ldots, n$ risulta

$$P(E_{j_1} \cap \ldots \cap E_{j_k}) = \frac{(n-k)!}{n!} .$$

Dal momento che $P(E_{j_1} \cap \ldots \cap E_{j_k})$ è costante per ogni k e che esistono $\binom{n}{k}$ possibili scelte di k eventi fra n eventi, si ha infine

$$S_{k,n} = \binom{n}{k} P(E_{j_1} \cap \ldots \cap E_{j_k}) = \frac{1}{k!} .$$

In base alla Formula di Inclusione ed Esclusione, la probabilità richiesta è quindi data da

$$P\left(\bigcup_{k=1}^{n} E_k\right) = \sum_{k=1}^{n} \frac{(-1)^{k+1}}{k!} .$$

Per n abbastanza elevato la precedente probabilità può essere approssimata da $1 - e^{-1} \simeq 0.632$ e quindi, in modo piuttosto controintuitivo, la probabilità che almeno una pallina sia inserita in una cella con lo stesso numero risulta maggiore di $\frac{1}{2}$. Per ulteriori problemi relativi a giochi di *rencontre* si veda Feller (1968). La prima analisi del presente problema è dovuta al matematico francese Pierre Raymond de Montmort (1678-1719), autore di *Essay d'analyse sur les jeux de hazard* (1708), una delle prime trattazioni del gioco d'azzardo con metodi probabilistici. □


ESSAY
D'ANALYSE
SUR
LES JEUX DE HAZARD.
SECONDE EDITION
A PARIS,
MDCCXIII

Figura 2.2.3. Frontespizio di
Essay d'analyse sur les jeux de hazard (1708).

Proposizione 2.2.6. *Sia $(\Omega, \mathcal{F}, P)$ uno spazio probabilizzato. Dati n eventi $(E_k)_{k=1}^{n} \in \mathcal{F}$, allora*

$$P\left(\bigcup_{k=1}^{n} E_k\right) \geq \max_k(P(E_k)) .$$

Dimostrazione. Dal momento che $E_k \subset \bigcup_{k=1}^{n} E_k$ per ogni k, allora $P(E_k) \leq P(\bigcup_{k=1}^{n} E_k)$ per la Proposizione 2.2.3. Se la disuguaglianza è valida per ogni $P(E_k)$, allora risulta valida anche per il corrispondente massimo. □

Proposizione 2.2.7. *Sia $(\Omega, \mathcal{F}, P)$ uno spazio probabilizzato. Dati n eventi $(E_k)_{k=1}^{n} \in \mathcal{F}$, allora*

$$P\left(\bigcap_{k=1}^{n} E_k\right) \leq \min_k(P(E_k)) .$$

Dimostrazione. Dal momento che $\bigcap_{k=1}^{n} E_k \subset E_k$ per ogni k, allora $P(\bigcap_{k=1}^{n} E_k) \leq P(E_k)$ per la Proposizione 2.2.3. Se la disuguaglianza è valida per ogni $P(E_k)$, allora risulta valida anche per il corrispondente minimo. □

Il prossimo Teorema fornisce due utili disuguaglianze sulla probabilità dell'unione di n eventi. Queste disuguaglianze hanno importanti applicazioni nella statistica matematica e furono introdotte dal matematico italiano Carlo Emilio Bonferroni (1892-1960).

Teorema 2.2.8. (Disuguaglianze di Bonferroni) *Sia $(\Omega, \mathcal{F}, P)$ uno spazio probabilizzato. Dati n eventi $(E_k)_{k=1}^{n} \in \mathcal{F}$, allora*

$$\sum_{k=1}^{n} P(E_k) - \sum_{1 \leq k < j \leq n} P(E_k \cap E_j) \leq P\left(\bigcup_{k=1}^{n} E_k\right) \leq \sum_{k=1}^{n} P(E_k) .$$

Dimostrazione. La maggiorazione è ovviamente verificata per $n = 1$ ed è valida per $n = 2$ per la Proposizione 2.2.4. Inoltre, dalla medesima Proposizione si ha

$$P\left(\bigcup_{k=1}^{n} E_k\right) = P\left(E_1 \cup \bigcup_{k=2}^{n} E_k\right) \leq P(E_1) + P\left(\bigcup_{k=2}^{n} E_k\right) .$$

Applicando di nuovo la Proposizione 2.2.4 risulta

$$P\left(\bigcup_{k=1}^{n} E_k\right) \leq P(E_1) + P\left(E_2 \cup \bigcup_{k=3}^{n} E_k\right) \leq P(E_1) + P(E_2) + P\left(\bigcup_{k=3}^{n} E_k\right) .$$

Iterando il procedimento si ottiene la maggiorazione per induzione. La seconda disuguaglianza è ovviamente verificata per $n = 1$ ed è valida per $n = 2$ per la Proposizione 2.2.4. Di nuovo per la medesima Proposizione e tenendo presente la precedente maggiorazione, si ha

$$\begin{aligned} P\left(\bigcup_{k=1}^{n} E_k\right) &= P(E_n) + P\left(\bigcup_{k=1}^{n-1} E_k\right) - P\left(\bigcup_{k=1}^{n-1} (E_n \cap E_k)\right) \\ &\geq P(E_n) + P\left(\bigcup_{k=1}^{n-1} E_k\right) - \sum_{k=1}^{n-1} P(E_n \cap E_k) . \end{aligned}$$

Applicando ancora la Proposizione 2.2.4, si ha

$$P\left(\bigcup_{k=1}^{n} E_k\right) \geq P(E_n) + P(E_{n-1}) + P\left(\bigcup_{k=1}^{n-2} E_k\right) - \sum_{k=1}^{n-1} P(E_n \cap E_k) - \sum_{k=1}^{n-2} P(E_{n-1} \cap E_k)\,.$$

Dunque, iterando il procedimento, si ottiene la seconda disuguaglianza per induzione. □

Proposizione 2.2.9. *Sia $(\Omega, \mathcal{F}, P)$ uno spazio probabilizzato. Dati n eventi $(E_k)_{k=1}^n \in \mathcal{F}$, si ha*

$$P\left(\bigcap_{k=1}^{n} E_k\right) \geq \sum_{k=1}^{n} P(E_k) - (n-1).$$

Dimostrazione. Dalla relazione di De Morgan, si ha

$$P\left(\bigcap_{k=1}^{n} E_k\right) = 1 - P\left(\bigcup_{k=1}^{n} E_k^c\right)\,.$$

Dalla precedente relazione e dal Teorema 2.2.8, si ottiene inoltre

$$P\left(\bigcap_{k=1}^{n} E_k\right) \geq 1 - \sum_{k=1}^{n} P(E_k^c) = 1 - \sum_{k=1}^{n} (1 - P(E_k)) = \sum_{k=1}^{n} P(E_k) - (n-1)$$

che è quanto si voleva dimostrare. □

Nel caso che vengano considerate successioni di eventi, si possono ottenere un insieme di ulteriori interessanti proprietà per la probabilità. I seguenti due Teoremi provano la "continuità" della misura di probabilità.

Teorema 2.2.10. (Teorema della Convergenza Monotona) *Se $(\Omega, \mathcal{F}, P)$ è uno spazio probabilizzato e se $(E_n)_{n\geq 1}$ è una successione crescente o decrescente di eventi di $\mathcal{F}$, allora*

$$\lim_n P(E_n) = P(\lim_n E_n)\,.$$

Dimostrazione. Si consideri una successione crescente. Si ha $E_n \subset E_{n+1}$ e quindi $P(E_n) \leq P(E_{n+1})$, ovvero la successione numerica $(P(E_n))_{n\geq 1}$ ammette limite essendo crescente e limitata superiormente da 1. Si esprima E_n come unione di eventi incompatibili, ovvero

$$E_n = E_1 \cup (E_2 \setminus E_1) \cup \ldots \cup (E_n \setminus E_{n-1}) = E_1 \cup \bigcup_{k=2}^{n} (E_k \setminus E_{k-1})\,,$$

da cui

$$\lim_n E_n = \bigcup_{n=1}^{\infty} E_n = E_1 \cup \bigcup_{k=2}^{\infty} (E_k \setminus E_{k-1})\,,$$

essendo $\lim_n E_n = \bigcup_{n=1}^{\infty} E_n$ dal momento che la successione è crescente. Inoltre, si ha

$$P(E_n) = P(E_1) + \sum_{k=2}^{n} P(E_k \setminus E_{k-1})\,,$$

e quindi, data la σ-additività di P, risulta

$$\lim_n P(E_n) = P(E_1) + \sum_{k=2}^{\infty} P(E_k \setminus E_{k-1}) = P\left(E_1 \cup \bigcup_{k=2}^{\infty} (E_k \setminus E_{k-1}) \right) = P(\lim_n E_n) \, .$$

Si consideri una successione decrescente. Risulta dunque $E_{n+1} \subset E_n$ e quindi $P(E_{n+1}) \leq P(E_n)$, ovvero la successione numerica $(P(E_n))_{n\geq 1}$ ammette limite essendo decrescente e limitata inferiormente da 0. Se $(E_n)_{n\geq 1}$ è decrescente la successione $(E_n^c)_{n\geq 1}$ è crescente, per cui dalla prima parte si ha

$$\lim_n P(E_n^c) = P(\lim_n E_n^c)$$

e quindi dalla relazione di De Morgan

$$\lim_n P(E_n) = \lim_n (1 - P(E_n^c)) = 1 - P(\lim_n E_n^c) = 1 - P\left(\bigcup_{n=1}^{\infty} E_n^c \right) = P\left(\bigcap_{n=1}^{\infty} E_n \right) = P(\lim_n E_n) \, ,$$

essendo $\lim_n E_n = \bigcap_{n=1}^{\infty} E_n$ dal momento che la successione è decrescente. □

Teorema 2.2.11. (Lemma di Fatou) *Se $(\Omega, \mathcal{F}, P)$ è uno spazio probabilizzato e se $(E_n)_{n\geq 1}$ è una successione di eventi di $\mathcal{F}$, allora*

$$P(\liminf_n E_n) \leq \liminf_n P(E_n) \leq \limsup_n P(E_n) \leq P(\limsup_n E_n) \, .$$

Se la successione ammette limite, allora

$$\lim_n P(E_n) = P(\lim_n E_n) \, .$$

Dimostrazione. Posto $F_n = \bigcap_{k=n}^{\infty} E_k$ e $D_n = \bigcup_{k=n}^{\infty} E_k$, $(F_n)_{n\geq 1}$ è una successione crescente di eventi e $(D_n)_{n\geq 1}$ è una successione decrescente di eventi. Inoltre, si ha $F_n \subset E_n \subset D_n$. Dunque, risulta

$$\lim_n F_n = \bigcup_{n=1}^{\infty} F_n = \liminf_n E_n$$

e

$$\lim_n D_n = \bigcap_{n=1}^{\infty} D_n = \limsup_n E_n \, ,$$

mentre dal Teorema 2.2.10 si ha

$$\lim_n P(F_n) = P(\lim_n F_n) = P(\liminf_n E_n)$$

e

$$\lim_n P(D_n) = P(\lim_n D_n) = P(\limsup_n E_n) \, .$$

Dal momento che $P(F_n) \leq P(E_n) \leq P(D_n)$ per ogni n, allora si ha la prima parte del Teorema. La seconda parte del Teorema segue immediatamente dalla prima, in quanto se la successione di eventi ammette limite, allora

$$P(\liminf_n E_n) = P(\limsup_n E_n) \, .$$ □

Proposizione 2.2.12. *Se $(\Omega, \mathcal{F}, P)$ è uno spazio probabilizzato e se $(E_n)_{n\geq 1}$ è una successione di eventi di $\mathcal{F}$ tale che $P(E_n) = 0$ per ogni n, allora*

$$P\left(\bigcup_{n=1}^{\infty} E_n\right) = 0 \, .$$

Inoltre, se $(E_n)_{n\geq 1}$ è una successione di eventi di $\mathcal{F}$ tale che $P(E_n) = 1$ per ogni n, allora

$$P\left(\bigcap_{n=1}^{\infty} E_n\right) = 1 \, .$$

Dimostrazione. Dalla prima disuguaglianza di Bonferroni (Teorema 2.2.8) per ogni n si ha

$$P\left(\bigcup_{k=1}^{n} E_k\right) \leq \sum_{k=1}^{n} P(E_k) = 0 \, ,$$

e quindi dal Teorema 2.2.10 e dall'assioma *i*) della Definizione 2.1.2 segue la prima parte della Proposizione. Dal momento che se $P(E_n) = 1$ allora $P(E_n^c) = 0$, dalla relazione di De Morgan e dalla prima parte della Proposizione si ha

$$P\left(\bigcap_{n=1}^{\infty} E_n\right) = 1 - P\left(\bigcup_{n=1}^{\infty} E_n^c\right) = 1 \, . \qquad \square$$

2.3. Costruzione di spazi probabilizzati

Dato uno spazio probabilizzabile $(\Omega, \mathcal{F})$, al fine di ottenere il relativo spazio probabilizzato $(\Omega, \mathcal{F}, P)$, occorre assegnare a ogni evento $E \in \mathcal{F}$ un numero reale $P(E)$ mediante una applicazione P in modo coerente con i tre assiomi della Definizione 2.1.2. A questo fine, è conveniente operare in modo che, assegnate le probabilità a certi particolari eventi di $\mathcal{F}$, risultino conseguentemente determinate le probabilità di tutti gli altri eventi di $\mathcal{F}$.

Se $\text{card}(\Omega) = n$, allora $\Omega = \{\omega_1, \ldots, \omega_n\}$ ed è sufficiente assegnare le probabilità a ciascuno degli eventi elementari, ovvero

$$P(\{\omega_k\}) = p_k \, , \, k = 1, \ldots, n \, ,$$

affinchè risultino determinate le probabilità di tutti i 2^n eventi di $\mathcal{F}$. Ovviamente, le probabilità di ogni evento elementare devono essere assegnate in modo da soddisfare i tre assiomi. In altri termini, occorre che $p_k \geq 0$ per ogni $k = 1, \ldots, n$ e che $\sum_{k=1}^{n} p_k = 1$. In questo caso, infatti, poichè ogni evento $E \in \mathcal{F}$ è costituito dall'unione di una scelta di eventi elementari ed essendo gli eventi elementari tra loro incompatibili, si ottiene che

$$P(E) = \sum_{k=1}^{n} \mathbf{1}_E(\omega_k) p_k \, ,$$

dove $\mathbf{1}_E$ rappresenta la funzione indicatrice dell'evento E, ovvero $\mathbf{1}_E(\omega) = 1$ se $\omega \in E$ e $\mathbf{1}_E(\omega) = 0$ altrimenti.

Nel caso in cui gli eventi elementari siano considerati equiprobabili, ovvero se si assegna $p_k = p$ per ogni $k = 1, \ldots, n$, allora occorre che risulti $p = n^{-1}$ affinchè $np = 1$. Quindi, risulta evidente che

$$P(E) = \frac{1}{n}\sum_{k=1}^{n} \mathbf{1}_E(\omega_k) = \frac{\text{card}(E)}{n} \, .$$

Questo tipo di assegnazione di probabilità è detta di Laplace, in quanto fu teorizzata in particolar modo dal matematico Pierre-Simon, marchese de Laplace (1749-1827), autore di alcune delle prime opere sistematiche di Teoria della Probabilità, ovvero *Théorie analytique des probabilités* (1812) e *Essai philosophique sur les probabilités* (1814).

THÉORIE

ANALYTIQUE

DES PROBABILITES;

PAR M. LE COMTE LAPLACE,

PARIS,

Figura 2.3.1. Pierre-Simon, marchese de Laplace (1749-1827) e frontespizio di *Théorie analytique des probabilités* (1812).

• **Esempio 2.3.1.** Si consideri il lancio di un dado. Se il dado non è truccato risulta naturale assegnare a ciascuno dei sei eventi elementari (corrispondenti a ciascuna faccia del dado) la stessa probabilità $p = \frac{1}{6}$. In questo caso, l'evento $E = \{1, 3, 5\}$ ha probabilità $P(E) = \frac{1}{2}$, essendo composto dall'unione di 3 eventi elementari. Se invece il dado è stato truccato e si assegnano le probabilità $P(\{\omega_k\}) = p_k$ tali che $p_1 = p_5 = \frac{1}{3}$ e $p_2 = p_3 = p_4 = p_6 = \frac{1}{12}$, si ottiene

$$P(E) = p_1 + p_3 + p_5 = \frac{3}{4} \, . \qquad \square$$

In modo analogo si procede nel caso in cui Ω è numerabile, ovvero se $\Omega = \{\omega_1, \omega_2, \dots\}$. Di nuovo, se si assegnano le probabilità a ciascuno degli eventi elementari, ovvero

$$P(\{\omega_n\}) = p_n \, , n = 1, 2, \dots \, ,$$

rispettando i vincoli $p_n \geq 0$ per ogni $n = 1, 2, \dots$ e $\sum_{n=1}^{\infty} p_n = 1$, risultano probabilizzati tutti gli eventi $E \in \mathcal{F}$, in quanto ciascuno di essi è costituito dall'unione di una scelta di un numero finito o di una infinità numerabile di eventi elementari. In altri termini, si ha

$$P(E) = \sum_{n=1}^{\infty} \mathbf{1}_E(\omega_n) p_n \, .$$

Si noti che, in caso di spazi fondamentali che contengono un'infinità numerabile di risultati, non è possibile assegnare a ciascun evento elementare la stessa probabilità $p > 0$. In questo modo, infatti, non converge la serie $\sum_{n=1}^{\infty} p$. Viceversa, assegnando $p = 0$ ad ogni evento elementare, la serie in questione converge ma risulta $P(\Omega) = 0$. In effetti, in questo caso l'equiprobabilità degli eventi elementari risulta possibile solo quando la misura di probabilità non è σ-additiva, ma semplicemente additiva, come risulta dal seguente esempio.

• **Esempio 2.3.2.** Si consideri uno spazio fondamentale che contiene una infinità numerabile di risultati, ovvero $\Omega = \{\omega_1, \omega_2, \dots\}$ e si supponga di considerare per ogni $E \in \mathcal{F}$, l'applicazione $\mu : \mathcal{F} \to [0, \infty[$ tale che

$$\mu(E) = \lim_k \frac{\text{card}(E \cap F_k)}{k} ,$$

dove $F_k = \{\omega_1, \dots, \omega_k\}$. L'applicazione μ rispetta l'assioma *i*) della Definizione 2.1.2 in quanto, trattandosi del limite di un rapporto tra quantità non negative, risulta $\mu(E) \geq 0$ per ogni $E \in \mathcal{F}$. Inoltre, l'applicazione μ rispetta anche l'assioma *ii*) della Definizione 2.1.2, in quanto $\text{card}(\Omega \cap F_k) = k$ e quindi

$$\mu(\Omega) = \lim_k \frac{\text{card}(\Omega \cap F_k)}{k} = 1 .$$

Tuttavia, l'applicazione μ non rispetta l'assioma *iii*) della Definizione 2.1.2. Infatti, per un evento elementare $\{\omega_n\}$ si ha $\text{card}(\{\omega_n\} \cap F_k) = 0$ se $n = k+1, k+2, \dots,$ e $\text{card}(\{\omega_n\} \cap F_k) = 1$ se $n = 1, \dots, k$, per cui

$$\mu(\{\omega_n\}) = \lim_k \frac{\text{card}(\{\omega_n\} \cap F_k)}{k} = 0 ,$$

ovvero μ assegna la stessa probabilità $p = 0$ ad ogni evento elementare e dal momento che

$$\sum_{n=1}^{\infty} \mu(\{\omega_n\}) = 0 < \mu(\Omega) = 1 ,$$

l'applicazione μ non è σ-additiva. L'applicazione μ risulta invece finitamente additiva, ovvero dati n eventi incompatibili $(E_k)_{k=1}^n \in \mathcal{F}$, si ha

$$\mu\left(\bigcup_{j=1}^n E_j\right) = \lim_k \frac{\text{card}(\bigcup_{j=1}^n (E_j \cap F_k))}{k} = \sum_{j=1}^n \lim_k \frac{\text{card}(E_j \cap F_k)}{k} = \sum_{j=1}^n \mu(E_j) ,$$

dal momento che

$$\text{card}\left(\bigcup_{j=1}^n (E_j \cap F_k)\right) = \sum_{j=1}^n \text{card}(E_j \cap F_k) .$$

Si osservi che se $E = \{\omega_2, \omega_4, \dots\}$ si ha $\mu(E) = \frac{1}{2}$, mentre se $E = \{\omega_3, \omega_6, \dots\}$ allora risulta $\mu(E) = \frac{1}{3}$. □

Infine, se Ω è non numerabile, in genere si sceglie un'opportuna classe iniziale di eventi $\mathcal{E}$ e si assegna la probabilità $P(E)$ a ciascun evento $E \in \mathcal{E}$ in modo da soddisfare i tre assiomi della Definizione 2.1.2. In questo caso, se la classe iniziale è scelta appropriatamente, si può infatti dimostrare che esiste una sola estensione di P da $\mathcal{E}$ alla σ-algebra $\sigma(\mathcal{E})$. In altri termini, esiste un solo modo di assegnare le probabilità agli eventi di $\sigma(\mathcal{E})$ in modo da rispettare gli assiomi e senza modificare le probabilità assegnate agli eventi di $\mathcal{E}$. Questo implica in pratica che, una volta probabilizzati gli eventi di $\mathcal{E}$, risultano probabilizzati univocamente anche tutti gli eventi di $\mathcal{F} = \sigma(\mathcal{E})$ (si veda Dudley, 2004, p.91). Si osservi che esistono eventi in Ω a cui non è possibile assegnare una probabilità, anche se la dimostrazione dell'esistenza di tali eventi richiede addirittura l'Assioma della Scelta (si veda Dudley, 2004, p.105).

• **Esempio 2.3.3.** Si consideri lo spazio fondamentale $\Omega =]0,1]$. In modo simile all'Esempio 1.4.3, la σ-algebra $\mathcal{F}$ può essere costruita a partire dalla classe $\mathcal{E}$ degli eventi del tipo $]a,b]$, dove $0 \leq a \leq b \leq 1$. In questo caso, si può verificare che

$$P(]a,b]) = b - a$$

è una misura di probabilità su $\mathcal{E}$ che ha un'unica estensione a $\mathcal{F} = \sigma(\mathcal{E})$ (si veda Dudley, 2004, p.94). Dunque, tenendo presente l'Esempio 1.4.3, in base al Teorema 2.2.10 si ha

$$P([a,b]) = P(\lim_n]a - n^{-1}, b]) = \lim_n P(]a - n^{-1}, b]) = \lim_n (b - a + n^{-1}) = b - a = P(]a,b]) \,.$$

In modo analogo, si ha

$$P(]a,b]) = P([a,b]) = P([a,b[) = P(]a,b[) \,.$$

Inoltre, si ha $P(\{a\}) = 0$, ovvero tutti gli eventi elementari hanno probabilità nulla. Di conseguenza, anche un'unione numerabile di eventi elementari ha probabilità nulla. In generale, mediante le proprietà viste nella Sezione 2.2, si possono ottenere le probabilità di tutti gli eventi ottenibili da un insieme numerabile di operazioni insiemistiche sugli eventi del tipo $]a,b[$, $]a,b]$, $[a,b[$ e $[a,b]$, ovvero si possono ottenere le probabilità di tutti gli eventi di $\mathcal{F}$. □

Nella Sezione 1.4 è stato visto che la σ-algebra di Borel su $\mathbb{R}$ può essere costruita a partire dalle classi di intervalli del tipo $]a,b]$. Nella Teoria della Misura, la misura λ che assegna la lunghezza ad ogni intervallo $]a,b]$ di $\mathbb{R}$, ovvero $\lambda(]a,b]) = b - a$, è detta *misura di Lebesgue*. Le basi della Teoria della Misura, che hanno permesso l'approccio assiomatico alla Teoria della Probabilità da parte di Kolmogorov, sono state appunto introdotte dal matematico francese Henri Léon Lebesgue (1875-1941) nella sua tesi di laurea *Intégrale, Longueur, Aire* (1902) sviluppata sotto la supervisione di Émile Borel.

Figura 2.3.2. Henri Léon Lebesgue (1875-1941).

Infine, una volta costruito lo spazio probabilizzato $(\Omega, \mathcal{F}, P)$ si dice che un evento E si verifica *quasi certamente* $(q.c.)$ rispetto a P se $P(E) = 1$. Analogamente, nel linguaggio della Teoria della Misura, si dice che una proprietà è valida *quasi ovunque* $(q.o.)$ rispetto alla misura di Lebesgue se la proprietà è verificata eccetto che su insieme di misura di Lebesgue nulla.

• **Esempio 2.3.4.** Dato lo spazio fondamentale $\Omega =]0,1]$ e la misura di probabilità P dell'Esempio 2.3.3, si consideri l'evento $E = \Omega \cap \mathbb{Q}$, dove $\mathbb{Q}$ rappresenta l'insieme dei numeri razionali. Dunque, E è l'insieme dei numeri razionali in $]0,1]$. Dal momento che $\text{card}(E) = \text{card}(\mathbb{Q})$, allora l'evento E è una unione numerabile di eventi elementari. Di conseguenza, si ha $P(E) = 0$ anche se l'insieme E è denso in $]0,1]$. Quindi, l'evento E^c si verifica $q.c.$ □

• **Esempio 2.3.5.** (Insieme di Cantor) Dato lo spazio fondamentale $\Omega =]0,1]$ e la misura di probabilità P dell'Esempio 2.3.3, si consideri l'evento E di Ω costruito in modo iterativo come segue. Si rimuove l'intervallo centrale $]\frac{1}{3}, \frac{2}{3}[$ da Ω, ottenendo l'evento

$$E_1 =]0, \tfrac{1}{3}] \cup [\tfrac{2}{3}, 1] \,,$$

per cui risulta $P(E_1) = \frac{2}{3}$. Successivamente, si rimuove l'intervallo centrale da $]0, \frac{1}{3}]$ e $[\frac{2}{3}, 1]$, ottenendo l'evento

$$E_2 =]0, \tfrac{1}{9}] \cup [\tfrac{2}{9}, \tfrac{1}{3}] \cup [\tfrac{2}{3}, \tfrac{7}{9}] \cup [\tfrac{8}{9}, 1] \,,$$

per cui risulta $P(E_2) = \frac{4}{9}$. Continuando il procedimento, alla n-esima iterazione l'evento E_n è composto dall'unione di 2^n intervalli disgiunti, per cui si ha $P(E_n) = 2^n 3^{-n}$. Si consideri dunque l'evento $E = \bigcap_{n=1}^{\infty} E_n$, che nel linguaggio di Teoria della Misura è detto insieme di Cantor. In effetti, questo insieme è stato introdotto dal matematico tedesco Georg Cantor (1845-1918), uno dei padri fondatori della Teoria degli Insiemi. Si osservi che $E_{n+1} \subset E_n$ e quindi la successione di eventi è decrescente, per cui, tenendo presente il Teorema 2.2.10, si ha

$$P(E) = P\left(\bigcap_{n=1}^{\infty} E_n\right) = \lim_n P(E_n) = \lim_n \frac{2^n}{3^n} = 0 \,.$$

Dal momento che si può dimostrare che E è non numerabile (si veda Dudley, 2004, p.105), allora E è un evento di probabilità nulla che contiene un'infinità non numerabile di risultati. In altre parole, l'evento E^c si verifica $q.c.$ □

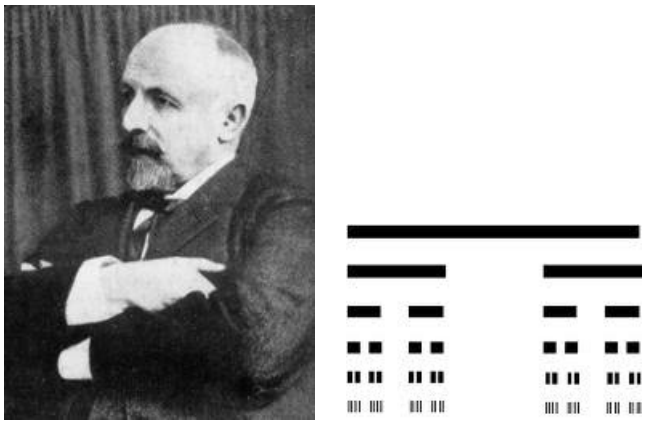


Figura 2.3.3. Georg Cantor (1845-1918) e
le prime cinque iterazioni per la costruzione dell'insieme di Cantor.

2.4. Probabilità condizionata

Dato lo spazio probabilizzato $(\Omega, \mathcal{F}, P)$, supponiamo che si sia verificato l'evento $E_0 \in \mathcal{F}$. Alla luce di questa nuova conoscenza, è ovvio che lo spazio fondamentale si riduce. I risultati possibili, infatti, non sono più tutte le eventualità che compongono Ω, ma solo le eventualità che compongono E_0. Risultano dunque impossibili, e dovranno quindi avere probabilità nulla, tutti gli eventi di $\mathcal{F}$ che sono incompatibili con E_0, mentre risulta certo, e dovrà quindi avere probabilità unitaria, l'evento E_0. In pratica, le probabilità dovranno essere riassegnate su $\mathcal{F}$ in modo da rispettare questi vincoli.

Definizione 2.4.1. Sia $(\Omega, \mathcal{F}, P)$ uno spazio probabilizzato. Se $E, E_0 \in \mathcal{F}$ e $P(E_0) > 0$, la *probabilità condizionata* di E dato E_0 è data da

$$P(E \mid E_0) = \frac{P(E \cap E_0)}{P(E_0)} \, . \qquad \square$$

Si noti che se $P(E_0) = 0$, la probabilità condizionata non è definita. Al variare di E in $\mathcal{F}$, la probabilità condizionata $P(E \mid E_0)$ descrive una misura di probabilità $P(\cdot \mid E_0)$ che è coerente con i tre assiomi della probabilità, come risulta dalla seguente Proposizione.

Proposizione 2.4.2. *Sia $(\Omega, \mathcal{F}, P)$ uno spazio probabilizzato. Se $E_0 \in \mathcal{F}$ e $P(E_0) > 0$, allora $P(\cdot \mid E_0)$ è una misura di probabilità su $\mathcal{F}$.*

Dimostrazione. Occorre dimostrare che $P(\cdot \mid E_0)$ soddisfa i tre assiomi della Definizione 2.1.2. Per ogni $E \in \mathcal{F}$

$$P(E \mid E_0) = \frac{P(E \cap E_0)}{P(E_0)} \geq 0 \, ,$$

in quanto $P(E \cap E_0) \geq 0$, mentre $P(E_0) > 0$ per ipotesi. Inoltre,

$$P(\Omega \mid E_0) = \frac{P(\Omega \cap E_0)}{P(E_0)} = \frac{P(E_0)}{P(E_0)} = 1 \, .$$

Infine, data successione di eventi incompatibili $(E_n)_{n \geq 1} \in \mathcal{F}$, si ha

$$\begin{aligned} P\left(\bigcup_{n=1}^{\infty} E_n \mid E_0 \right) &= \frac{P(\bigcup_{n=1}^{\infty} E_n \cap E_0)}{P(E_0)} = \frac{P(\bigcup_{n=1}^{\infty} (E_n \cap E_0))}{P(E_0)} \\ &= \frac{\sum_{n=1}^{\infty} P(E_n \cap E_0)}{P(E_0)} = \sum_{n=1}^{\infty} P(E_n \mid E_0) \, , \end{aligned}$$

in quanto $(E_n \cap E_0)_{n \geq 1}$ è una successione di eventi incompatibili. $\square$

Dalla definizione di probabilità condizionata segue il cosiddetto *Principio delle Probabilità Composte*, il quale permette di esprimere la probabilità dell'intersezione di due eventi $E_1, E_2 \in \mathcal{F}$ nel seguente modo

$$P(E_1 \cap E_2) = P(E_2 \mid E_1) P(E_1)$$

o, alternativamente, come

$$P(E_1 \cap E_2) = P(E_1 \mid E_2) P(E_2) \, .$$

Il Principio delle Probabilità Composte può essere applicato in modo ricorsivo per determinare la probabilità dell'intersezione di più di due eventi. Per esempio, la probabilità dell'intersezione di tre eventi $E_1, E_2, E_3 \in \mathcal{F}$ risulta

$$\begin{aligned} P(E_1 \cap E_2 \cap E_3) &= P((E_1 \cap E_2) \cap E_3) = P(E_3 \mid E_1 \cap E_2) P(E_1 \cap E_2) \\ &= P(E_3 \mid E_1 \cap E_2) P(E_2 \mid E_1) P(E_1) \, . \end{aligned}$$

In generale, dati n eventi $(E_k)_{k=1}^{n} \in \mathcal{F}$, si verifica per induzione che la probabilità della loro intersezione risulta

$$P\left(\bigcap_{n=1}^{\infty} E_n \right) = P(E_n \mid E_1 \cap E_2 \cap \ldots \cap E_{n-1}) \times \cdots \times P(E_2 \mid E_1) P(E_1) \, .$$

Teorema 2.4.3. (Legge delle Probabilità Totali) *Se $(\Omega, \mathcal{F}, P)$ è uno spazio probabilizzato e se gli eventi $(E_k)_{k=1}^n \in \mathcal{F}$ costituiscono una partizione di Ω con $P(E_n) > 0$ per ogni $k = 1, \dots, n$, allora per ogni $E \in \mathcal{F}$ si ha*

$$P(E) = \sum_{k=1}^{n} P(E \mid E_k) P(E_k) \, .$$

Dimostrazione. Dal momento che $E \subset \Omega$, si ha

$$E = E \cap \Omega = E \cap \bigcup_{k=1}^{n} E_k = \bigcup_{k=1}^{n} (E \cap E_k) \, .$$

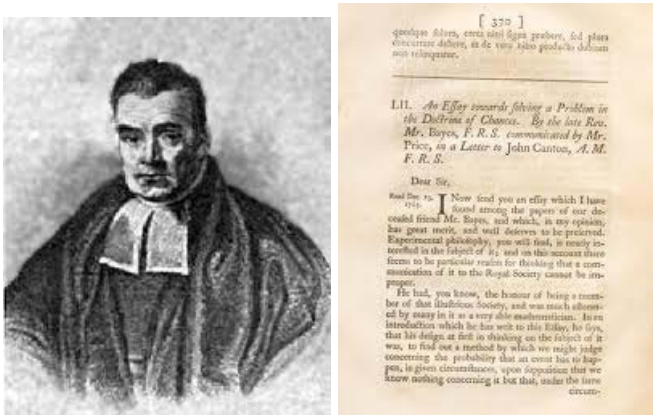
Inoltre, gli eventi $(E_k)_{k=1}^n$ sono tra loro incompatibili e quindi anche le loro intersezioni con l'evento E risultano incompatibili, per cui si ha

$$P(E) = P\left(\bigcup_{k=1}^{n} (E \cap E_k) \right) = \sum_{k=1}^{n} P(E \cap E_k) = \sum_{k=1}^{n} P(E \mid E_k) P(E_k) \, . \qquad \square$$

Dalla definizione di probabilità condizionata si può esprimere la probabilità di un evento, condizionata al verificarsi di un ulteriore evento, in termini della probabilità di quest'ultimo condizionata al verificarsi del primo, ovvero se $E_1, E_2 \in \mathcal{F}$ e $P(E_1), P(E_2) > 0$, allora

$$P(E_2 \mid E_1) = \frac{P(E_1 \mid E_2) P(E_2)}{P(E_1)} \, .$$

Questa riscrittura della probabilità condizionata dà luogo alla celebrata Formula di Bayes, che prende nome dal reverendo Thomas Bayes (1702-1761), matematico inglese autore di un saggio (pubblicato postumo nel 1763) che ha posto le basi di moderne correnti di pensiero nella Teoria della Probabilità e nella Statistica, ovvero *Essay Towards Solving a Problem in the Doctrine of Chances* (si veda von Plato, 1994).



[370]

LII. *An Essay towards solving a Problem in the Doctrine of Chances. By the late Rev. Mr. Bayes, F. R. S. communicated by Mr. Price, in a Letter to* John Canton, *A. M. F. R. S.*

Dear Sir,

Read Dec. 23, 1763. I Now send you an essay which I have found among the papers of our deceased friend Mr. Bayes, and which, in my opinion, has great merit, and well deserves to be preserved. Experimental philosophy, you will find, is nearly interested in the subject of it; and on this account there seems to be particular reason for thinking that a communication of it to the Royal society cannot be improper.

He had, you know, the honour of being a member of that illustrious Society, and was much esteemed by many in it as a very able mathematician. In an introduction which he has writ to this Essay, he says, that his design at first in thinking on the subject of it was, to find out a method by which we might judge concerning the probability that an event has to happen, in given circumstances, upon supposition that we know nothing concerning it but that, under the same circum-

Figura 2.4.1. Probabile ritratto di Thomas Bayes (1702-1761) e frontespizio di *Essay Towards Solving a Problem in the Doctrine of Chances* (1763).

Teorema 2.4.4. (Formula di Bayes) *Sia $(\Omega, \mathcal{F}, P)$ uno spazio probabilizzato. Se gli eventi $(E_k)_{k=1}^n \in \mathcal{F}$ costituiscono una partizione di Ω con $P(E_k) > 0$ per ogni $k = 1, \dots, n$, allora dato $E \in \mathcal{F}$ con $P(E) > 0$ si ha*

$$P(E_k \mid E) = \frac{P(E \mid E_k) P(E_k)}{P(E)} = \frac{P(E \mid E_k) P(E_k)}{\sum_{k=1}^{n} P(E \mid E_k) P(E_k)} \, .$$

Dimostrazione. Dalla definizione di probabilità condizionata, dal Principio delle Probabilità Composte e dalla Legge delle Probabilità Totali (Teorema 2.4.3) si ottiene immediatamente la tesi. □

Da un punto di vista formale, la Formula di Bayes sembra essere una mera riformulazione della probabilità condizionata, ottenuta sulla base del Principio delle Probabilità Composte e della Legge delle Probabilità Totali. Tuttavia, a causa di alcune sue interpretazioni logiche, come evidenziato in precedenza, la formula di Bayes ha assunto una rilevante importanza applicativa, soprattutto nell'ambito della statistica inferenziale (vedi von Plato, 1994). Per quanto riguarda la generalizzazione della Legge delle Probabilità Totali e della Formula di Bayes, si rimanda al Capitolo 5.

• **Esempio 2.4.1.** Si considerino 1000 palline che sono inserite in 3 urne che contengono rispettivamente 200, 700 e 100 palline. Se si suppone di considerare un'estrazione di una pallina da un'urna scelta in modo casuale, allora lo spazio fondamentale risulta $\Omega = \{\omega_1, \ldots, \omega_{1000}\}$, dove ω_k rappresenta il risultato relativo all'estrazione della pallina k-esima. Se si indicano con E_1, E_2 e E_3 rispettivamente l'evento che la pallina sia estratta dalla prima, seconda e terza urna, allora si ha $P(E_1) = \frac{1}{5}$, $P(E_2) = \frac{7}{10}$ e $P(E_3) = \frac{1}{10}$. Inoltre, si supponga che la prima urna contenga 20 palline rosse, che la seconda urna contenga 35 palline rosse e che l'ultima urna non contenga nessuna pallina rossa. Se si indica con E l'evento che una pallina rossa è stata estratta, allora si ha $P(E \mid E_1) = \frac{1}{10}$, $P(E \mid E_2) = \frac{1}{20}$ e $P(E \mid E_3) = 0$. Dal momento che gli eventi E_1, E_2 e E_3 costituiscono una partizione dello spazio fondamentale, applicando la Legge delle Probabilità Totali si ottiene

$$P(E) = \sum_{k=1}^{3} P(E \mid E_k)P(E_k) = \frac{11}{200}\,.$$

Applicando inoltre la Formula di Bayes, si è in grado di ottenere la probabilità $P(E_k \mid E)$, ovvero la probabilità dell'evento che la pallina sia stata estratta dalla k-esima urna, condizionata all'evento che sia stata estratta una pallina rossa. Si ottengono dunque le seguenti probabilità

$$P(E_1 \mid E) = \frac{P(E \mid E_1)P(E_1)}{P(E)} = \frac{4}{11}$$

e

$$P(E_2 \mid E) = \frac{P(E \mid E_2)P(E_2)}{P(E)} = \frac{7}{11}\,,$$

mentre

$$P(E_3 \mid E) = \frac{P(E \mid E_3)P(E_3)}{P(E)} = 0\,.$$

In pratica, se è stata estratta una pallina rossa, allora la pallina è stata estratta dalla seconda urna con la maggiore probabilità e dalla terza urna con la minore probabilità. □

• **Esempio 2.4.2.** (Paradosso delle scatole di Bertrand) Si considerino tre scatole in modo tale che la prima scatola contenga due monete d'oro, la seconda scatola contenga due monete d'argento e la terza scatola una moneta d'oro e una d'argento. L'esperimento aleatorio consiste nel considerare l'estrazione di una moneta da una scatola scelta in modo equiprobabile. In questo caso, se la moneta estratta è d'oro si vuole conoscere la probabilità che anche la restante moneta sia d'oro. Ingenuamente, il neofita è portato a pensare che questa probabilità sia pari a $\frac{1}{2}$, non tenendo presente il presente contesto di probabilità condizionata. Formalmente, se si indicano con E_1, E_2 e E_3 rispettivamente gli eventi che la moneta sia estratta dalla prima, seconda e terza scatola, allora si ha $P(E_1) = P(E_2) = P(E_3) = \frac{1}{3}$. Inoltre, se E rappresenta l'evento che la moneta estratta sia d'oro, allora $P(E \mid E_1) = 1$ e

$P(E \mid E_2) = 0$, mentre $P(E \mid E_3) = \frac{1}{2}$. Dunque, applicando la Legge delle Probabilità Totali si ottiene $P(E) = \frac{1}{2}$, da cui la probabilità richiesta risulta

$$P(E_1 \mid E) = \frac{P(E \mid E_1)P(E_1)}{P(E)} = \frac{2}{3} ,$$

ovvero un risultato controintuitivo. Questo è il motivo per cui il problema è detto paradosso delle scatole di Bertrand, dal momento che fu introdotto dal probabilista Joseph Bertrand (1822-1900) nel suo testo *Calcul des probabilités*, pubblicato nel 1889. □

2.5. Indipendenza stocastica

Si considerino due eventi E_1 e E_2 tali che $P(E_1), P(E_2) > 0$. Dal momento che, sulla base della definizione di probabilità condizionata, il verificarsi dell'evento E_1 modifica generalmente la valutazione probabilistica relativa all'evento E_2, si può allora concludere che

$$P(E_2 \mid E_1) = \frac{P(E_2 \cap E_1)}{P(E_1)} \neq P(E_2) .$$

Se questo non avviene, si deve verificare necessariamente che

$$P(E_2 \mid E_1) = \frac{P(E_2 \cap E_1)}{P(E_1)} = P(E_2) .$$

In base a queste considerazioni è naturale introdurre la seguente definizione di indipendenza stocastica.

Definizione 2.5.1. Sia $(\Omega, \mathcal{F}, P)$ uno spazio probabilizzato. Se $E_1, E_2 \in \mathcal{F}$, gli eventi E_1 e E_2 sono detti *stocasticamente indipendenti* (o semplicemente *indipendenti*) se

$$P(E_1 \cap E_2) = P(E_1)P(E_2) . \qquad \square$$

Si noti che la definizione di indipendenza dipende dalla misura di probabilità che viene adottata. In altri termini, due eventi che risultano indipendenti con una assegnazione di probabilità, potrebbero non esserlo con una differente assegnazione di probabilità.

• **Esempio 2.5.1.** Si consideri il lancio di un dado e gli eventi $E_1 = \{1, 2, 3, 4\}$, $E_2 = \{3, 4, 5, 6\}$ e $E_3 = \{2, 4, 6\}$. Se gli eventi elementari sono equiprobabili, ovvero se $P(\{k\}) = \frac{1}{6}$ per ogni $k = 1, \ldots, 6$, allora $P(E_1) = P(E_2) = \frac{2}{3}$ e $P(E_3) = \frac{1}{2}$. Gli eventi E_1 e E_2 non sono indipendenti in quanto, risultando $E_1 \cap E_2 = \{3, 4\}$, si ha

$$P(E_1 \cap E_2) = \frac{1}{3} \neq P(E_1)P(E_2) = \frac{4}{9} .$$

Al contrario, gli eventi E_1 e E_3 sono indipendenti in quanto, risultando $E_1 \cap E_3 = \{2, 4\}$, si ha

$$P(E_1 \cap E_3) = \frac{1}{3} = P(E_1)P(E_3) .$$

Se invece si considera una ulteriore assegnazione di probabilità $P(\{k\}) = p_k$, dove $p_1 = \frac{1}{2}$ e $p_2 = \ldots = p_6 = \frac{1}{10}$, allora si ha $P(E_1) = \frac{4}{5}$ e $P(E_3) = \frac{3}{10}$. In questo caso, al contrario di quanto avviene con la precedente assegnazione di probabilità, gli eventi E_1 e E_3 non sono indipendenti in quanto si ha

$$P(E_1 \cap E_3) = \frac{1}{5} \neq P(E_1)P(E_2) = \frac{6}{25} \,. \qquad \square$$

Proposizione 2.5.2. *Sia $(\Omega, \mathcal{F}, P)$ uno spazio probabilizzato. Se $E_1, E_2 \in \mathcal{F}$ sono eventi indipendenti, allora anche le coppie di eventi E_1^c e E_2, E_1 e E_2^c, E_1^c e E_2^c sono indipendenti.*

Dimostrazione. Si considerino gli eventi E_1^c e E_2. Dal momento che

$$E_2 = (E_1 \cap E_2) \cup (E_1^c \cap E_2) \,,$$

dove $(E_1 \cap E_2)$ e $(E_1^c \cap E_2)$ sono eventi incompatibili, allora risulta

$$P(E_2) = P(E_1 \cap E_2) + P(E_1^c \cap E_2) = P(E_1)P(E_2) + P(E_1^c \cap E_2) \,,$$

da cui

$$P(E_1^c \cap E_2) = P(E_2) - P(E_1)P(E_2) = P(E_2)(1 - P(E_1)) = P(E_2)P(E_1^c) \,,$$

che prova l'indipendenza di E_1^c e E_2. In modo del tutto analogo, dal momento che

$$E_1 = (E_1 \cap E_2) \cup (E_1 \cap E_2^c) \,,$$

si dimostra l'indipendenza di E_1 e E_2^c. Infine, dal momento che $E_1^c \cap E_2^c = (E_1 \cup E_2)^c$, si ottiene

$$\begin{aligned} P(E_1^c \cap E_2^c) &= 1 - P(E_1 \cup E_2) = 1 - (P(E_1) + P(E_2) - P(E_1 \cap E_2)) \\ &= 1 - P(E_1) - P(E_2) + P(E_1)P(E_2) \\ &= (1 - P(E_1))(1 - P(E_2)) = P(E_1^c)P(E_2^c) \,, \end{aligned}$$

ovvero l'indipendenza di E_1^c e E_2^c. $\square$

Proposizione 2.5.3. *Se $(\Omega, \mathcal{F}, P)$ è uno spazio probabilizzato, gli eventi $\emptyset$ e Ω sono indipendenti da ogni altro evento $E \in \mathcal{F}$.*

Dimostrazione. Si ha

$$P(E \cap \emptyset) = P(\emptyset) = P(E)P(\emptyset) \,,$$

mentre

$$P(E \cap \Omega) = P(E) = P(E)P(\Omega) \,,$$

dal momento che $P(\emptyset) = 0$ e $P(\Omega) = 1$. $\square$

Proposizione 2.5.4. *Se $(\Omega, \mathcal{F}, P)$ è uno spazio probabilizzato, un evento $E \in \mathcal{F}$ è indipendente da se stesso se e solo se $P(E) = 0$ o $P(E) = 1$.*

Dimostrazione. Se E è indipendente da se stesso si ha

$$P(E) = P(E \cap E) = P(E)^2 \,,$$

da cui segue che necessariamente che $P(E) = 0$ o $P(E) = 1$. L'affermazione inversa è immediata. $\square$

Definizione 2.5.5. Sia $(\Omega, \mathcal{F}, P)$ uno spazio probabilizzato. Gli eventi $(E_k)_{k=1}^n \in \mathcal{F}$ sono detti *reciprocamente indipendenti* se

$$P(E_k \cap E_j) = P(E_k)P(E_j)$$

per ogni $k \neq j = 1, \ldots, n$, mentre sono detti *completamente indipendenti* se per ogni possibile scelta $1 \leq j_1 < \ldots < j_k \leq n$ di k eventi con $k = 2, \ldots, n$, si ha

$$P\left(\bigcap_{l=1}^{k} E_{j_l}\right) = \prod_{l=1}^{k} P(E_{j_l}) \,. \qquad \square$$

Evidentemente l'indipendenza completa è una condizione molto più restrittiva dell'indipendenza reciproca. In effetti, al fine di verificare che n eventi sono reciprocamente indipendenti devono essere soddisfatte $\binom{n}{2}$ condizioni, mentre se gli eventi sono completamente indipendenti devono sussistere $(2^n - n - 1)$ condizioni.

• **Esempio 2.5.2.** Si consideri un'urna contenente 8 palline contrassegnate dai primi 8 numeri interi. Lo spazio fondamentale relativo all'estrazione aleatoria di una pallina dall'urna è dato da $\Omega = \{1, \ldots, 8\}$, dove ogni risultato rappresenta l'estrazione della pallina contraddistinta dal corrispondente numero. Se ciascuna pallina ha la stessa probabilità di essere estratta, ovvero se $P(\{k\}) = \frac{1}{8}$ per ogni $k = 1, \ldots, 8$, gli eventi $E_1 = \{1, 2, 3, 4\}$, $E_2 = \{2, 4, 6, 8\}$ e $E_3 = \{3, 6\}$ hanno probabilità $P(E_1) = P(E_2) = \frac{1}{2}$ e $P(E_3) = \frac{1}{4}$. Essendo $E_1 \cap E_2 = \{2, 4\}$, si ha inoltre

$$P(E_1 \cap E_2) = P(E_1)P(E_2) = \frac{1}{4} \,,$$

mentre, essendo $E_1 \cap E_3 = \{3\}$, allora

$$P(E_1 \cap E_3) = P(E_1)P(E_3) = \frac{1}{8} \,.$$

Infine, essendo $E_2 \cap E_3 = \{6\}$, si ha

$$P(E_2 \cap E_3) = P(E_2)P(E_3) = \frac{1}{8} \,.$$

Sulla base di questi risultati si può concludere che i tre eventi sono reciprocamente indipendenti. Tuttavia, gli stessi eventi non sono completamente indipendenti in quanto essendo $E_1 \cap E_2 \cap E_3 = \emptyset$, risulta

$$P(E_1 \cap E_2 \cap E_3) = 0 \neq P(E_1)P(E_2)P(E_3) = \frac{1}{16} \,.$$

Con la medesima assegnazione di probabilità, si considerino inoltre gli eventi $E_1 = \{1, 2, 3, 4\}$, $E_2 = \{1, 5, 6, 7\}$ e $E_3 = \{1, 5, 6, 8\}$, per cui risulta ovviamente $P(E_1) = P(E_2) = P(E_3) = \frac{1}{2}$. In questo caso si ha

$$P(E_1 \cap E_2 \cap E_3) = P(\{1\}) = \frac{1}{8} = P(E_1)P(E_2)P(E_3) \,.$$

Tuttavia gli eventi non sono completamente indipendenti in quanto

$$P(E_1 \cap E_2) = P(\{1\}) = \frac{1}{8} \neq P(E_1)P(E_2) = \frac{1}{4} \,. \qquad \square$$

2.6. Costruzione di spazi probabilizzati prodotto

Si considerino n esperimenti aleatori che danno luogo ad altrettanti spazi probabilizzati $(\Omega_k, \mathcal{F}_k, P_k)$, dove $k = 1, \ldots, n$. Sia dunque $\Omega = \Omega_1 \times \cdots \times \Omega_n$ il relativo spazio fondamentale prodotto. Come evidenziato nella Sezione 1.5, se Ω ha cardinalità finita o numerabile, la σ-algebra

prodotto $\mathcal{F}$ può essere scelta come l'insieme delle parti $\mathcal{P}(\Omega)$. In questo caso, per definire una misura di probabilità P sugli eventi di $\mathcal{F}$, è allora sufficiente probabilizzare gli eventi elementari di Ω. Se Ω è non numerabile, allora il problema risulta notevolmente più complesso. In effetti, se la σ-algebra prodotto $\mathcal{F}$ è scelta come la σ-algebra generata dalla classe $\mathcal{E}$ degli eventi rettangolari, allora si può dimostrare che è sufficiente probabilizzare gli eventi di $\mathcal{E}$ per ottenere in modo univoco una misura di probabilità P anche sugli eventi di $\mathcal{F}$ (si veda Dudley, 2004, p.255).

Una situazione di particolare interesse è quella in cui ciascun esperimento aleatorio non ha influenza sui rimanenti esperimenti aleatori, ovvero quando si richiede che

$$P(E_1 \times \cdots \times E_n) = \prod_{k=1}^{n} P_k(E_k) \,,$$

per ogni $E_1 \in \mathcal{F}_1, \ldots, E_n \in \mathcal{F}_n$. La misura P è detta *probabilità prodotto* ed è indicata con $P = P_1 \otimes \cdots \otimes P_n$. Si può dimostrare l'esistenza e l'unicità della misura di probabilità prodotto (si veda Dudley, 2004, p.255). La precedente relazione definisce in effetti il concetto di indipendenza per la classe di σ-algebre $(\mathcal{F}_k)_{k=1}^n$.

Nella Teoria della Misura, dato lo spazio misurabile prodotto $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$, la misura di Lebesgue λ^n in $\mathbb{R}^n$ assegna un "ipervolume" ad ogni insieme rettangolare $]a_1, b_1] \times \cdots \times]a_n, b_n]$ di $\mathbb{R}^n$, ovvero

$$\lambda^n(]a_1, b_1] \times \cdots \times]a_n, b_n]) = \prod_{k=1}^{n} \lambda(]a_k, b_k]) \,.$$

La terna $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \lambda^n)$ è uno *spazio misurato prodotto*.

Se in generale $((\Omega_k, \mathcal{F}_k, P_k))_{k\in I}$, dove I è un insieme eventualmente non numerabile, è una classe di spazi probabilizzati, si può dimostrare che sulla σ-algebra prodotto $\mathcal{F}$ esiste un'unica misura di probabilità P tale che

$$P\left(\prod_{k\in J} E_k\right) = \prod_{k\in J} P_k(E_k) \,,$$

dove $(E_k)_{k\in J}$ è una qualsiasi classe di eventi cilindrici di Ω (si veda Dudley, 2004, p.255). In questo caso, P è detta *probabilità prodotto* e viene indicata con $P = \bigotimes_{k\in I} P_k$. Inoltre, la relazione definisce il concetto di indipendenza per la classe di σ-algebre $(\mathcal{F}_k)_{k\in J}$. Si noti infine che la terna $(\Omega, \mathcal{F}, P)$ è detta *spazio probabilizzato prodotto*.

• **Esempio 2.6.1.** Si lanci in modo indipendente un dado e una moneta e si consideri come unico esperimento aleatorio la combinazione dei due esperimenti aleatori. Si supponga che il dado e la moneta non siano truccati e quindi che le probabilità nei singoli spazi fondamentali Ω_1 e Ω_2 siano assegnate ad ogni evento elementare come $P_1(\{k\}) = \frac{1}{6}$, dove $k = 1, \ldots, 6$, e $P_2(\{j\}) = \frac{1}{2}$, dove $j = t, c$. Per ogni evento elementare $\{(k, j)\} \in \Omega = \Omega_1 \times \Omega_2$, la probabilità prodotto $P = P_1 \otimes P_2$ risulta dunque

$$P(\{(k, j)\}) = P_1(\{k\})P_2(\{j\}) = \frac{1}{12} \,.$$

Una volta assegnate le probabilità agli eventi elementari, risultano determinate anche le probabilità dei 2^{12} eventi che compongono la σ-algebra $\mathcal{F} = \mathcal{P}(\Omega)$ relativa all'esperimento aleatorio combinato. In particolare, la probabilità dell'evento $E = \{(1, c), (2, c)\}$ è data da $P(E) = \frac{1}{6}$ e la probabilità dell'evento $E = \{(3, t), (4, t), (3, c)\}$ è data da $P(E) = \frac{1}{4}$. Si osservi come l'assegnazione sia coerente con le probabilità relative ai singoli esperimenti aleatori. Per esempio, l'evento cilindrico $E = \{(2, t), (2, c)\}$ ha probabilità $P(E) = \frac{1}{6}$, che equivale appunto alla probabilità che nel lancio del dado si sia verificata la faccia contrassegnata dal simbolo 2. □

• **Esempio 2.6.2.** (Problema del Cavalier de Méré) Antoine Gombaud (1607-1684), detto Cavalier de Méré, fu uno scrittore francese e accanito giocatore d'azzardo che pose uno dei primi problemi di calcolo delle probabilità al matematico francese Blaise Pascal (1623-1662). Il Cavalier de Méré ripeteva 4 volte il lancio di un dado scommettendo che la faccia contrassegnata con 6 si sarebbe verificata almeno una volta. In questo caso, lo spazio fondamentale è dato da $\Omega = \Omega_1 \times \cdots \times \Omega_4$, dove Ω_k rappresenta lo spazio fondamentale relativo al k-esimo lancio. Supponendo che i lanci siano effettuati in modo indipendente, si consideri l'evento E_k che non si sia verificata la faccia 6 al k-esimo lancio. Quindi, se E è l'evento che non si sia verificata la faccia 6 a nessun lancio, e considerando un'assegnazione equiprobabile per ogni lancio, si ha

$$P(E) = P(E_1 \times \cdots \times E_4) = \prod_{k=1}^{4} P_k(E_k) = \left(\frac{5}{6}\right)^4 .$$

Dunque, la probabilità che la faccia contrassegnata da 6 si sia verificata almeno in un lancio è data da

$$P(E^c) = 1 - \left(\frac{5}{6}\right)^4 \simeq 0.518 .$$

Figura 2.6.1. Blaise Pascal (1623-1662).

Figura 2.6.2. Pierre de Fermat (1601 o 1607/8-1665).

Il gioco era quindi leggermente favorevole per il Cavalier de Méré, che aveva intuito questo risultato semplicemente dalla sua esperienza di giocatore d'azzardo senza ovviamente conoscere l'esatta probabilità. Ingenuamente, il Cavalier de Méré suppose che, dal momento che i possibili risultati nel lancio di due dadi sono 36 (ovvero 6 volte i risultati nel lancio del singolo dado), allora ripetendo $4 \times 6 = 24$ volte il lancio di 2 dadi, sarebbe stato ugualmente favorevole scommettere che “il doppio

6” si sarebbe presentato almeno una volta. Tuttavia, con grande sorpresa, il Cavalier de Méré si accorse empiricamente che la giocata non era favorevole e fu spinto a porre il problema a Blaise Pascal, che successivamente ebbe uno scambio epistolare con l'altro grande matematico francese Pierre de Fermat (1601 o 1607/8-1665), in cui viene ottenuta la corretta soluzione. Formalmente, lo spazio fondamentale è dato da $\Omega = \Omega_1 \times \cdots \times \Omega_{24}$, dove in questo caso Ω_k rappresenta lo spazio fondamentale relativo al k-esimo lancio congiunto dei due dadi e che quindi può essere considerato a sua volta uno spazio prodotto di due spazi fondamentali. Se E rappresenta l'evento che non si sia verificato “il doppio 6” su nessun lancio congiunto dei due dadi, considerando un'assegnazione equiprobabile per ogni lancio, allora si ha

$$P(E) = P(E_1 \times \cdots \times E_{24}) = \prod_{k=1}^{24} P_k(E_k) = \left(\frac{35}{36}\right)^{24} .$$

Dunque, la probabilità che si sia verificato almeno in un lancio congiunto “il doppio 6” è data da

$$P(E^c) = 1 - \left(\frac{35}{36}\right)^{24} \simeq 0.491 .$$

Come aveva intuito il Cavalier de Méré, il gioco era quindi leggermente sfavorevole. Resta comunque sorprendente come il Cavalier de Méré fosse giunto a queste conclusioni solo sulla base dell'esperienza fatta al tavolo di gioco. Il problema posto dal Cavalier de Méré viene considerato comunemente come il vero inizio della Teoria della Probabilità. Fra l'altro si deve notare che Pascal incoraggiò il matematico, astronomo e fisico Christiaan Huygens (1629-1695) a scrivere il trattato *De ratiociniis in ludo aleae* (1657), largamente basato sullo studio probabilistico del gioco dei dadi. □

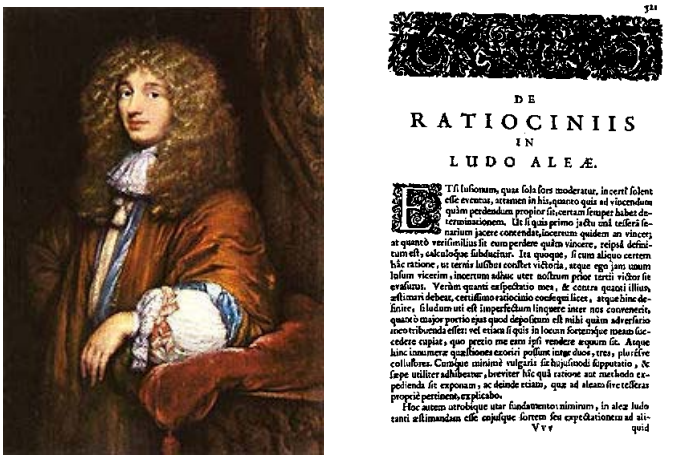

DE
RATIOCINIIS
IN
LUDO ALEÆ.

Figura 2.6.3. Christiaan Huygens (1629-1695) e frontespizio di *De ratiociniis in ludo aleae* (1657).

• **Esempio 2.6.3.** (Gioco della roulette) La *roulette* è un gioco d'azzardo la cui introduzione è stata attribuita, almeno nella sua forma primitiva, a Blaise Pascal. Nella versione europea della *roulette*, detta di tipo Monte Carlo, il gioco è basato su un disco diviso in 37 settori di uguale ampiezza e che sono numerati con i primi 36 interi e con lo zero. Alternativamente, nella versione americana della *roulette*, detta di tipo Las Vegas, il disco è diviso in 38 settori di uguale ampiezza e che sono numerati con i primi 36 interi, con lo zero e con il cosiddetto doppio zero. I settori sono colorati alternativamente in rosso e nero, mentre lo zero, così come il doppio zero quando è presente, sono normalmente colorati di verde. Il gioco consiste nel far ruotare il disco dal gestore del banco (il cosiddetto *croupier*), che successivamente vi lancia una pallina. La pallina viene fatta ruotare in senso opposto a quello della *roulette* e il numero vincente è quello relativo al settore in cui cade la pallina.

Le puntate vengono effettuate su un tavolo verde, su cui sono riportati i numeri della *roulette*, oltre a varie possibilità di differenti scommesse.

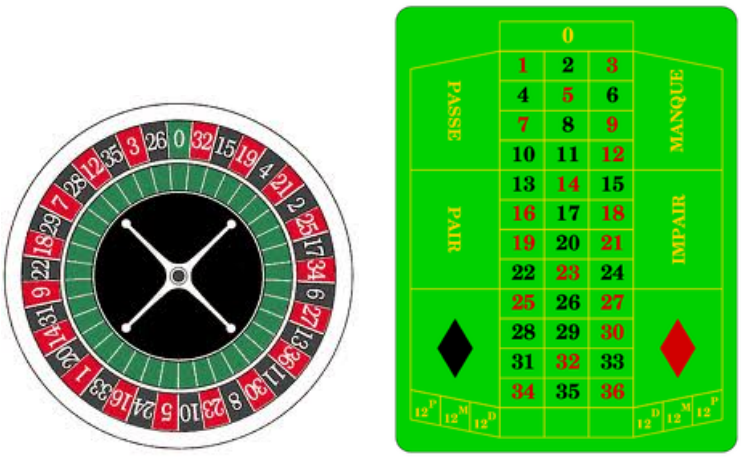


Figura 2.6.4. Disposizione dei numeri sulla *roulette* e sul tavole verde per una *roulette* di tipo Monte Carlo.

Ad esempio, si può scommettere su *Pair ou Impair*, ovvero che il numero uscito sia pari o dispari, su *Manque ou Passe*, ovvero che sia uscito un numero da 1 a 18 o un numero da 19 a 36, o su *Rouge ou Noir*, ovvero sul colore del numero uscito. Esistono ulteriori possibilità di scommessa, che sono usualmente definite con termini specifici nel linguaggio caratteristico del giocatore d'azzardo. Ovviamente, la procedura viene ripetuta numerose volte durante una serata di gioco. Anche se apparentemente è possibile considerare alla *roulette* una pluralità di differenti scommesse, in generale il gioco può essere riportato allo schema delle prove ripetute introdotto nell'Esempio 1.5.2. Ad esempio, se si decide di puntare sul rosso durante la serata di gioco, allora si ha in effetti uno schema di Bernoulli basato su n ripetizioni di un esperimento aleatorio con esito dicotomico (si veda l'Esempio 1.5.2), ognuna delle quali con spazio fondamentale $\Omega_k = \{\omega_1, \omega_2\}$, dove ω_1 è il risultato che rappresenta l'uscita di un numero colorato in rosso, mentre ω_2 è il risultato che rappresenta l'uscita di un numero colorato in nero o dello zero. Supponendo che il *croupier* sia onesto e la *roulette* sia bilanciata, si può considerare l'assegnazione di probabilità $P_k(\{\omega_1\}) = \frac{18}{37}$ e $P_k(\{\omega_2\}) = \frac{19}{37}$ sullo spazio fondamentale Ω_k. Dal momento che lo spazio fondamentale prodotto Ω ha cardinalità finita, la misura di probabilità prodotto P si ottiene probabilizzando i 2^n eventi elementari $\{(\omega_{j_1}, \ldots, \omega_{j_n})\}$ di Ω, ovvero

$$P(\{(\omega_{j_1}, \ldots, \omega_{j_n})\}) = \prod_{k=1}^{n} P_k(\{\omega_{j_k}\}) \,,$$

dove $j_k = 1, 2$ e $k = 1, \ldots, n$. Per un approfondimento delle strategie di gioco ottimali alla *roulette* (nel senso di strategie che permettono di limitare le perdite) si consideri la trattazione di Billingsley (1995, p.92). Si noti che nelle case da gioco europee la precedente assegnazione di probabilità può risultare modificata dalla possibilità di recuperare almeno in parte la giocata quando si presenta lo zero (mediante le regole basate su *la partage* o *en prison*) e che nel presente esempio non è stata adottata per semplicità di esposizione. □

2.7. Lemma di Borel-Cantelli e Legge zero-uno di Kolmogorov

In questa Sezione vengono considerati due risultati molto celebrati che riguardano successioni di eventi. Il seguente Lemma, che è fondamentale per ottenere risultati teorici di convergenza, è dovuto al probabilista italiano Francesco Paolo Cantelli (1875-1966) e a Émile Borel.

Teorema 2.7.1. (Lemma di Borel-Cantelli) *Dato lo spazio probabilizzato* $(\Omega, \mathcal{F}, P)$*, se* $(E_n)_{n\geq 1}$ *è una successione di eventi di* $\mathcal{F}$ *tale che* $\sum_{n=1}^{\infty} P(E_n) < \infty$*, allora*

$$P(\limsup_n E_n) = 0 \,.$$

Inoltre, se $(E_n)_{n\geq 1}$ *è una successione di eventi indipendenti di* $\mathcal{F}$ *tale che* $\sum_{n=1}^{\infty} P(E_n) = \infty$*, allora*

$$P(\limsup_n E_n) = 1 \,.$$

Dimostrazione. Tenendo presente la Proposizione 2.2.7 e il Teorema 2.2.8, dalla definizione di $\limsup_n E_n$ si ha

$$P(\limsup_n E_n) \leq P\left(\bigcup_{k=n}^{\infty} E_k\right) \leq \sum_{k=n}^{\infty} P(E_k)$$

per ogni n. Dal momento che $\sum_{n=1}^{\infty} P(E_n) < \infty$ per ipotesi, allora deve risultare

$$\lim_n \sum_{k=n}^{\infty} P(E_k) = 0 \,,$$

da cui segue la prima parte del Teorema. Inoltre, tenendo presente la relazione di De Morgan, si ha

$$\limsup_n E_n = \bigcap_{n=1}^{\infty} \left(\bigcap_{k=n}^{\infty} E_k^c\right)^c = \left(\bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} E_k^c\right)^c = (\liminf_n E_n^c)^c \,,$$

da cui

$$P(\limsup_n E_n) = 1 - P(\liminf_n E_n^c) \,.$$

Si noti che $\liminf_n E_n^c = \bigcup_{n=1}^{\infty} F_n$, dove $F_n = \bigcap_{k=n}^{\infty} E_k^c$, mentre $(F_n)_{n\geq 1}$ è una successione crescente di eventi. Dunque, dal Teorema 2.2.10 e dall'indipendenza degli eventi della successione risulta

$$P(\limsup_n E_n) = 1 - P(\lim_n F_n) = 1 - \lim_n P(F_n) = 1 - \lim_n \prod_{k=n}^{\infty} P(E_k^c) = 1 - \lim_n \prod_{k=n}^{\infty} (1 - P(E_k)) \,.$$

Tenendo presente la disuguaglianza $1 - x \leq e^{-x}$, per ogni n si ha

$$\prod_{k=n}^{\infty} (1 - P(E_k)) \leq \prod_{k=n}^{\infty} e^{-P(E_k)} = e^{-\sum_{k=n}^{\infty} P(E_k)} \,.$$

Quindi dal momento che $\sum_{n=1}^{\infty} P(E_n) = \infty$ e che $P(E_n) \leq 1$ per ogni n, allora deve risultare

$$\lim_n \sum_{k=n}^{\infty} P(E_k) = \infty \,,$$

e dunque

$$\lim_n \prod_{k=n}^{\infty} (1 - P(E_k)) = 0 \,,$$

da cui segue la seconda parte del Teorema. □

• **Esempio 2.7.1.** Si consideri lo schema di Bernoulli relativo ad un'infinità numerabile di ripetizioni del lancio di una moneta bilanciata e il relativo spazio probabilizzato prodotto $(\Omega, \mathcal{F}, P)$. Se E_n rappresenta l'evento di ottenere una sequenza di almeno n volte il simbolo “testa”, si ha

$$P(E_n) = 1 - \sum_{k=1}^{n} 2^{-k} = 2^{-n} .$$

Se si considera la successione di eventi $(E_n)_{n\geq 1}$, allora risulta $\sum_{n=1}^{\infty} P(E_n) = 1$ e quindi la prima parte del Lemma di Borel-Cantelli implica che $P(\limsup_n E_n) = 0$. In altre parole, una sequenza di simboli "testa" di lunghezza arbitraria si presenta *infinitamente spesso* con probabilità nulla. Se inoltre E_n è l'evento di ottenere il simbolo "testa" al $(2n-1)$-esimo e al $(2n)$-esimo lancio, si ha $P(E_n) = \frac{1}{4}$. Gli eventi della successione $(E_n)_{n\geq 1}$ sono indipendenti e risulta $\sum_{n=1}^{\infty} P(E_n) = \infty$. Quindi, dalla seconda parte del Lemma di Borel-Cantelli si ha $P(\limsup_n E_n) = 1$. In altre parole, una sequenza di due simboli "testa" si presenta *infinitamente spesso* con probabilità unitaria. Analogamente, si può verificare che una qualsiasi prefissata sequenza di lunghezza finita di simboli "testa" e "croce" si verifica infinitamente spesso. Se i simboli "testa" e "croce" vengono messi in corrispondenza biunivoca con i simboli "punto" e "linea" di un alfabeto Morse, allora il risultato appena ottenuto è stato interpretato in maniera pittoresca. Infatti, qualsiasi opera letteraria può essere rappresentata come una sequenza di simboli nell'alfabeto Morse. Quindi, adottando uno schema di Bernoulli di lanci ripetuti della moneta, questa opera "verrà scritta" dal lancio della moneta con probabilità unitaria, purché il lanciatore sia abbastanza paziente. □

Teorema 2.7.2. (Legge zero-uno di Kolmogorov) *Dato lo spazio probabilizzato* $(\Omega, \mathcal{F}, P)$*, se* $(E_n)_{n\geq 1}$ *è una successione di eventi indipendenti di* $\mathcal{F}$*, si consideri la successione di* σ*-algebre* $(\mathcal{F}_n)_{n\geq 1}$*, dove* $\mathcal{F}_n = \sigma(\mathcal{E}_n)$ *e* $\mathcal{E}_n = \{E_n, E_{n+1}, \dots\}$*. Se* $E \in \bigcap_{n=1}^{\infty} \mathcal{F}_n$*, allora si ha* $P(E) = 0$ *o* $P(E) = 1$.

Dimostrazione. Si veda Gut (2005, p.20). □

• **Esempio 2.7.2.** Si consideri una successione di eventi indipendenti $(E_n)_{n\geq 1}$ e sia dato l'ulteriore evento

$$E_x = \left\{ \omega \in \Omega : \lim_n \frac{1}{n} \sum_{k=1}^{n} \mathbf{1}_{E_k}(\omega) \leq x \right\},$$

dove $x \in \mathbb{R}$. In pratica, l'evento E_x si verifica se il limite della frequenza relativa degli eventi della successione che si verificano è minore o uguale a x. Si noti che dalla definizione di limite si ha

$$E_x = \bigcap_{m=1}^{\infty} \bigcup_{k=j}^{\infty} \bigcap_{n=k}^{\infty} \left\{ \omega \in \Omega : \frac{1}{n} \sum_{k=j}^{n} \mathbf{1}_{E_k}(\omega) \leq x + \frac{1}{m} \right\}$$

Inoltre, tenendo presente la notazione del Teorema 2.7.2, si ha

$$\left\{ \omega \in \Omega : \frac{1}{n} \sum_{k=j}^{n} \mathbf{1}_{E_k}(\omega) \leq x + \frac{1}{m} \right\} \in \mathcal{F}_n .$$

Dal momento che E_x può essere espresso come unione e intersezione numerabile di questi eventi, allora dal Teorema 2.7.2 si ha $P(E_x) = 0$ o $P(E_x) = 1$ per ogni x. □

2.8. Riferimenti bibliografici

L'approccio assiomatico è quello seguito nella maggioranza dei testi di Teoria della Probabilità, come ad esempio quelli segnalati nella Sezione 1.6. Si deve comunque evidenziare che l'approccio soggettivo alla probabilità, per quanto minoritario nei testi, ha notevole importanza in ambito

statistico. Per una trattazione di questo approccio si dovrebbero consultare i classici testi di de Finetti (1970, volume I e II) e Dubins e Savage (1965). Per una revisione degli approcci moderni alla probabilità si veda il testo di von Plato (1994). I problemi tecnici connessi alla costruzione della misura di probabilità sono estesamente considerati in Dudley (2004). Per quanto riguarda l'esemplificazione del calcolo delle probabilità con spazi finiti o numerabili si dovrebbero consultare i testi di Blom (1994), Gorroochurn (2012), Isaac (1995), Mosteller (1987) e Petkovic (2009), che contengono numerosi problemi classici. Erickson (2010), Graham *et al.* (1994) e Lovász (2003) sono testi introduttivi al calcolo combinatorio con parti dedicate al calcolo delle probabilità con spazi finiti. Infine, per una trattazione di paradossi e controesempi in probabilità, si veda i testi di Székely (1986), Romano e Siegel (1986) e Wise e Hall (1993).

2.9. Esercizi svolti

Sezione 2.1

• **Esercizio 2.1.1.** Si consideri lo spazio fondamentale $\Omega = \{\omega_1, \omega_2, \dots\}$ e la σ-algebra $\mathcal{F} = \mathcal{P}(\Omega)$. Inoltre, si assuma che

$$P(E) = \sum_{k=1}^{\infty} \mathbf{1}_E(\omega_k) p_k \, ,$$

dove $E \in \mathcal{F}$, mentre $p_k \geq 0$ per ogni $k = 1, 2, \dots$ con $\sum_{k=1}^{\infty} p_k = 1$. Si verifichi che $(\Omega, \mathcal{F}, P)$ è uno spazio probabilizzato.

Soluzione. Si deve dimostrare che P è in effetti una misura di probabilità. Il primo assioma è verificato, in quanto P è una applicazione a valori non negativi. Per quanto riguarda il secondo assioma, si ha

$$P(\Omega) = \sum_{k=1}^{\infty} \mathbf{1}_{\Omega}(\omega_k) p_k = \sum_{k=1}^{\infty} p_k = 1 \, .$$

Infine, per quanto riguarda il terzo assioma, se $(E_n)_{n \geq 1} \in \mathcal{F}$ è una successione di eventi incompatibili si ha

$$P\left(\bigcup_{n=1}^{\infty} E_n\right) = \sum_{k=1}^{\infty} \mathbf{1}_{\bigcup_{n=1}^{\infty} E_n}(\omega_k) p_k = \sum_{k=1}^{\infty} \sum_{n=1}^{\infty} \mathbf{1}_{E_n}(\omega_k) p_k = \sum_{n=1}^{\infty} \sum_{k=1}^{\infty} \mathbf{1}_{E_n}(\omega_k) p_k = \sum_{n=1}^{\infty} P(E_n) \, ,$$

dove la permutazione delle sommatorie è possibile in base al Teorema di Fubini. Dunque, $(\Omega, \mathcal{F}, P)$ è uno spazio probabilizzato. □

Sezione 2.2

• **Esercizio 2.2.1.** (Problema della segretaria distratta) In una segreteria vi sono n lettere e n buste preparate con l'indirizzo del relativo destinatario. Una segretaria distratta pone le lettere in modo casuale nelle buste. Si determini la probabilità che esattamente m lettere raggiungano correttamente il destinatario. Al fine della soluzione dell'esercizio, si tenga presente la formula di Waring ovvero, se si considera n eventi $(E_k)_{k=1}^n$, la probabilità dell'evento F_m che si verifica quando esattamente m di questi eventi si verificano è data da

$$P(F_m) = \sum_{k=m}^{n} \binom{k}{m} (-1)^{k-m} S_{k,n} \, ,$$

dove $m \in \{0, 1, \ldots, n\}$ e $S_{k,n}$ è definito nella Formula di Inclusione ed Esclusione con l'assunzione che $S_{0,n} = 1$. La formula prende il nome dal matematico inglese Edward Waring (1736-1798).

Soluzione. In questo caso, lo spazio fondamentale Ω è equivalente a quello relativo a tutte le possibili configurazioni di n palline distinguibili in n celle in modo tale che vi sia al massimo una pallina per cella. Dunque, risulta che $\text{card}(\Omega) = n!$, ovvero gli eventi elementari di Ω sono in corrispondenza biunivoca con le possibili permutazioni dei primi n interi. Inoltre, si può scegliere $\mathcal{F} = \mathcal{P}(\Omega)$. Dalle assunzioni fatte le $n!$ configurazioni sono ugualmente probabili, ovvero ad ogni evento elementare $\omega_k \in \Omega$ si assegna la probabilità $P(\{\omega_k\}) = (n!)^{-1}$. Si denoti con E_k l'evento che la k-esima lettera sia correttamente posta nella rispettiva busta. Dunque, per ogni scelta di indici $1 \leq j_1 < \ldots < j_k \leq n$ si ha

$$P(E_{j_1} \cap \ldots \cap E_{j_k}) = \frac{(n-k)!}{n!}$$

e quindi, dal momento che vi sono $\binom{n}{k}$ possibili scelte di questi indici, risulta

$$S_{k,n} = \binom{n}{k} \frac{(n-k)!}{n!} = \frac{1}{k!} \, .$$

La probabilità richiesta è dunque data da

$$P(F_m) = \sum_{k=m}^{n} \binom{k}{m} \frac{(-1)^{k-m}}{k!} \, .$$

Ovviamente, dal momento che la classe di eventi $(F_m)_{m=0}^n$ costituisce una partizione di Ω, risulta $\sum_{m=0}^n P(F_m) = 1$. Inoltre, si osservi che la probabilità che nessuna lettera raggiunga correttamente il destinatario è data da

$$P(F_0) = \sum_{k=0}^{n} \frac{(-1)^k}{k!}$$

e quindi la probabilità che almeno una lettera raggiunga il destinatario è data da

$$P\left(\bigcup_{k=1}^{n} E_k\right) = 1 - P(F_0) = \sum_{k=1}^{n} \frac{(-1)^{k+1}}{k!} \, ,$$

ovvero si è ottenuto in modo alternativo il risultato fornito dalla Formula di Inclusione ed Esclusione. Sulla base della serie esponenziale, per n abbastanza elevato, la probabilità che nessuna lettera raggiunga correttamente il destinatario può essere approssimata da $e^{-1} \simeq 0.368$. Quindi, la probabilità che almeno una lettera raggiunga correttamente il destinatario può essere approssimata da $1 - e^{-1} \simeq 0.632$. Per una semplice dimostrazione della formula di Waring si veda l'Esercizio 7.4.3. □

• **Esercizio 2.2.2.** (Problema della segretaria distratta, seconda parte) Con le medesime assunzioni dell'Esercizio 2.2.1, si determini la probabilità che almeno m lettere raggiungano correttamente il destinatario. Al fine di ottenere la soluzione dell'esercizio, si dimostri il seguente risultato. Considerati n eventi $(E_k)_{k=1}^n$, la probabilità dell'evento A_m che si verifica quando almeno m di questi eventi si verificano è data da

$$P(A_m) = \sum_{k=m}^{n} \binom{k-1}{m-1} (-1)^{k-m} S_{k,n} \, ,$$

dove $m \in \{1, \ldots, n\}$ e $S_{k,n}$ è definito nella Formula di Inclusione ed Esclusione.

Soluzione. Tenendo presente l'Esercizio 2.2.1, risulta $P(A_m) = \sum_{k=m}^n P(F_k)$. L'espressione richiesta si ottiene dunque verificando la relazione

$$P(A_{m+1}) = P(A_m) - P(F_m)\ .$$

per $m = 0, 1, \dots n-1$. In effetti, per le proprietà dei coefficienti binomiali, si ha

$$\begin{aligned} P(A_m) - P(F_m) &= \sum_{k=m}^{n} \binom{k-1}{m-1} (-1)^{k-m} S_{k,n} - \sum_{k=m}^{n} \binom{k}{m} (-1)^{k-m} S_{k,n} \\ &= \sum_{k=m}^{n} \left[\binom{k-1}{m-1} - \binom{k}{m} \right] (-1)^{k-m} S_{k,n} \\ &= \sum_{k=m+1}^{n} \binom{k-1}{m} (-1)^{k-m-1} S_{k,n} = P(A_{m+1})\ . \end{aligned}$$

Inoltre, tenendo presente l'Esercizio 2.2.1 si ha $S_{k,n} = (k!)^{-1}$ e quindi la probabilità che almeno m lettere raggiungano correttamente il destinatario è data da

$$P(A_m) = \sum_{k=m}^{n} \binom{k-1}{m-1} \frac{(-1)^{k-m}}{k!}\ . \qquad \square$$

• **Esercizio 2.2.3.** Dato lo spazio probabilizzato $(\Omega, \mathcal{F}, P)$ e considerati due eventi $E_1, E_2 \in \mathcal{F}$, si determini la probabilità che si verifichino esattamente m dei due eventi con $m = 0, 1, 2$ e almeno m dei due eventi con $m = 1, 2$.

Soluzione. Assumendo le notazioni degli Esercizi 2.2.1 e 2.2.2, si ottiene $S_{0,2} = 1$ e $S_{1,2} = P(E_1) + P(E_2)$, mentre $S_{2,2} = P(E_1 \cap E_2)$. Dunque, sulla base dell'Esercizio 2.2.1, la probabilità degli eventi F_0, F_1 e F_2 è data da

$$P(F_0) = 1 - S_{1,2} + S_{2,2} = 1 - P(E_1 \cup E_2)$$

e

$$P(F_1) = S_{1,2} - 2S_{2,2} = P(E_1 \cup E_2) - P(E_1 \cap E_2)\ ,$$

mentre

$$P(F_2) = S_{2,2} = P(E_1 \cap E_2)\ .$$

Inoltre, sulla base dell'Esercizio 2.2.2, la probabilità degli eventi A_1 e A_2 è data da

$$P(A_1) = S_{1,2} - S_{2,2} = P(E_1 \cup E_2)$$

e

$$P(A_2) = S_{2,2} = P(E_1 \cap E_2)\ . \qquad \square$$

• **Esercizio 2.2.4.** Un dado bilanciato viene lanciato n volte. Si determini la probabilità che almeno una delle facce non si verifichi.

Soluzione. Lo spazio fondamentale Ω è equivalente a quello dato da tutte le possibili configurazioni di n palline distinguibili in 6 celle. Dunque, si ha $\text{card}(\Omega) = 6^n$ e si può scegliere $\mathcal{F} = \mathcal{P}(\Omega)$. Dalle assunzioni fatte le 6^n configurazioni sono ugualmente probabili, ovvero ad ogni evento elementare $\omega_k \in \Omega$ si assegna la probabilità $P(\{\omega_k\}) = 6^{-n}$. Si denoti con E_k l'evento che la k-esima faccia non si verifichi negli n lanci, dove $k = 1, \dots, 6$. Se si considerano indici distinti, risulta $P(E_j) = 5^n 6^{-n}$, $P(E_{j_1} \cap E_{j_2}) = 4^n 6^{-n}$, $\dots$, $P(\bigcap_{k=1}^{5} E_{j_k}) = 6^{-n}$ e $P(\bigcap_{k=1}^{6} E_k) = 0$. In base alla Formula di Inclusione ed Esclusione, la probabilità richiesta è quindi data da

$$P\left(\bigcup_{k=1}^{6} E_k\right) = \binom{6}{1}\left(\frac{5}{6}\right)^n - \binom{6}{2}\left(\frac{4}{6}\right)^n + \ldots + \binom{6}{5}\left(\frac{1}{6}\right)^n$$
$$= \sum_{k=1}^{5} (-1)^{k+1}\binom{6}{k}\left(\frac{6-k}{6}\right)^n,$$

dal momento che vi sono $\binom{6}{k}$ possibili scelte di k facce fra 6 facce. □

Sezione 2.3

• **Esercizio 2.3.1.** (Dimostrazione probabilistica à la Paul Erdös) Una circonferenza è suddivisa in due insiemi che vengono rispettivamente colorati di bianco e di nero. La proporzione di bianco è un quinto della circonferenza. Senza avere informazione sulla struttura dei due insiemi, determinare se è possibile inscrivere nella circonferenza un quadrato in modo tale che tutti i suoi vertici appartengano all'insieme colorato di nero.

Soluzione. Una dimostrazione probabilistica è un metodo di dimostrazione non costruttiva dell'esistenza certa di un oggetto matematico mediante considerazioni probabilistiche. La tecnica è stata introdotta dal matematico Paul Erdös (1913-1996) e viene spesso impiegata nella Teoria dei Grafi. Per quanto riguarda il presente esercizio, senza perdita di generalità si consideri lo spazio fondamentale $\Omega = [0, 2\pi[$, in modo tale che ogni elemento di Ω rappresenta un punto sulla circonferenza, e la relativa σ-algebra $\mathcal{F} = \mathcal{B}([0, 2\pi[)$. Si consideri inoltre la misura di probabilità uniforme su Ω, ovvero si assegni la probabilità P in modo tale che

$$P(B) = \frac{\lambda(B)}{\lambda(\Omega)} = \frac{\lambda(B)}{2\pi}$$

per ogni $B \in \mathcal{F}$, e si assuma che i due insiemi considerati siano probabilizzabili rispetto a questa misura. In questo caso, si selezioni in modo uniforme un evento elementare di Ω in modo tale che il punto ottenuto rappresenta la posizione sulla circonferenza del vertice di riferimento del quadrato. Inoltre, si indichi con E_k l'evento che il vertice k-esimo del quadrato appartenga all'insieme colorato di bianco, per cui si ha $P(E_k) = \frac{1}{5}$ per le assunzioni fatte. Dunque, l'evento che tutti i vertici del quadrato appartengano all'insieme colorato di bianco è dato da $E = \bigcap_{k=1}^{4} E_k$. Dal momento che risulta $E \subset \bigcup_{k=1}^{4} E_k$ e tenendo presente la disuguaglianza di Bonferroni, si ha

$$P(E) \leq P\left(\bigcup_{k=1}^{4} E_k\right) \leq \sum_{k=1}^{4} P(E_k) = \frac{4}{5}\,.$$

Dunque, la probabilità di selezionare un quadrato con tutti i vertici che appartengono all'insieme colorato di nero è almeno pari ad $\frac{1}{5}$. Si deve concludere che è possibile inscrivere nella circonferenza un quadrato del tipo richiesto anche se non se ne può determinare la posizione. □

Sezione 2.4

• **Esercizio 2.4.1.** (Un problema di Lewis Carroll) Un'urna contiene una pallina di cui è noto che può essere bianca o nera. Un'ulteriore pallina bianca viene inserita nell'urna e successivamente una delle due palline viene estratta. Se la pallina estratta è bianca, si determini la probabilità che anche la pallina rimasta nell'urna sia bianca.

Soluzione. Si indichi con E_1 l'evento relativo alla presenza di una pallina bianca nell'urna prima dell'inserimento dell'ulteriore pallina bianca e si assuma che $P(E_1) = p$ dove $p \in [0, 1]$. Ovviamente, la probabilità che la pallina nell'urna sia nera è data da $P(E_1^c) = 1 - p$. Inoltre, sia E_2 l'evento

relativo all'estrazione di una pallina bianca dall'urna. Dunque, risulta $P(E_2 \mid E_1) = 1$ e $P(E_2 \mid E_1^c) = \frac{1}{2}$. Dal momento che gli eventi E_1 e E_1^c costituiscono una partizione dello spazio fondamentale, applicando la Legge delle Probabilità Totali si ottiene

$$P(E_2) = P(E_2 \mid E_1)P(E_1) + P(E_2 \mid E_1^c)P(E_1^c) = \frac{p+1}{2} \, .$$

Considerando la Formula di Bayes si ha la probabilità richiesta, ovvero

$$P(E_1 \mid E_2) = \frac{P(E_2 \mid E_1)P(E_1)}{P(E_2)} = \frac{2p}{p+1} \, .$$

Dunque, in modo controintuitivo per un neofita risulta $P(E_1 \mid E_2) = \frac{2}{3}$ nel caso in cui $p = \frac{1}{2}$. Infine, si osservi che in questo esercizio è stato considerato in effetti uno spazio fondamentale Ω costituito da quattro eventi di una opportuna partizione in modo simile all'Esercizio 1.3.2, ovvero

$$\Omega = \{E_1 \cap E_2, E_1 \cap E_2^c, E_1^c \cap E_2, E_1^c \cap E_2^c\} \, ,$$

con $\mathcal{F} = \mathcal{P}(\Omega)$, mentre l'assegnazione di probabilità è data da $P(E_1 \cap E_2) = p$, $P(E_1 \cap E_2^c) = 0$, $P(E_1^c \cap E_2) = \frac{1-p}{2}$ e $P(E_1^c \cap E_2^c) = \frac{1-p}{2}$. Si noti che la sola conoscenza di $P(E_1)$, $P(E_2 \mid E_1)$ e $P(E_2 \mid E_1^c)$ permette di effettuare in modo coerente l'assegnazione di probabilità. Lewis Carroll, pseudonimo di Charles Lutwidge Dodgson (1832-1898), oltre ad essere un famoso scrittore, autore fra l'altro del celeberrimo *Alice's Adventures in Wonderland*, è stato anche un matematico con molti interessi nella Teoria della Probabilità. □

• **Esercizio 2.4.2.** (Un test clinico) Un test ematico rivela la presenza di una patologia con probabilità pari a p_1, mentre produce falsi positivi con probabilità pari a p_2. Se in una popolazione la probabilità di avere la patologia è pari a p, si determini la probabilità di avere la patologia dato che il test ematico è risultato positivo.

Soluzione. Si indichi con E_1 l'evento relativo alla presenza della patologia, mentre sia E_2 l'evento relativo alla positività al test. Si ha $P(E_1) = p$ e quindi $P(E_1^c) = 1 - p$. Inoltre, $P(E_2 \mid E_1) = p_1$ e $P(E_2 \mid E_1^c) = p_2$. Dal momento che gli eventi E_1 e E_1^c costituiscono una partizione dello spazio fondamentale, applicando la Legge delle Probabilità Totali si ottiene

$$P(E_2) = P(E_2 \mid E_1)P(E_1) + P(E_2 \mid E_1^c)P(E_1^c) = p_1 p + p_2(1-p) \, .$$

Considerando la Formula di Bayes, la probabilità richiesta è data da

$$P(E_1 \mid E_2) = \frac{p_1 p}{p_1 p + p_2(1-p)} = \frac{1}{1 + \left(\frac{1}{p} - 1\right)\frac{p_2}{p_1}} \, .$$

Si osservi che la probabilità $P(E_1 \mid E_2)$ decresce al diminuire di p e all'aumentare del rapporto $\frac{p_2}{p_1}$. Dal momento che si ha $P(E_1^c \mid E_2) = 1 - P(E_1 \mid E_2)$, per questa ultima probabilità valgono considerazioni inverse. Come esempio pratico, si assuma che $p_1 = \frac{99}{100}$ e $p_2 = \frac{1}{100}$, mentre $p = \frac{1}{1000}$. Dal momento che si ha $P(E_1 \mid E_2) = \frac{11}{122} \simeq 0.09$, il test ematico è fortemente inefficiente in questo caso. Dunque, quando si considera una patologia abbastanza rara, si deve avere valori di p_1 molto elevati e valori di p_2 molto bassi. In particolare, se si richiede che $P(E_1 \mid E_2) \geq \alpha$ per $\alpha \in]0, 1[$, deve risultare

$$\frac{p_2}{p_1} \leq \frac{1-\alpha}{\alpha} \, \frac{p}{1-p} \, .$$

Dunque, se si pone $\alpha = \frac{9}{10}$ e $p = \frac{1}{1000}$, si deve avere

$$\frac{p_2}{p_1} \leq \frac{1}{8991} \simeq 0.0001 \, . \qquad \square$$

• **Esercizio 2.4.3.** (Paradosso di Simpson) Due trattamenti vengono assegnati in modo aleatorio ad un gruppo di pazienti che comprendono sia donne che uomini. Se E_1 rappresenta l'evento che il trattamento sia efficace, E_2 rappresenta l'evento che sia stato assegnato il primo trattamento (quindi, E_2^c è l'evento che sia stato assegnato il secondo trattamento) e E_3 rappresenta l'evento che sia stata selezionata una donna (quindi, E_3^c è l'evento che sia stato selezionato un uomo), si verifichi che si può avere

$$P(E_1 \mid E_2 \cap E_3) > P(E_1 \mid E_2^c \cap E_3)$$

e

$$P(E_1 \mid E_2 \cap E_3^c) > P(E_1 \mid E_2^c \cap E_3^c) \, ,$$

anche se

$$P(E_1 \mid E_2) < P(E_1 \mid E_2^c) \, .$$

In altre parole, il primo trattamento può essere più efficace del secondo nei singoli gruppi di donne e uomini, anche se è vera l'affermazione opposta per l'intero gruppo di pazienti.

Soluzione. Si noti che sulla base dell'Esercizio 1.3.2 si può considerare lo spazio fondamentale Ω costituito dagli otto eventi di una opportuna partizione, ovvero

$$\begin{aligned} \Omega = \{ & E_1 \cap E_2 \cap E_3, E_1^c \cap E_2 \cap E_3, E_1 \cap E_2^c \cap E_3, E_1^c \cap E_2^c \cap E_3, \\ & E_1 \cap E_2 \cap E_3^c, E_1^c \cap E_2 \cap E_3^c, E_1 \cap E_2^c \cap E_3^c, E_1^c \cap E_2^c \cap E_3^c \} \end{aligned}$$

con $\mathcal{F} = \mathcal{P}(\Omega)$. Si assuma l'assegnazione di probabilità

$$\begin{aligned} & P(E_1 \cap E_2 \cap E_3) = p_1 \, , P(E_1^c \cap E_2 \cap E_3) = p_2 \, , P(E_1 \cap E_2^c \cap E_3) = p_3 \, , P(E_1^c \cap E_2^c \cap E_3) = p_4 \, , \\ & P(E_1 \cap E_2 \cap E_3^c) = p_5 \, , P(E_1^c \cap E_2 \cap E_3^c) = p_6 \, , P(E_1 \cap E_2^c \cap E_3^c) = p_7 \, , P(E_1^c \cap E_2^c \cap E_3^c) = p_8 \, , \end{aligned}$$

dove $p_k \geq 0$ con $k = 1, \ldots, 8$ e $\sum_{k=1}^{8} p_k = 1$. Dalla precedente assegnazione risulta in particolare

$$\begin{aligned} & P(E_2 \cap E_3) = p_1 + p_2 \, , P(E_2^c \cap E_3) = p_3 + p_4 \, , P(E_2 \cap E_3^c) = p_5 + p_6 \, , P(E_2^c \cap E_3^c) = p_7 + p_8 \, , \\ & P(E_1 \cap E_2) = p_1 + p_5 \, , P(E_1^c \cap E_2) = p_2 + p_6 \, , P(E_1 \cap E_2^c) = p_3 + p_7 \, , P(E_1^c \cap E_2^c) = p_4 + p_8 \, , \end{aligned}$$

da cui si ha anche

$$P(E_2) = p_1 + p_2 + p_5 + p_6 \, , P(E_2^c) = p_3 + p_4 + p_7 + p_8 \, .$$

Quindi, le prime due disuguaglianze considerate nell'esercizio sono equivalenti alle ulteriori disuguaglianze

$$\frac{p_1}{p_1 + p_2} > \frac{p_3}{p_3 + p_4}$$

e

$$\frac{p_5}{p_5 + p_6} > \frac{p_7}{p_7 + p_8} \, ,$$

mentre la terza disuguaglianza è equivalente a

$$\frac{p_1 + p_5}{p_1 + p_2 + p_5 + p_6} < \frac{p_3 + p_7}{p_3 + p_4 + p_7 + p_8} \, .$$

Queste tre disuguaglianze possono essere contemporaneamente soddisfatte. In effetti, non si può in generale sommare ordinatamente i numeratori e i denominatori di due disuguaglianze fra rapporti al fine di ottenere una nuova valida disuguaglianza. Ad esempio, l'assegnazione di probabilità $p_1 = \frac{3}{30}$, $p_2 = \frac{1}{30}$, $p_3 = \frac{8}{30}$, $p_4 = \frac{3}{30}$, $p_5 = \frac{3}{30}$, $p_6 = \frac{8}{30}$, $p_7 = \frac{1}{30}$, $p_8 = \frac{3}{30}$ verifica contemporaneamente le tre disuguaglianze descritte. Il paradosso è stato considerato in dettaglio dallo statistico Edward Hugh Simpson (1922-2019), anche se era già noto al famoso statistico George Udny Yule (1871-1951). □

• **Esercizio 2.4.4.** (Paradosso di Berkson) In una popolazione vengono scelti soggetti in modo aleatorio al fine analizzare la presenza di due patologie. Sia E_1 l'evento che il soggetto selezionato presenti la prima patologia e E_2 l'evento che presenti la seconda patologia. Dunque, $(E_1 \cup E_2)$ rappresenta l'evento che il soggetto sia nel gruppo dei malati con almeno una patologia. Assumendo che gli eventi E_1 ed E_2 siano indipendenti e che risulti $P(E_1), P(E_2) \in]0,1[$, si verifichi che

$$P(E_1 \mid E_2) < P(E_1 \mid E_1 \cup E_2) \, .$$

In altre parole, la probabilità che un soggetto nel gruppo dei malati manifesti la prima patologia è maggiore della probabilità questo soggetto manifesti la prima patologia data la presenza della seconda patologia. Dunque, apparentemente la disuguaglianza sembra suggerire una dipendenza fra le patologie, anche se queste sono eventi indipendenti.

Soluzione. Dal momento che gli eventi E_1 e E_2 sono indipendenti si ha $P(E_1 \mid E_2) = P(E_1)$. Inoltre, dal momento che

$$P(E_1 \cup E_2) \geq \max(P(E_1), P(E_2)) > 0 \, ,$$

risulta

$$P(E_1 \mid E_1 \cup E_2) = \frac{P(E_1)}{P(E_1 \cup E_2)} > P(E_1) \, ,$$

essendo $P(E_1 \cup E_2) \in]0,1[$. L'apparente paradosso deriva dal fatto che le probabilità vengono calcolate rispetto al gruppo dei malati, escludendo in effetti il gruppo dei soggetti sani, ovvero eventi indipendenti possono non essere tali quando si considera eventi condizionati. Il paradosso prende il nome dallo statistico Joseph Berkson (1899-1982), che ne aveva osservato gli effetti in alcuni studi clinici. □

Sezione 2.5

• **Esercizio 2.5.1.** Si consideri lo spazio fondamentale $\Omega = \{1, \ldots, n\}$, dove n è un numero primo, e sia $\mathcal{F} = \mathcal{P}(\Omega)$. Si assuma inoltre l'assegnazione di probabilità

$$P(E) = \frac{\text{card}(E)}{n}$$

per ogni $E \in \mathcal{F}$. Si verifichi che, se E_1 e E_2 sono eventi indipendenti, almeno uno di questi eventi è l'evento impossibile o l'evento certo.

Soluzione. Sia $\text{card}(E_1) = a$ e $\text{card}(E_2) = b$, mentre $\text{card}(E_1 \cap E_2) = c$. Dal momento che $P(E_1 \cap E_2) = P(E_1)P(E_2)$, si ha $cn = ab$. Se $ab = 0$, allora almeno un evento tra E_1 e E_2 è l'evento impossibile. Al contrario, se $ab \neq 0$, allora n deve essere un divisore di ab essendo c un numero intero. Tuttavia, n è un numero primo e quindi n è un divisore di a o di b. Dal momento che $a, b \leq n$, deve sussistere almeno una delle relazioni $a = n$ e $b = n$, ovvero almeno un evento tra E_1 e E_2 è l'evento certo. Se E_1 e E_2 non sono l'evento impossibile o l'evento certo, risulta controintuitivo il fatto che per un numero primo quale $n = 2^{31} - 1$ non esistono eventi indipendenti, mentre per $n = 2^{31}$ esistono numerosi eventi indipendenti. Incidentalmente, si osservi che $2^{31} - 1$ è un numero primo di Mersenne (che prende il nome dal matematico francese Marin Mersenne, 1588-1648). □

• **Esercizio 2.5.2.** Sia dato lo spazio probabilizzato $(\Omega, \mathcal{F}, P)$, dove $\Omega = [0,1]$ e $\mathcal{F} = \mathcal{B}([0,1])$, mentre $P = \lambda$ è la misura di Lebesgue su $[0,1]$. Si consideri inoltre la successione di eventi $(E_n)_{n\geq 0} \in \mathcal{F}$ con

$$E_n = \bigcup_{k=1}^{2^n} [(2k-1)2^{-n-1}, k2^{-n}] \, .$$

Si verifichi che gli eventi della successione sono completamente indipendenti.

Soluzione. L'evento E_n è un'unione di 2^n eventi incompatibili di misura di Lebesgue pari a 2^{-n-1}, ovvero per ogni n si ha

$$P(E_n) = \sum_{k=1}^{2^n} \lambda([(2k-1)2^{-n-1}, k2^{-n}]) = 2^n \times 2^{-n-1} = \frac{1}{2} \, .$$

Se $n_1 < n_2$, l'evento $E_{n_1} \cap E_{n_2}$ è l'unione di $2^{n_1} \times 2^{n_2-n_1-1} = 2^{n_2-1}$ eventi incompatibili di misura di Lebesgue pari a 2^{-n_2-1}, e quindi risulta

$$P(E_{n_1} \cap E_{n_2}) = 2^{n_2-1} \times 2^{-n_2-1} = \frac{1}{4} = P(E_{n_1})P(E_{n_2}) \, .$$

Iterando il procedimento, se $n_1 < n_2 < n_3$, si osservi che l'evento $E_{n_1} \cap E_{n_2} \cap E_{n_3}$ è l'unione di $2^{n_2-1} \times 2^{n_3-n_2-1} = 2^{n_3-2}$ eventi incompatibili di misura di Lebesgue pari a 2^{-n_3-1}, da cui

$$P(E_{n_1} \cap E_{n_2} \cap E_{n_3}) = 2^{n_3-2} \times 2^{-n_3-1} = \frac{1}{8} = P(E_{n_1})P(E_{n_2})P(E_{n_3}) \, .$$

In generale, se $n_1 < \ldots < n_k$ dove $k > 1$, l'evento $\bigcap_{l=1}^{k} E_{n_l}$ è l'unione di 2^{n_k-k+1} eventi incompatibili di misura di Lebesgue pari a 2^{-n_k-1}, da cui

$$P\left(\bigcap_{l=1}^{k} E_{n_l}\right) = 2^{n_k-k+1} \times 2^{-n_k-1} = 2^{-k} = \prod_{l=1}^{k} P(E_{n_l}) \, .$$

Quindi, gli eventi della successione sono completamente indipendenti. □

Sezione 2.6

• **Esercizio 2.6.1.** Si consideri un'urna che contiene N palline numerate da 1 a N. Dall'urna vengono estratte con reimmissione n palline. Si determini la probabilità che il massimo numero estratto nelle n estrazioni sia pari a j, dove $j = 1, \ldots, N$.

Soluzione. L'esperimento aleatorio analizzato è relativo ad uno spazio fondamentale prodotto dato da $\Omega = \Omega_1 \times \cdots \times \Omega_n$, dove $\Omega_k = \{1, \ldots, N\}$ rappresenta lo spazio fondamentale relativo alla k-esima estrazione per $k = 1, \ldots, n$. Dal momento Ω ha cardinalità finita, la σ-algebra prodotto $\mathcal{F}$ può essere scelta come $\mathcal{P}(\Omega)$. Inoltre, dal momento che le estrazioni sono indipendenti, si può considerare la probabilità prodotto $P = P_1 \otimes \cdots \otimes P_n$ con un'assegnazione equiprobabile su ogni spazio fondamentale, ovvero $P_k(\{j\}) = N^{-1}$ per $k = 1, \ldots, n$. Se $E_{k,j}$ rappresenta l'evento che venga estratto un numero minore o uguale a j nella k-esima estrazione, ovvero $E_{k,j} = \{1, \ldots, j\}$, si ha $P_k(E_{k,j}) = \frac{j}{N}$. Dunque, se $E_j = E_{1,j} \times \cdots \times E_{n,j}$, ovvero E_j rappresenta l'evento che venga estratto un numero minore o uguale a j nelle n estrazioni, si ha

$$P(E_j) = \prod_{k=1}^{n} P_k(E_{k,j}) = \left(\frac{j}{N}\right)^n \, .$$

Infine, se E rappresenta l'evento che il massimo numero estratto nelle n estrazioni sia pari a j, dal momento che $E = E_j \setminus E_{j-1}$ con $E_{j-1} \subset E_j$, la probabilità richiesta è data da

$$P(E) = P(E_j) - P(E_{j-1}) = \left(\frac{j}{N}\right)^n - \left(\frac{j-1}{N}\right)^n . \qquad \square$$

• **Esercizio 2.6.2.** (Test simultanei) In un procedimento giudiziario, un campione di DNA viene confrontato con quelli presenti in un archivio relativo a n soggetti. Supponendo che la probabilità di avere una corrispondenza del campione di DNA con quello di un soggetto dell'archivio per effetto della casualità sia pari a p, si determini la probabilità di avere almeno una corrispondenza.

Soluzione. L'esperimento aleatorio analizzato è relativo allo spazio fondamentale prodotto dato da $\Omega = \Omega_1 \times \cdots \times \Omega_n$, dove $\Omega_k = \{0, 1\}$ è lo spazio fondamentale relativo al k-esimo confronto fra campioni di DNA, supponendo che gli eventi elementari $\{0\}$ e $\{1\}$ rappresentino rispettivamente la mancanza e la presenza della corrispondenza. La σ-algebra prodotto $\mathcal{F}$ può essere scelta come $\mathcal{P}(\Omega)$. Inoltre, dal momento che si può assumere che i campioni presenti nell'archivio diano luogo ad eventi indipendenti, si considera la probabilità prodotto $P = P_1 \otimes \cdots \otimes P_n$ con l'assegnazione $P_k(\{1\}) = p$ e $P_k(\{0\}) = 1 - p$ per $k = 1, \ldots, n$. Dunque, la probabilità di non avere nessuna corrispondenza è data da $P(\{(0, \ldots, 0)\}) = (1 - p)^n$ e dunque la probabilità richiesta risulta

$$1 - P(\{(0, \ldots, 0)\}) = 1 - (1 - p)^n .$$

Anche nel caso in cui p è un valore apparentemente trascurabile, si noti che per n elevato la precedente probabilità potrebbe essere prossima all'unità. Ad esempio, per $p = 10^{-5}$ e $n = 10^6$, questa probabilità è circa pari a 1. Dunque, basandosi solamente sulla corrispondenza fra campioni di DNA, per effetto della casualità una persona innocente potrebbe essere sottoposta a giudizio se i risultati probabilistici non vengono correttamente interpretati nel procedimento giudiziario. $\square$

Sezione 2.7

• **Esercizio 2.7.1.** Dato lo spazio probabilizzato $(\Omega, \mathcal{F}, P)$, si consideri una successione di eventi $(E_n)_{n\geq 1}$ tale che $P(E_n) = n^{-2}$ e si determini $P(\limsup_n E_n)$.

Soluzione. Si ha

$$\sum_{n=1}^{\infty} P(E_n) = \sum_{n=1}^{\infty} \frac{1}{n^2} < 1 + \sum_{n=2}^{\infty} \frac{1}{n(n-1)} = 1 + \sum_{n=2}^{\infty} \left(\frac{1}{n-1} - \frac{1}{n}\right) = 2 < \infty ,$$

dal momento che la serie considerata nell'ultimo passaggio è telescopica. Quindi, dalla prima parte del Lemma di Borel-Cantelli si ha $P(\limsup_n E_n) = 0$. $\square$

• **Esercizio 2.7.2.** Si consideri lo schema di Bernoulli relativo ad una infinità numerabile di ripetizioni di un gioco per cui alla n-esima ripetizione si vince 1 unità con probabilità $\frac{2^n}{2^n+1}$ e si perde 2^n unità con probabilità $\frac{1}{2^n+1}$. Evidentemente, il gioco è equo dal momento che la perdita attesa è pari alla vincita attesa. Si commenti riguardo alla convenienza del gioco.

Soluzione. Se E_n rappresenta l'evento di perdere alla n-esima ripetizione del gioco, tenendo presente le proprietà della serie geometrica si ha

$$\sum_{n=1}^{\infty} P(E_n) = \sum_{n=1}^{\infty} \frac{1}{2^n + 1} < \sum_{n=1}^{\infty} \frac{1}{2^n} = 1 < \infty .$$

Dalla prima parte del Lemma di Borel-Cantelli risulta $P(\limsup_n E_n) = 0$, ovvero il giocatore perderà infinitamente spesso con probabilità nulla. Inoltre, dall'Esercizio 1.2.1 si ha

$$P(\liminf_n E_n^c) = 1 - P(\limsup_n E_n) = 1$$

e $\liminf_n E_n^c \subseteq \limsup_n E_n^c$, per cui $P(\limsup_n E_n^c) = 1$, ovvero il giocatore vincerà infinitamente spesso con probabilità unitaria. Quindi, anche se il gioco è equo la quantità di denaro vinta dal giocatore tende a diventare infinitamente elevata. Tuttavia, si noti che il giocatore deve avere una grande disponibilità per sostenere l'eventuale ingente perdita alla n-esima ripetizione del gioco. □

• **Esercizio 2.7.3.** (Valori record) Si consideri i record personali di un atleta che possiede prestazioni costanti nel tempo e in modo tale che i risultati in una gara sono indipendenti dalle restanti gare. Si commenti riguardo alla probabilità che l'atleta migliori il record personale nel tempo.

Soluzione. Si consideri la successione di eventi $(E_n)_{n\geq 1}$ dove E_n rappresenta l'evento che l'atleta ottenga il record personale nella n-esima gara. Evidentemente, alla prima gara si ha $P(E_1) = 1$. Se si considera le prime due gare, dal momento che le prestazioni dell'atleta sono costanti e i due risultati sono indipendenti, il record può essere ottenuto nella prima o nella seconda gara con la medesima probabilità, ovvero $P(E_2) = \frac{1}{2}$. Sulla base di questo ragionamento, in generale si ha dunque che $P(E_n) = n^{-1}$. Quindi, tenendo presente che la serie armonica diverge, si ha

$$\sum_{n=1}^{\infty} P(E_n) = \sum_{n=1}^{\infty} \frac{1}{n} = \infty$$

e dalla seconda parte del Lemma di Borel-Cantelli risulta $P(\limsup_n E_n) = 1$, ovvero l'atleta migliora il record personale infinitamente spesso se riesce a mantenere le proprie prestazioni costanti nel tempo. Si consideri l'ulteriore successione $(E_n)_{n\geq 2}$, dove E_n rappresenta l'evento che l'atleta ottenga il doppio record, ovvero due record personali in due gare successive. Con considerazioni simili a quelle fatte in precedenza, si ha

$$P(E_n) = \frac{1}{n(n-1)} \,,$$

per cui

$$\sum_{n=2}^{\infty} P(E_n) = \sum_{n=2}^{\infty} \frac{1}{n(n-1)} = \sum_{n=2}^{\infty} \left(\frac{1}{n-1} - \frac{1}{n} \right) = 1 < \infty \,.$$

Quindi, dalla prima parte del Lemma di Borel-Cantelli si ha $P(\limsup_n E_n) = 0$, ovvero il doppio record si presenta infinitamente spesso con probabilità nulla, in modo controintuitivo rispetto alla prima parte dell'esercizio. Si noti che in questo caso gli eventi della successione non sono indipendenti, ma questa assunzione non è necessaria nella prima parte del Lemma di Borel-Cantelli. □

Capitolo 3

Variabili e vettori aleatori

3.1. Variabili aleatorie

Quando si analizzano esperimenti o fenomeni aleatori, le quantità di interesse sono comunemente applicazioni dallo spazio fondamentale all'insieme dei numeri reali, che nella Teoria della Probabilità sono dette variabili aleatorie. Formalmente, si ha la seguente definizione.

Definizione 3.1.1. Se $(\Omega, \mathcal{F}, P)$ è uno spazio probabilizzato, un'applicazione $X : \Omega \to \mathbb{R}$ è detta *variabile aleatoria* (v.a.) se l'evento

$$X^{-1}(B) = \{\omega \in \Omega : X(\omega) \in B\}$$

è tale che $X^{-1}(B) \in \mathcal{F}$ per ogni $B \in \mathcal{B}(\mathbb{R})$. □

Quindi, un'applicazione da Ω a $\mathbb{R}$ è una v.a. se l'immagine inversa di ogni $B \in \mathcal{B}(\mathbb{R})$ appartiene alla σ-algebra $\mathcal{F}$. Nel linguaggio della Teoria della Misura una v.a. è detta *funzione misurabile*. Inoltre, si osservi che la definizione di v.a. non coinvolge la scelta della misura di probabilità effettuata su $\mathcal{F}$. In altre parole, una v.a. continua a rimanere tale anche se viene considerata una differente scelta della misura di probabilità su $\mathcal{F}$.

Se si considera lo spazio probabilizzato $(\Omega, \mathcal{F}, P)$, la misura di probabilità indotta dalla v.a. X è data da

$$P_X(B) = P(X^{-1}(B)) .$$

La misura di probabilità P_X è detta *legge* (o *distribuzione*) della v.a. X.

Per quanto riguarda la notazione, si noti che le v.a. sono generalmente indicate con lettere maiuscole (ad esempio X, Y e Z) e le rispettive leggi sono indicizzate di conseguenza. Inoltre, per semplicità di notazione, l'evento $X^{-1}(B)$ viene semplicemente indicato con $\{X \in B\}$ e la probabilità $P(\{X \in B\})$ è comunemente indicata con $P(X \in B)$. Questi leggeri abusi in notazione sono universalmente adottati nei testi di Teoria della Probabilità. Inoltre, è usuale affermare che una v.a. si "distribuisce" con una particolare legge, adottando una denominazione e una simbologia specifica per la legge.

Proposizione 3.1.2. *$(\mathbb{R}, \mathcal{B}(\mathbb{R}), P_X)$ è uno spazio probabilizzato.*

Dimostrazione. Occorre dimostrare che P_X soddisfa i tre assiomi della Definizione 2.1.2. Per ogni $B \in \mathcal{B}(\mathbb{R})$ si ha ovviamente

$$P_X(B) = P(X \in B) \geq 0 .$$

Inoltre, risulta

$$P_X(\mathbb{R}) = P(X \in \mathbb{R}) = 1 .$$

Infine, data una successione di insiemi disgiunti $(B_n)_{n\geq 1} \in \mathcal{B}(\mathbb{R})$, dal momento che $(\{X \in B_n\})_{n\geq 1}$ è una successione di eventi incompatibili, si ha

$$P_X\left(\bigcup_{n=1}^{\infty} B_n\right) = P\left(X \in \bigcup_{n=1}^{\infty} B_n\right) = P\left(\bigcup_{n=1}^{\infty} \{X \in B_n\}\right) = \sum_{n=1}^{\infty} P(X \in B_n) = \sum_{n=1}^{\infty} P_X(B_n)\,,$$

da cui segue la tesi. □

Se l'interesse è dunque focalizzato sulla v.a. X, lo spazio probabilizzato $(\mathbb{R}, \mathcal{B}(\mathbb{R}), P_X)$ diventa centrale e, da un punto di vista pratico, lo spazio probabilizzato $(\Omega, \mathcal{F}, P)$ può essere tralasciato. Tuttavia, si dovrebbe notare che, mentre una v.a. determina univocamente la rispettiva legge, l'affermazione contraria è falsa. In altre parole, v.a. differenti potrebbero avere la medesima legge, perfino se sono definite su spazi probabilizzati differenti. In particolare, se X e Y sono v.a. con leggi date da P_X e P_Y e se $P_X(B) = P_Y(B)$ per ogni $B \in \mathcal{B}(\mathbb{R})$, allora si dice che le due v.a. sono *uguali in legge* e si scrive

$$X \stackrel{\mathcal{L}}{=} Y\,.$$

Si osservi inoltre che, se si ha un'applicazione $X : \Omega \to S$, dove $S \subset \mathbb{R}$ è un insieme numerabile, affinchè X sia una v.a. è sufficiente che l'evento

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

appartenga a $\mathcal{F}$ per ogni $x \in S$. In questo caso, la v.a. è detta *discreta*. Per definire la legge di una v.a. discreta è sufficiente dunque considerare le probabilità $P(X = x)$ per $x \in S$, che definiscono la cosiddetta *legge essenziale* della v.a. X. Inoltre, se $\text{card}(S) = 1$, la v.a. discreta è detta *degenere*. Una precisa classificazione delle v.a. sarà considerata nella Sezione 3.2, così come le v.a. discrete saranno considerate in dettaglio nella Sezione 3.3. Si noti inoltre che nel linguaggio della Teoria della Misura una v.a. discreta che assume un numero finito di valori, ovvero quando si ha $\text{card}(S) = n$, è detta *funzione misurabile semplice*.

• **Esempio 3.1.1.** Si consideri lo spazio probabilizzato $(\Omega, \mathcal{F}, P)$, dove $\Omega = \{\omega_1, \omega_2\}$ con $\mathcal{F} = \mathcal{P}(\Omega)$ e $P(\{\omega_k\}) = \frac{1}{2}$ per $k = 1, 2$. Si consideri inoltre la v.a. discreta X definita su Ω e tale che $X(\omega_1) = 0$ e $X(\omega_2) = 1$. Essendo $S = \{0, 1\}$, la legge essenziale di X risulta $P(X = x) = \frac{1}{2}$ per $x = 0, 1$. Sia data l'ulteriore v.a. discreta Y definita su Ω e tale che $Y(\omega_1) = 1$ e $Y(\omega_2) = 0$. Dunque, anche la legge essenziale di Y è data da $P(Y = y) = \frac{1}{2}$ per $y = 0, 1$. Risulta evidente che le leggi essenziali di X e Y sono identiche e quindi $X \stackrel{\mathcal{L}}{=} Y$, perfino se le v.a. X e Y differiscono per ogni evento elementare di Ω. Dato l'ulteriore spazio probabilizzato $(\Omega_1, \mathcal{F}_1, P_1)$, dove $\Omega_1 = \{\omega_1, \omega_2, \omega_3, \omega_4\}$ con $\mathcal{F}_1 = \mathcal{P}(\Omega_1)$ e $P_1(\{\omega_k\}) = \frac{1}{4}$ per $k = 1, 2, 3, 4$, se Z è la v.a. discreta definita su Ω_1 tale che $Z(\omega_1) = Z(\omega_2) = 0$ e $Z(\omega_3) = Z(\omega_4) = 1$, la legge essenziale della v.a. Z è identica a quella della v.a. X. Quindi $X \stackrel{\mathcal{L}}{=} Z$, anche se la v.a. Z è definita su un differente spazio fondamentale. La precedente legge è una delle più semplici che si possono considerare e si ottiene per una specifica parametrizzazione della legge di Bernoulli, che è stata introdotta da Jakob Bernoulli. La legge di Bernoulli è un caso particolare della legge Binomiale analizzata nella Sezione 6.1. □

Dal momento che la costruzione della σ-algebra di Borel $\mathcal{B}(\mathbb{R})$ può essere fatta a partire dalla classe di insiemi del tipo $]-\infty, x]$, allora X è una v.a. se l'evento $\{X \leq x\} = \{X \in]-\infty, x]\}$ appartiene ad $\mathcal{F}$ per ogni $x \in \mathbb{R}$. In questo caso, a P_X è associata in modo unico la cosiddetta *funzione di ripartizione* (f.r.), che viene indicata con F_X ed è data da

$$F_X(x) = P_X(]-\infty, x]) = P(X \leq x)\,.$$

Dunque, l'uguaglianza in legge equivale all'uguaglianza delle f.r., ovvero se X e Y sono v.a. con f.r. date rispettivamente da F_X e F_Y, allora $X \stackrel{\mathcal{L}}{=} Y$ se $F_X(x) = F_Y(x)$ per ogni $x \in \mathbb{R}$.

• **Esempio 3.1.2.** Si consideri di nuovo la v.a. discreta X introdotta nell'Esempio 3.1.1. La f.r. di X è data da

$$F_X(x) = \begin{cases} 0 & x < 0 \\ \frac{1}{2} & 0 \leq x < 1 \\ 1 & x \geq 1 \end{cases}$$

o, in forma più concisa, da

$$F_X(x) = \frac{1}{2}\,\mathbf{1}_{[0,1[}(x) + \mathbf{1}_{[1,\infty[}(x)\,.$$

Il grafico di F_X è riportato nella Figura 3.1.1. □

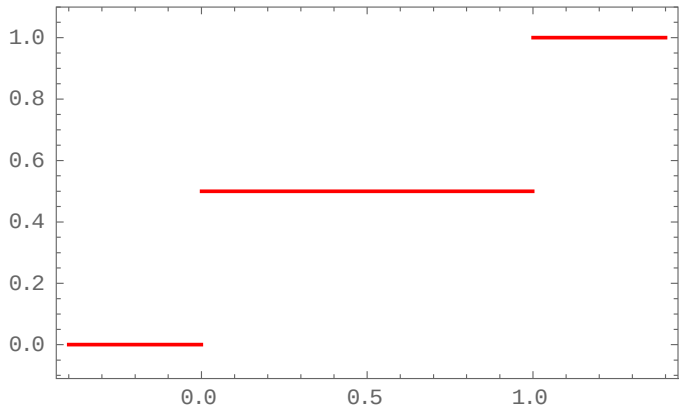


Figura 3.1.1. Funzione di ripartizione relativa alla legge di Bernoulli.

3.2. Proprietà della funzione di ripartizione

Per quanto detto nella Sezione 3.1, risulta evidente che la descrizione probabilistica della v.a. X è incentrata sulle caratteristiche della corrispondente f.r. Sulla base della definizione, è quindi conveniente ricavare alcune proprietà relative alla f.r.

Proposizione 3.2.1. *Se F_X è la f.r. relativa alla v.a. X definita sullo spazio probabilizzato $(\Omega, \mathcal{F}, P)$, si ha $0 \leq F_X(x) \leq 1$ per ogni $x \in \mathbb{R}$.*

Dimostrazione. Per definizione sussiste la relazione $F_X(x) = P(X \leq x)$ con $\{X \leq x\} \in \mathcal{F}$ e quindi risulta $0 \leq P(X \leq x) \leq 1$ per le proprietà della probabilità. □

Proposizione 3.2.2. *Se F_X è la f.r. relativa alla v.a. X definita sullo spazio probabilizzato $(\Omega, \mathcal{F}, P)$, F_X è non decrescente.*

Dimostrazione. Occorre dimostrare che se $x < y$, allora $F_X(x) \leq F_X(y)$. Per definizione si hanno le relazioni $F_X(x) = P(X \leq x)$ e $F_X(y) = P(X \leq y)$. Inoltre, se $\omega \in \{X \leq x\}$ deve risultare anche $\omega \in \{X \leq y\}$. Dunque, si ha $\{X \leq x\} \subset \{X \leq y\}$, e dalla Proposizione 2.2.3 si ottiene infine

$$F_X(x) = P(X \leq x) \leq P(X \leq y) = F_X(y)\,.$$ □

Proposizione 3.2.3. *Se F_X è la f.r. relativa alla v.a. X definita sullo spazio probabilizzato $(\Omega, \mathcal{F}, P)$, allora $\lim_{x\to-\infty} F_X(x) = 0$ e $\lim_{x\to\infty} F_X(x) = 1$.*

Dimostrazione. Si ha

$$\lim_{x\to-\infty} F_X(x) = \lim_{x\to-\infty} P(X \leq x) = \lim_{n} P(E_n)\,,$$

dove $(E_n)_{n\geq 1}$ è una successione decrescente di eventi il cui generico elemento è dato da $E_n = \{X \leq -n\}$. Dunque, tenendo presente il Teorema 2.2.10 e la definizione di limite di successione decrescente di eventi, risulta

$$\lim_n P(E_n) = P(\lim_n E_n) = P\left(\bigcap_{n=1}^{\infty} E_n\right) = P(\emptyset) = 0\,.$$

In modo analogo, si ha

$$\lim_{x\to\infty} F_X(x) = \lim_{x\to\infty} P(X \leq x) = \lim_n P(E_n)\,,$$

dove $(E_n)_{n\geq 1}$ è una successione crescente di eventi il cui generico elemento è $E_n = \{X \leq n\}$. Dunque, risulta

$$\lim_{x\to\infty} F_X(x) = P\left(\bigcup_{n=1}^{\infty} E_n\right) = P(\Omega) = 1\,. \qquad \square$$

Proposizione 3.2.4. *Se F_X è la f.r. di una v.a. X definita sullo spazio probabilizzato $(\Omega, \mathcal{F}, P)$, allora F_X è continua a destra.*

Dimostrazione. Per ogni $x \in \mathbb{R}$, si deve dimostrare che per $\varepsilon > 0$ si ha

$$\lim_{\varepsilon\to 0^+} F_X(x+\varepsilon) = F_X(x)\,,$$

ovvero che

$$\lim_{\varepsilon\to 0^+} P(X \leq x+\varepsilon) = P(X \leq x)\,.$$

Per un dato $x \in \mathbb{R}$, si consideri che

$$\lim_{\varepsilon\to 0^+} P(X \leq x+\varepsilon) = \lim_n P(E_n)\,,$$

dove $(E_n)_{n\geq 1}$ è una successione decrescente di eventi il cui generico elemento è dato da $E_n = \{X \leq x + n^{-1}\}$. Tenendo presente il Teorema 2.2.10 e la definizione di limite di successione decrescente di eventi, si ha

$$\lim_n P(E_n) = P\left(\bigcap_{n=1}^{\infty} E_n\right) = P(X \leq x)\,. \qquad \square$$

Proposizione 3.2.5. *Sia F_X la f.r. di una v.a. X definita sullo spazio probabilizzato $(\Omega, \mathcal{F}, P)$ e per $\varepsilon > 0$ sia*

$$\triangle F_X(x) = \lim_{\varepsilon\to 0^+} F_X(x+\varepsilon) - \lim_{\varepsilon\to 0^+} F_X(x-\varepsilon) = F_X(x) - \lim_{\varepsilon\to 0^+} F_X(x-\varepsilon)\,,$$

ovvero $\triangle F_X(x)$ rappresenta l'ampiezza del salto di $F_X(x)$ nel punto x. Si ha

$$\triangle F_X(x) = P(X = x)\,.$$

Dimostrazione. Risulta

$$\lim_{\varepsilon\to 0^+} F_X(x-\varepsilon) = \lim_{\varepsilon\to 0^+} P(X \leq x-\varepsilon) = \lim_n P(E_n)\,,$$

dove $(E_n)_{n\geq 1}$ è una successione crescente di eventi il cui generico elemento è dato da $E_n = \{X < x - n^{-1}\}$. Inoltre, dal Teorema 2.2.10 e dalla definizione di limite di successione crescente di eventi, risulta

$$\lim_n P(E_n) = P\left(\bigcup_{n=1}^{\infty} E_n\right) = P(X < x)\,,$$

da cui

$$\triangle F_X(x) = P(X \leq x) - P(X < x) = P(X = x)\,. \qquad \square$$

Proposizione 3.2.6. *Sia F_X la f.r. di una v.a. X definita sullo spazio probabilizzato $(\Omega, \mathcal{F}, P)$. L'insieme di punti di discontinuità di F_X è al più numerabile.*

Dimostrazione. Per ogni punto di salto x si consideri l'intervallo aperto

$$I_x = \,]\lim_{\varepsilon \to 0^+} F_X(x-\varepsilon), F_X(x)[\,,$$

dove $\varepsilon > 0$. Se y è un ulteriore punto di salto tale che $x < y$, allora dalla Proposizione 3.2.2 si ha

$$F_X(x) \leq \lim_{\varepsilon \to 0^+} F_X(y-\varepsilon)\,,$$

per cui gli intervalli I_x e I_y sono disgiunti. Quindi, l'insieme dei punti di salto può essere messo in corrispondenza biunivoca con un insieme di intervalli disgiunti, che per la Proposizione 3.2.1 sono sottoinsiemi dell'intervallo $[0, 1]$. Quest'insieme è necessariamente numerabile, dal momento che ogni intervallo dell'insieme contiene almeno un numero razionale e i numeri razionali sono densi in ogni intervallo di $\mathbb{R}$. Dunque, l'insieme dei punti di salto è in corrispondenza biunivoca con un sottoinsieme dei numeri razionali e quindi è finito o numerabile. $\square$

Si può quindi concludere che la f.r. F_X di una v.a. X non è in generale continua a sinistra e che risulta continua in x se e solo se $P(X = x) = 0$, mentre l'insieme dei punti di discontinuità, ovvero l'insieme dei punti di probabilità non nulla, deve essere finito o al più numerabile.

Ogni f.r. F_X può essere espressa in modo unico come la combinazione convessa di tre tipi fondamentali di f.r., ovvero

$$F_X = \alpha_d F_d + \alpha_{ac} F_{ac} + \alpha_s F_s\,,$$

dove $\alpha_d, \alpha_{ac}, \alpha_s \geq 0$ e $\alpha_d + \alpha_{ac} + \alpha_s = 1$ (si veda Chung, 2001, p.1). In questo caso, F_d è una f.r. costante a tratti con un insieme numerabile di punti di salto ed è detta f.r. *discreta*. Formalmente, F_d è una f.r. discreta se esiste un insieme numerabile S, tale che $P(X = x) > 0$ solo se $x \in S$, per cui si ha

$$F_d(x) = \sum_{u \in S} \mathbf{1}_{]-\infty, x]}(u) P(X = u)\,.$$

Inoltre, F_{ac} è una f.r. *assolutamente continua*, ovvero esiste una classe di funzioni non negative, il cui generico elemento è indicato con f_X, che coincidono $q.o.$ rispetto alla misura di Lebesgue, integrabili su $\mathbb{R}$ rispetto alla misura di Lebesgue, e per cui si ha

$$F_{ac}(x) = \int_{-\infty}^{x} f_X(u)\, du\,.$$

Infine, F_s è una f.r. *singolare*, se non è identicamente nulla e se la derivata di F_s esiste ed è nulla $q.o.$ Una f.r. singolare è continua, ma non è assolutamente continua e non ammette la precedente rappresentazione integrale.

Risulta immediato verificare che una v.a. *discreta*, come definita nella Sezione 3.1, possiede una f.r. del primo tipo. Inoltre, una v.a. è detta *assolutamente continua* se possiede una f.r. del secondo tipo. Infine, una v.a. è detta *singolare* se possiede una f.r. del terzo tipo. Una v.a. è detta *mista* se possiede una f.r. che è data da una combinazione convessa di almeno due tipi di f.r. Proprietà ed esempi specifici di v.a. discrete e assolutamente continue saranno discusse a lungo nel presente capitolo e nei prossimi capitoli. Al contrario, v.a. singolari sono poco impiegate nella pratica e non saranno ulteriormente considerate. Tuttavia, si noti che questo tipo di v.a. riveste un'importanza notevole nelle strategie di gioco ottimali in giochi d'azzardo quale la roulette. Questi argomenti sono considerati in grande dettaglio da Billingsley (1995, p.101).

• **Esempio 3.2.1.** Si consideri la v.a. X con f.r. data da

$$F_X(x) = \begin{cases} 0 & x < 0 \\ x & 0 \leq x < 1 \\ 1 & x \geq 1 \end{cases}$$

o, in forma più concisa, da

$$F_X(x) = x\mathbf{1}_{[0,1[}(x) + \mathbf{1}_{[1,\infty[}(x)\,.$$

Dal momento che

$$F_X(x) = \int_{-\infty}^{x} \mathbf{1}_{]0,1[}(u)\,du\,,$$

allora X è una v.a. assolutamente continua. Evidentemente, in questo caso si è scelto $f_X(x) = \mathbf{1}_{]0,1[}(x)$, anche se sarebbe stata ugualmente corretta una scelta del tipo $f_X(x) = \mathbf{1}_{]0,1[\setminus\mathbb{Q}}(x)$. In effetti, le due scelte coincidono $q.o.$, dal momento che l'insieme dei numeri razionali in $]0,1[$ è numerabile e la sua misura di Lebesgue risulta nulla. La legge associata a questa v.a. si ottiene per una particolare parametrizzazione della cosiddetta legge Uniforme, che a sua volta è un caso particolare della legge Beta, che sarà considerata in dettaglio nella Sezione 6.9. Il grafico della f.r. considerata è riportato nella Figura 3.2.1. □

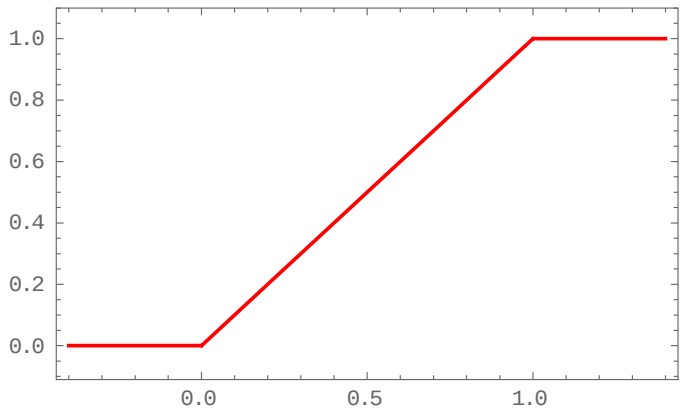


Figura 3.2.1. Funzione di ripartizione relativa alla legge Uniforme.

• **Esempio 3.2.2.** Si consideri la v.a. X con f.r. data da

$$F_X(x) = \begin{cases} 0 & x < 0 \\ \frac{2x+1}{4} & 0 \leq x < 1 \\ 1 & x \geq 1 \end{cases}$$

o, in forma più concisa, da

$$F_X(x) = \frac{2x+1}{4}\,\mathbf{1}_{[0,1[}(x) + \mathbf{1}_{[1,\infty[}(x)\,.$$

Si osservi che

$$F_X(x) = \frac{1}{2}\,F_d(x) + \frac{1}{2}\,F_{ac}(x)\,,$$

dove F_d è la f.r. discreta considerata nell'Esempio 3.1.2, mentre F_{ac} è la f.r. assolutamente continua considerata nell'Esempio 3.2.1. Dunque, X è una v.a. mista. Il grafico della f.r. considerata è riportato nella Figura 3.2.2. □

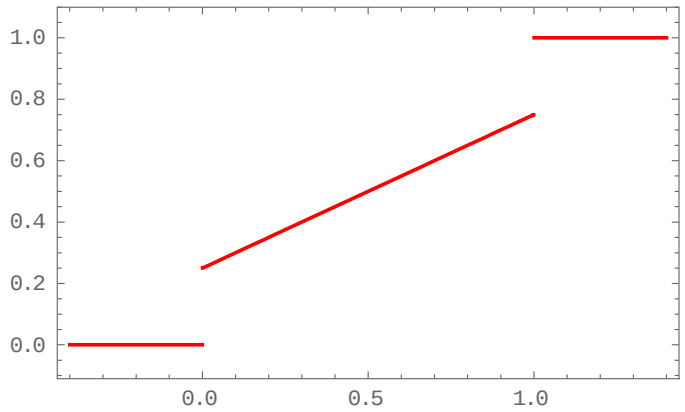


Figura 3.2.2. Funzione di ripartizione relativa alla legge mista.

• **Esempio 3.2.3.** Si consideri l'espansione ternaria di $x \in]0,1[$, ovvero

$$x = \sum_{n=1}^{\infty} \frac{c_n}{3^n}\,,$$

dove $c_n = 0, 1, 2$. Se x appartiene all'insieme di Cantor analizzato nell'Esempio 2.3.5 risulta $c_n = 0, 2$, ovvero x ha un'espansione ternaria che non contiene la cifra 1. Sia inoltre $N = \min\{n \in \mathbb{N} : c_n = 1\}$, mentre sia $N = \infty$ se non esiste tale n, e si ponga

$$b_n = \begin{cases} \frac{c_n}{2} & n < N \\ 1 & n = N \end{cases}.$$

Si consideri la f.r. data da

$$F_X(x) = \sum_{n=1}^{\infty} \frac{b_n}{2^n} + \mathbf{1}_{[1,\infty[}(x)\,,$$

dove i valori dei c_n (e quindi dei b_n) sono quelli relativi all'espansione ternaria di x. Questa f.r. è in effetti la funzione di Cantor, che in modo piuttosto pittoresco viene detta anche "scala del diavolo". Si può dimostrare che la derivata di F_X è nulla eccetto che nei punti che appartengono all'insieme di Cantor. Dal momento che l'insieme di Cantor ha misura di Lebesgue nulla (vedi Esempio 2.3.5), la derivata di F_X è nulla *q.o.* e quindi F_X è singolare (si veda Chung, 2001, p.13). Il grafico (ovviamente approssimato) della f.r. è riportato nella Figura 3.2.3. □

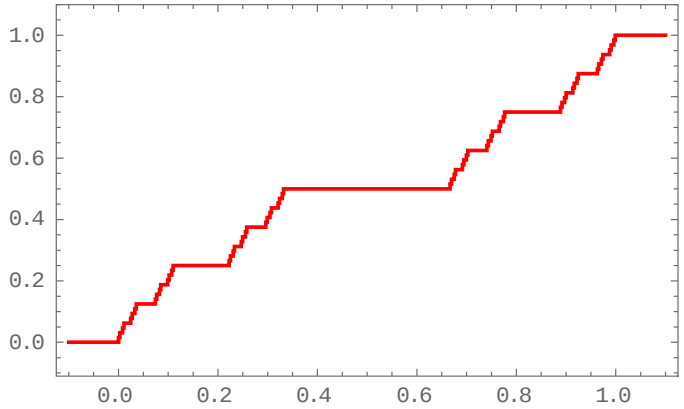


Figura 3.2.3. Funzione di ripartizione di Cantor.

3.3. Variabili aleatorie discrete

Come evidenziato nella Sezione 3.1, la v.a. X è detta discreta se è a valori su insieme numerabile S. Se si definisce la *funzione di probabilità* (f.p.) della v.a. X come

$$p_X(x) = P(X = x) \,,$$

in base a quanto visto nella Sezione 3.2, la f.r. di X può essere scritta come

$$F_X(x) = \sum_{u \in S} \mathbf{1}_{]-\infty,x]}(u) p_X(u) \,.$$

La f.p. p_X è non nulla solo se $x \in S$. L'insieme delle probabilità $(p_X(x))_{x \in S}$ costituisce la legge essenziale di X definita nella Sezione 3.1. Inoltre, la f.r. F_X è costante a tratti, con punti di salto sugli elementi $x \in S$ per cui si ha $p_X(x) = \triangle F_X(x)$. Per ovvie ragioni di semplicità è in pratica conveniente lavorare con la f.p. piuttosto che con la f.r., dal momento che la f.p. definisce comunque la legge essenziale della v.a. X.

Si noti che la legge della v.a. X può essere anche scritta sotto forma di integrale, ovvero per ogni $B \in \mathcal{B}(\mathbb{R})$ si ha

$$P_X(B) = \int_B p_X \, d\nu \,,$$

dove $\nu(B) = \text{card}(B \cap S)$, ovvero ν è la misura che enumera gli elementi di un sottoinsieme di S. Equivalentemente, risulta dunque

$$P(X \in B) = \sum_{x \in S} \mathbf{1}_B(x) p_X(x) \,.$$

Dal momento che $\Omega = \bigcup_{x \in S} \{X = x\}$, è immediato verificare che

$$\sum_{x \in S} p_X(x) = 1 \,.$$

• **Esempio 3.3.1.** Si consideri la v.a. X con f.p. data da

$$p_X(x) = 2^{-x-1} \, \mathbf{1}_{\mathbb{N}}(x) \,.$$

In questo caso si ha $S = \mathbb{N}$, mentre la legge essenziale è data dall'insieme delle probabilità $(2^{-x-1})_{x \geq 0}$. Inoltre, considerando la serie geometrica, per $a \in]0, 1[$ è noto che

$$\sum_{n=0}^{\infty} a^n = \frac{1}{1-a}$$

e quindi, adoperando la precedente relazione con $a = \frac{1}{2}$, si ha

$$\sum_{x=0}^{\infty} p_X(x) = \frac{1}{2} \sum_{x=0}^{\infty} 2^{-x} = 1 \,.$$

Tenendo presente che per $a \in]0,1[$ si ha inoltre

$$\sum_{k=0}^{n} a^k = \frac{1-a^{n+1}}{1-a}$$

e, adoperando questa relazione con $a = \frac{1}{2}$, la f.r. di X risulta

$$F_X(x) = \sum_{u \in S} \mathbf{1}_{]-\infty,x]}(u)\, 2^{-u-1} = \mathbf{1}_{[0,\infty[}(x) \sum_{u=0}^{\lfloor x \rfloor} 2^{-u-1} = (1 - 2^{-\lfloor x \rfloor - 1})\, \mathbf{1}_{[0,\infty[}(x) \,,$$

dove $\lfloor x \rfloor$ denota il più grande numero intero minore o uguale a x. Infine, la probabilità che la v.a. X assuma un valore intero dispari è data da

$$\sum_{x=0}^{\infty} p_X(2x+1) = \sum_{x=0}^{\infty} 2^{-2x-2} = \frac{1}{4} \sum_{x=0}^{\infty} 4^{-x} = \frac{1}{3} \,.$$

La legge associata a questa v.a. si ottiene per una particolare parametrizzazione della legge Geometrica, che a sua volta è un caso particolare della legge Binomiale Negativa considerata nella Sezione 6.3. □

• **Esempio 3.3.2.** Sia $S = (x_n)_{n \geq 1}$ una enumerazione dei numeri razionali e si consideri la v.a. X con legge essenziale data dall'insieme delle probabilità $(2^{-n})_{n \geq 1}$, mentre la relativa f.r. risulta

$$F_X(x) = \sum_{n=1}^{\infty} \mathbf{1}_{]-\infty,x]}(x_n)\, 2^{-n} \,.$$

Si noti che le probabilità relative alla legge essenziale sono in effetti identiche a quelle dell'Esempio 3.3.1, anche se definite su un differente insieme S. Tuttavia, la v.a. X del presente esempio è in qualche modo patologica, in quanto prende valori su una enumerazione dei numeri razionali che non è unica e neppure "cronologica". Inoltre, S è un insieme denso su $\mathbb{R}$. □

3.4. Variabili aleatorie assolutamente continue

Come detto nella Sezione 3.2., una v.a. X è assolutamente continua se possiede una f.r. assolutamente continua rispetto alla misura di Lebesgue. In questo caso, la f.r. della v.a. X può essere espressa come

$$F_X(x) = \int_{-\infty}^{x} f_X(u)\, du \,,$$

dove la funzione misurabile non negativa $f_X : \mathbb{R} \to [0, \infty[$ è detta *densità di probabilità* (d.p.) o semplicemente *densità*. Come evidenziato nella Sezione 3.2, la d.p. non è unica e quindi si dovrebbe parlare più propriamente di una versione della densità. Tenendo presente la mancanza di univocità, per

l'utilizzo pratico è comunque conveniente considerare la d.p. in quanto mostra generalmente un'espressione molto più semplice della relativa f.r. Per comodità si sceglie di solito la versione continua di f_X o, in alternativa, quella con il minor numero di punti di discontinuità. In questo caso, si dice che X ammette una d.p. f_X. Inoltre, dal momento che una f.r. è differenziabile $q.o.$ su $\mathbb{R}$ (si veda Chung, 2001, p.11), allora risulta

$$f_X(x) = \frac{d}{dx} F_X(x) \, .$$

Tenendo presente il Teorema di Radon-Nikodym (Teorema A.9), la legge della v.a. X può essere scritta sotto forma di integrale, ovvero per ogni $B \in \mathcal{B}(\mathbb{R})$ si ha

$$P_X(B) = P(X \in B) = \int_B f_X(x) \, dx \, .$$

In particolare, se $B =]a, b]$ allora risulta

$$P(X \in]a, b]) = F_X(b) - F_X(a) = \int_a^b f_X(x) \, dx \, .$$

Dal momento che $P(X = x) = 0$ per ogni $x \in \mathbb{R}$, essendo la v.a. X assolutamente continua, allora si ha anche

$$P(X \in [a, b]) = P(X \in]a, b[) = P(X \in [a, b[) = \int_a^b f_X(x) \, dx \, .$$

Risulta infine

$$P(X \in \mathbb{R}) = \int_{-\infty}^{\infty} f_X(x) \, dx = 1 \, .$$

• **Esempio 3.4.1.** Si consideri la v.a. X con f.r. data da

$$F_X(x) = x^3 \, \mathbf{1}_{[0,1[}(x) + \mathbf{1}_{[1,\infty[}(x) \, .$$

La v.a. X è assolutamente continua e si ha

$$F_X(x) = \int_{-\infty}^{x} 3u^2 \, \mathbf{1}_{]0,1[}(u) \, du \, .$$

Inoltre, F_X è derivabile $q.o.$, dal momento che non è derivabile solo se $x = 1$, per cui si ha $P(X = 1) = 0$. Dunque, una versione della d.p. è data da

$$f_X(x) = \frac{d}{dx} F_X(x) = 3x^2 \, \mathbf{1}_{]0,1[}(x) \, .$$

Inoltre, si ha

$$P(X \in]0, \tfrac{1}{2}[) = P(X \in [0, \tfrac{1}{2}]) = F_X\left(\frac{1}{2}\right) - F_X(0) = \frac{1}{8} \, ,$$

mentre

$$P(X \in]-1, 2]) = F_X(2) - F_X(-1) = 1 \, .$$

La legge associata alla v.a. considerata si ottiene per una particolare parametrizzazione della legge Beta (si veda la Sezione 6.9). □

3.5. Vettori aleatori

Quando si considera un'applicazione dallo spazio fondamentale a $\mathbb{R}^n$, il concetto di v.a. può essere esteso immediatamente a quello di vettore aleatorio. Formalmente, si ha la seguente definizione.

Definizione 3.5.1. Se $(\Omega, \mathcal{F}, P)$ è uno spazio probabilizzato, un'applicazione $X : \Omega \to \mathbb{R}^n$ è detta *vettore di variabili aleatorie* (v.v.a.) (o semplicemente *vettore aleatorio*) se l'evento

$$X^{-1}(B) = \{\omega \in \Omega : X(\omega) \in B\}$$

è tale che $X^{-1}(B) \in \mathcal{F}$ per ogni $B \in \mathcal{B}(\mathbb{R}^n)$. □

Un v.v.a. è quindi dato da un vettore del tipo $X = (X_1, \ldots, X_n)^{\mathrm{T}}$, dove $X_1, \ldots, X_n$ sono v.a. La v.a. X_k è detta *componente marginale* del v.v.a. X, mentre la misura di probabilità P_X indotta dal v.v.a. è detta *legge congiunta*. Se $B = B_1 \times \cdots \times B_n$ con $B_k \in \mathcal{B}(\mathbb{R})$, si definisca inoltre l'evento

$$\{X \in B\} = \{X_1 \in B_1, \ldots, X_n \in B_n\} = \bigcap_{k=1}^{n} \{X_k \in B_k\} .$$

In particolare, si definisce

$$\{X \leq x\} = \{X_1 \leq x_1, \ldots, X_n \leq x_n\} = \bigcap_{k=1}^{n} \{X_k \leq x_k\} ,$$

dove $x = (x_1, \ldots, x_n)^{\mathrm{T}}$. Tenendo presente che la σ-algebra di Borel su $\mathbb{R}^n$ può essere costruita a partire dagli insiemi rettangolari del tipo $]-\infty, x_1] \times \cdots \times]-\infty, x_n]$, alla legge congiunta P_X è associata in modo unico la cosiddetta *funzione di ripartizione congiunta* (f.r.c.), data da

$$F_X(x) = F_X(x_1, \ldots, x_n) = P(X \leq x) = P(X_1 \leq x_1, \ldots, X_n \leq x_n) .$$

Inoltre, la *funzione di ripartizione marginale* (f.r.m.) della k-esima componente marginale X_k è data da

$$F_{X_k}(x_k) = P(X_k \leq x_k) = P(X_1 \in \mathbb{R}, \ldots, X_k \leq x_k, \ldots, X_n \in \mathbb{R}) ,$$

per $k = 1, \ldots, n$.

Si può verificare facilmente che la f.r.c. di un v.v.a. ha proprietà analoghe a quelle della f.r. di una v.a. In effetti, si ha $0 \leq F_X(x) \leq 1$ per ogni $x \in \mathbb{R}^n$, mentre $F_X(x) \leq F_X(y)$ se $x = (x_1, \ldots, x_n)^{\mathrm{T}}$ e $y = (y_1, \ldots, y_n)^{\mathrm{T}}$ sono tali che $x_k \leq y_k$ per ogni $k = 1, \ldots, n$. Inoltre, risulta

$$\lim_{x_k \to -\infty} F_X(x_1, \ldots, x_n) = 0$$

per ogni $k = 1, \ldots, n$, mentre

$$\lim_{x_1 \to \infty, \ldots, x_n \to \infty} F_X(x_1, \ldots, x_n) = 1 .$$

Infine, per $\varepsilon > 0$ si ha

$$\lim_{\varepsilon \to 0^+} F_X(x_1, \ldots, x_k + \varepsilon, \ldots, x_n) = F_X(x) ,$$

da cui si può dimostrare che F_X è continua nel punto x se e solo se $P(X = x) = 0$.

In modo simile a quanto fatto per le v.a., si può ottenere una classificazione dei v.v.a. a secondo delle varie tipologie delle f.r.c. Tuttavia, per brevità saranno considerate nel seguito solo le due più importanti classi di v.v.a. Per quanto riguarda la prima classe, un v.v.a. è detto *discreto* se è a valori su insieme $S \in \mathbb{R}^n$ finito o numerabile. La *funzione di probabilità congiunta* (f.p.c.) del v.v.a. X è

definita come

$$p_X(x) = p_X(x_1, \dots, x_n) = P(X = x) = P(X_1 = x_1, \dots, X_n = x_n) \, .$$

La f.p.c. $p_X(x)$ è non nulla solo se $x \in S$ e l'insieme delle probabilità $(p_X(x))_{x \in S}$ costituisce la legge essenziale di X. Per semplicità di notazione e di calcolo, risulta generalmente conveniente lavorare con la legge essenziale di un v.v.a. discreto piuttosto che con la relativa f.r.c. In ogni caso, in modo simile a quanto visto nella Sezione 3.3, si può ottenere la legge del v.v.a. X, ovvero per ogni $B \in \mathcal{B}(\mathbb{R}^n)$ si ha

$$P(X \in B) = \int_B p_X \, d\nu = \sum_{(x_1,\dots,x_n) \in S} \mathbf{1}_B(x_1, \dots, x_n) p_X(x_1, \dots, x_n) \, ,$$

dove $\nu(B) = \text{card}(B \cap S)$ e quindi l'espressione della f.r.c. è data da

$$F_X(x) = \sum_{(u_1,\dots,u_n) \in S} \mathbf{1}_{]-\infty, x_1] \times \cdots \times]-\infty, x_n]}(u_1, \dots, u_n) p_X(u_1, \dots, u_n) \, .$$

Se il v.v.a. è discreto, tale è anche ogni componente marginale X_k. In questo caso, la *funzione di probabilità marginale* (f.p.m.) della k-esima componente marginale X_k è data da

$$p_{X_k}(x_k) = P(X_1 \in \mathbb{R}, \dots, X_k = x_k, \dots, X_n \in \mathbb{R}) = \sum_{j \neq k = 1}^{n} \sum_{x_j \in S_j} p_X(x_1, \dots, x_n) \, ,$$

dove S_k rappresenta la proiezione di S sul k-esimo asse cartesiano. L'insieme delle probabilità $(p_{X_k}(x_k))_{x_k \in S_k}$ costituisce la legge essenziale della v.a. X_k.

Un v.v.a. X è assolutamente continuo se possiede una f.r.c. assolutamente continua rispetto alla misura di Lebesgue in $\mathbb{R}^n$. In questo caso, per il Teorema di Radon-Nikodym (Teorema A.9), la legge della v.a. X può essere scritta sotto forma di integrale, ovvero per ogni $B \in \mathcal{B}(\mathbb{R}^n)$ si ha

$$P_X(B) = \int_B f_X(x_1, \dots, x_n) \, dx_1 \dots dx_n$$

e quindi la f.r.c. del v.v.a. X può essere espressa come

$$F_X(x) = \int_{-\infty}^{x_1} \dots \int_{-\infty}^{x_n} f_X(u_1, \dots, u_n) \, du_1 \dots du_n \, ,$$

dove la funzione misurabile non negativa $f_X : \mathbb{R}^n \to [0, \infty[$ è detta *densità di probabilità congiunta* (d.p.c.) o semplicemente *densità congiunta* del v.v.a. X. In modo simile a quanto visto nella Sezione 3.3., la d.p.c. non è unica e quindi si dovrebbe parlare più propriamente di una versione della densità congiunta. Inoltre, dal momento che la f.r.c. è differenziabile $q.o.$ su $\mathbb{R}^n$, allora risulta

$$f_X(x) = f_X(x_1, \dots, x_n) = \frac{\partial^n}{\partial x_1 \dots \partial x_n} F_X(x_1, \dots, x_n) \, .$$

Di nuovo, si deve notare che è conveniente lavorare con la d.p.c. di un v.v.a. assolutamente continuo piuttosto che con la relativa f.r.c. Se il v.v.a. è assolutamente continuo, tale è anche ogni componente marginale X_k. In questo caso, la k-esima componente marginale X_k ammette *densità di probabilità marginale* (d.p.m.) data da

$$f_{X_k}(x_k) = \int_{\mathbb{R}^{n-1}} f_X(x_1, \dots, x_n) \, dx_1 \dots dx_{k-1} dx_{k+1} \dots dx_n \, .$$

• **Esempio 3.5.1.** Si consideri il v.v.a. discreto $X = (X_1, X_2)^{\mathrm{T}}$ con f.p.c.

$$p_X(x_1,x_2)=\left(\frac{1}{4}\right)^{x_1}\left(\frac{1}{2}\right)^{x_2}\left(\frac{1}{4}\right)^{1-x_1-x_2}\mathbf{1}_{\{0,1\}}(x_1)\mathbf{1}_{\{0,1\}}(x_2)\mathbf{1}_{\{0,1\}}(x_1+x_2)\,.$$

Evidentemente, risulta $S=\{(0,0),(1,0),(0,1)\}$, mentre $S_1=S_2=\{0,1\}$. Dunque, la f.p.m. relativa alla prima componente marginale X_1 è data da

$$p_{X_1}(x_1)=\sum_{x_2\in S_2}p_X(x_1,x_2)=\left(\frac{1}{4}\right)^{x_1}\left(\frac{3}{4}\right)^{1-x_1}\mathbf{1}_{\{0,1\}}(x_1)\,,$$

mentre la f.p.m. relativa alla seconda componente marginale X_2 risulta

$$p_{X_2}(x_2)=\sum_{x_1\in S_1}p_X(x_1,x_2)=\frac{1}{2}\,\mathbf{1}_{\{0,1\}}(x_2)\,.$$

La legge associata al v.v.a. X è un caso particolare della legge Multinomiale, mentre le leggi associate alle v.a. X_1 e X_2 si hanno per particolari parametrizzazioni della legge Binomiale (si veda la Sezione 6.5). □

• **Esempio 3.5.2.** Si consideri il v.v.a. assolutamente continuo $X=(X_1,X_2)^{\mathrm{T}}$ che ammette la seguente d.p.c.

$$f_X(x_1,x_2)=24x_1(1-x_1-x_2)\,\mathbf{1}_{]0,1[}(x_1)\mathbf{1}_{]0,1[}(x_2)\mathbf{1}_{]0,1[}(x_1+x_2)\,.$$

Le d.p.m. della prima e della seconda componente marginale sono rispettivamente date da

$$f_{X_1}(x_1)=\int_0^{1-x_1}24x_1(1-x_1-x_2)\,\mathbf{1}_{]0,1[}(x_1)\,dx_2=12x_1(1-x_1)^2\,\mathbf{1}_{]0,1[}(x_1)$$

e

$$f_{X_2}(x_2)=\int_0^{1-x_2}24x_1(1-x_1-x_2)\,\mathbf{1}_{]0,1[}(x_2)\,dx_1=4(1-x_2)^3\,\mathbf{1}_{]0,1[}(x_2)\,.$$

La legge associata al v.v.a. X è legata alla legge di Dirichlet, mentre le leggi associate alle v.a. X_1 e X_2 si hanno per particolari parametrizzazioni della legge Beta (si veda la Sezione 6.12). □

• **Esempio 3.5.3.** Si consideri il v.v.a. $X=(X_1,X_2)^{\mathrm{T}}$ che possiede f.r.c.

$$F_X(x_1,x_2)=\min(x_1,x_2)\,\mathbf{1}_{[0,1]}(\min(x_1,x_2))+\mathbf{1}_{]1,\infty[}(\min(x_1,x_2))\,.$$

Il v.v.a. X non è ovviamente discreto e non è neppure assolutamente continuo in quanto la f.r.c. non è assolutamente continua rispetto alla misura di Lebesgue in $\mathbb{R}^2$. In effetti, si ha che $\frac{\partial^2 F_X}{\partial x_1\partial x_2}$ è nulla $q.o.$ Inoltre, si ha

$$F_{X_1}(x_1)=P(X_1\le x_1,X_2\in\mathbb{R})=x_1\,\mathbf{1}_{[0,1]}(x_1)+\mathbf{1}_{]1,\infty[}(x_1)\,,$$

da cui

$$f_{X_1}(x_1)=\mathbf{1}_{[0,1]}(x_1)\,,$$

ovvero la prima componente marginale è una v.a. assolutamente continua. In modo simile si verifica che anche la seconda componente marginale è assolutamente continua e che $X_1\overset{\mathcal{L}}{=}X_2$. Dunque, il fatto che le componenti marginali siano assolutamente continue non implica che il v.v.a. sia assolutamente continuo. □

3.6. Indipendenza di variabili aleatorie

Risulta molto importante estendere il concetto di indipendenza che è stato introdotto per un insieme di eventi ad un insieme di v.a. definite sullo stesso spazio fondamentale. Si ha la seguente definizione formale che in effetti è una conseguenza della Definizione 2.5.5.

Definizione 3.6.1. Le v.a. $X_1, \dots, X_n$ definite sullo spazio probabilizzato $(\Omega, \mathcal{F}, P)$ sono dette *stocasticamente indipendenti* (o semplicemente *indipendenti*) se si ha

$$P\left(\bigcap_{k=1}^{n} \{X_k \in B_k\}\right) = \prod_{k=1}^{n} P(X_k \in B_k) \,,$$

per ogni $B_k \in \mathcal{B}(\mathbb{R})$ con $k = 1, \dots, n$. □

Si noti che la precedente Definizione implica che ogni possibile scelta di k v.a. $X_{j_1}, \dots, X_{j_k}$ con $k = 2, \dots, n$ risulta indipendente. Inoltre, posto $X = (X_1, \dots, X_n)^{\mathrm{T}}$, dal momento che la costruzione della σ-algebra di Borel $\mathcal{B}(\mathbb{R})$ può essere fatta a partire dalla classe di insiemi del tipo $]-\infty, x]$, allora la condizione data nella Definizione 3.6.1 è equivalente alla condizione

$$F_X(x_1, \dots, x_n) = P\left(\bigcap_{k=1}^{n} \{X_k \leq x_k\}\right) = \prod_{k=1}^{n} P(X_k \leq x_k) = \prod_{k=1}^{n} F_{X_k}(x_k) \,.$$

Proposizione 3.6.2. *Le v.a. discrete $X_1, \dots, X_n$ definite sullo spazio probabilizzato $(\Omega, \mathcal{F}, P)$ sono indipendenti se e solo se la f.p.c. p_X del v.v.a. $X = (X_1, \dots, X_n)^{\mathrm{T}}$ è data da*

$$p_X(x_1, \dots, x_n) = \prod_{k=1}^{n} p_{X_k}(x_k) \,,$$

dove p_{X_k} è la f.p.m. della k-esima componente.

Dimostrazione. Se $X_1, \dots, X_n$ sono indipendenti, dalla Definizione 3.6.1 si ha come caso particolare

$$p_X(x_1, \dots, x_n) = P\left(\bigcap_{k=1}^{n} \{X_k = x_k\}\right) = \prod_{k=1}^{n} P(X_k = x_k) = \prod_{k=1}^{n} p_{X_k}(x_k) \,.$$

Inversamente, se è vera la fattorizzazione, se il v.v.a. X prende valori su $S \in \mathbb{R}^n$ e S_k rappresenta la proiezione di S sul k-esimo asse cartesiano, sommando opportunamente si ha

$$\begin{aligned} P\left(\bigcap_{k=1}^{n} \{X_k \in B_k\}\right) &= \sum_{x_1 \in S_1} \dots \sum_{x_n \in S_n} \mathbf{1}_{B_1}(x_1) \dots \mathbf{1}_{B_n}(x_n) p_X(x_1, \dots, x_n) \\ &= \sum_{x_1 \in S_1} \dots \sum_{x_n \in S_n} \prod_{k=1}^{n} \mathbf{1}_{B_k}(x_k) \prod_{k=1}^{n} p_{X_k}(x_k) \\ &= \prod_{k=1}^{n} \sum_{x_k \in S_k} \mathbf{1}_{B_k}(x_k) p_{X_k}(x_k) = \prod_{k=1}^{n} P(X_k \in B_k) \,. \end{aligned}$$

La precedente relazione è equivalente alla condizione della Definizione 3.6.1 e implica la tesi. □

Proposizione 3.6.3. *Le v.a. assolutamente continue $X_1, \dots, X_n$ definite sullo spazio probabilizzato $(\Omega, \mathcal{F}, P)$ sono indipendenti se e solo se la d.p.c. f_X del v.v.a. $X = (X_1, \dots, X_n)^{\mathrm{T}}$ è data da*

$$f_X(x_1,\dots,x_n) = \prod_{k=1}^{n} f_{X_k}(x_k)\,,$$

dove f_{X_k} è la d.p.m. della k-esima componente.

Dimostrazione. Dal momento che la Definizione 3.6.1 è equivalente alla relazione $F_X(x_1,\dots,x_n) = \prod_{k=1}^{n} F_{X_k}(x_k)$, allora è sufficiente verificare che

$$\begin{aligned} f_X(x_1,\dots,x_n) &= \frac{\partial^n}{\partial x_1 \dots \partial x_n} F_X(x_1,\dots,x_n) = \frac{\partial^n}{\partial x_1 \dots \partial x_n} \prod_{k=1}^{n} F_{X_k}(x_k) \\ &= \prod_{k=1}^{n} \frac{d}{dx_k} F_{X_k}(x_k) = \prod_{k=1}^{n} f_{X_k}(x_k)\,, \end{aligned}$$

che è quanto si voleva dimostrare. □

• **Esempio 3.6.1.** Si consideri il v.v.a. $X = (X_1, X_2)^{\mathrm{T}}$ che ammette f.p.c.

$$p_X(x_1,x_2) = \frac{1}{2}\left(\frac{1}{4}\right)^{x_2}\left(\frac{3}{4}\right)^{1-x_2} \mathbf{1}_{\{0,1\}}(x_1)\mathbf{1}_{\{0,1\}}(x_2)\,.$$

Evidentemente, risulta $S = \{(0,0),(1,0),(0,1),(1,1)\}$ e $S_1 = S_2 = \{0,1\}$. La v.a. X_1 ammette f.p.m. data da

$$p_{X_1}(x_1) = \sum_{x_2 \in S_2} p_X(x_1,x_2) = \frac{1}{2}\,\mathbf{1}_{\{0,1\}}(x_2)\,,$$

mentre la v.a. X_2 ammette f.p.m. data da

$$p_{X_2}(x_2) = \sum_{x_1 \in S_1} p_X(x_1,x_2) = \left(\frac{1}{4}\right)^{x_2}\left(\frac{3}{4}\right)^{1-x_2} \mathbf{1}_{\{0,1\}}(x_2)\,.$$

Incidentalmente, si noti che il v.v.a. X possiede componenti marginali che hanno le stesse f.p.m. del v.v.a. dell'Esempio 3.5.1, anche se la f.p.c. risulta differente. Essendo

$$p_X(x_1,x_2) = p_{X_1}(x_1)p_{X_2}(x_2)\,,$$

le v.a. X_1 e X_2 sono indipendenti. Al contrario, è immediato verificare che le v.a. X_1 e X_2 dell'Esempio 3.5.1 non sono indipendenti. □

• **Esempio 3.6.2.** Si consideri il v.v.a. $X = (X_1, X_2)^{\mathrm{T}}$ che ammette d.p.c.

$$f_X(x_1,x_2) = 4x_1x_2\,\mathbf{1}_{]0,1[}(x_1)\mathbf{1}_{]0,1[}(x_2)\,.$$

La v.a. X_1 ammette d.p.m. data da

$$f_{X_1}(x_1) = \int_0^1 4x_1x_2\,\mathbf{1}_{]0,1[}(x_1)\,dx_2 = 2x_1\,\mathbf{1}_{]0,1[}(x_1)$$

e quindi per simmetria la v.a. X_2 ammette d.p.m.

$$f_{X_2}(x_2) = 2x_2\,\mathbf{1}_{]0,1[}(x_2)\,.$$

Essendo

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

le v.a. X_1 e X_2 sono indipendenti. Le leggi associate alle v.a. X_1 e X_2 si hanno per particolari parametrizzazioni della legge Beta (si veda Sezione 6.9). □

• **Esempio 3.6.3.** Si consideri il v.v.a. assolutamente continuo $X = (X_1, X_2)^{\mathrm{T}}$ che ammette d.p.c.

$$f_X(x_1, x_2) = \frac{1}{\pi}\, \mathbf{1}_{[0,1[}(x_1^2 + x_2^2)\,.$$

La legge associata al v.v.a. è detta Uniforme sul cerchio unitario. La v.a. X_1 ammette d.p.m. data da

$$f_{X_1}(x_1) = \int_{-\sqrt{1-x_1^2}}^{\sqrt{1-x_1^2}} \frac{1}{\pi}\, \mathbf{1}_{]-1,1[}(x_1)\, dx_2 = \frac{2}{\pi}\, \sqrt{1 - x_1^2}\, \mathbf{1}_{]-1,1[}(x_1)$$

e quindi per simmetria la v.a. X_2 ammette d.p.m.

$$f_{X_2}(x_2) = \frac{2}{\pi}\, \sqrt{1 - x_2^2}\, \mathbf{1}_{]-1,1[}(x_2)\,.$$

Le v.a. X_1 e X_2 non sono indipendenti in quanto

$$f_X(x_1, x_2) \neq f_{X_1}(x_1) f_{X_2}(x_2)\,.$$ □

3.7. Trasformate di variabili aleatorie

In molti casi è necessario considerare una funzione di una certa v.a. (o di un certo v.v.a.), piuttosto che la v.a. (o il v.v.a.) originale. Formalmente, dato lo spazio probabilizzato $(\Omega, \mathcal{F}, P)$ e la v.a. $X : \Omega \to \mathbb{R}$, se $g : \mathbb{R} \to \mathbb{R}$ è una funzione misurabile, allora $Y = g(X)$ è detta *trasformata* della v.a. X. Dal momento che g è una funzione misurabile e che per ogni $B \in \mathcal{B}(\mathbb{R})$ si ha

$$Y^{-1}(B) = X^{-1}(g^{-1}(B))\,,$$

allora segue immediatamente che Y è una v.a. in base alla Definizione 3.1.1. Inoltre, la legge indotta dalla v.a. Y è data da

$$P_Y(B) = P_X(g^{-1}(B)) = P(g(X) \in B) = P(X \in g^{-1}(B))\,.$$

Dalla precedente relazione si può ottenere la f.r. della v.a. Y qualora si ponga $B =]-\infty, y]$. Inoltre, identici risultati possono adeguati a trasformate di v.v.a., assumendo in questo caso che $g : \mathbb{R}^n \to \mathbb{R}^m$ sia una funzione misurabile. Anche se il problema di determinare la legge e la f.r. di una trasformata è concettualmente semplice, la gestione di ogni caso specifico può richiedere tuttavia una certa abilità di calcolo.

• **Esempio 3.7.1.** Data la v.a. X, si consideri la trasformata $Y = |X|$. In questo caso, per ogni $y \geq 0$ si ha

$$P(Y \leq y) = P(X \in [-y, y])$$

e dunque

$$F_Y(y) = P(Y \leq y) = (F_X(y) - F_X(-y) + P(X = -y))\, \mathbf{1}_{[0,\infty[}(y)\,.$$

Quando la v.a. X è assolutamente continua, allora risulta

$$F_Y(y) = (F_X(y) - F_X(-y))\, \mathbf{1}_{[0,\infty[}(y)$$

e quindi anche la v.a. Y è assolutamente continua e ammette d.p. data da

$$f_Y(y) = \frac{d}{dy} F_Y(y) = (f_X(y) + f_X(-y))\, \mathbf{1}_{[0,\infty[}(y) \,.$$

Infine, se la v.a. X è simmetrica rispetto all'origine, ovvero $X \stackrel{\mathcal{L}}{=} -X$, allora

$$f_Y(y) = 2 f_X(y)\, \mathbf{1}_{[0,\infty[}(y) \,.$$

Nel caso particolare in cui

$$f_X(x) = \frac{1}{2}\, \mathbf{1}_{]-1,1[}(x) \,,$$

si ha dunque

$$f_Y(y) = \mathbf{1}_{]0,1[}(y) \,.$$ □

• **Esempio 3.7.2.** Dato il v.v.a. $X = (X_1, X_2)^{\mathrm{T}}$, si consideri la trasformata $Y = X_1 + X_2$. Si ha

$$F_Y(y) = P(Y \leq y) = P(X_1 + X_2 \leq y) \,.$$

Quando il v.v.a. X è assolutamente continuo, dalla precedente espressione risulta

$$F_Y(y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_X(x_1, x_2)\, \mathbf{1}_{]-\infty,y[}(x_1 + x_2)\, dx_1 dx_2 = \int_{-\infty}^{\infty}\int_{-\infty}^{y-x_1} f_X(x_1, x_2)\, dx_1 dx_2$$

e quindi anche la v.a. Y è assolutamente continua con d.p. data da

$$f_Y(y) = \frac{d}{dy} F_Y(y) = \int_{-\infty}^{\infty} f_X(x_1, y - x_1)\, dx_1 \,.$$

Nel caso particolare in cui

$$f_X(x_1, x_2) = \mathbf{1}_{]0,1[}(x_1)\mathbf{1}_{]0,1[}(x_2) \,,$$

si ha

$$f_Y(y) = \int_0^1 \mathbf{1}_{]0,1[}(y - x_1)\, dx_1 = \min(y, 2 - y)\, \mathbf{1}_{]0,2[}(y) = (1 - |y - 1|)\, \mathbf{1}_{]0,2[}(y) \,.$$

In questo caso, la legge associata alla v.a. Y si ottiene per una particolare parametrizzazione della cosiddetta legge Triangolare. Si noti inoltre che un metodo più elegante per ottenere la legge di una somma di v.a. sarà considerato nel Capitolo 7. □

Di seguito vengono considerati alcuni risultati utili per determinare la f.r. e la d.p. di una trasformata di una v.a. assolutamente continua. Successivamente, vengono analizzate estensioni di questi metodi a trasformate di v.v.a. assolutamente continui.

Proposizione 3.7.1. *Sia data la trasformata $Y = g(X)$ della v.a. X assolutamente continua definita sullo spazio probabilizzato $(\Omega, \mathcal{F}, P)$, con g funzione misurabile biunivoca. Se g è una funzione crescente, allora si ha*

$$F_Y(y) = F_X(g^{-1}(y)) \,,$$

mentre se g è una funzione decrescente, allora si ha

$$F_Y(y) = 1 - F_X(g^{-1}(y)) \,.$$

Inoltre, la d.p. della v.a. Y è data da

$$f_Y(y) = f_X(g^{-1}(y)) \left| \frac{d}{dy} g^{-1}(y) \right| .$$

Dimostrazione. Se g è crescente si ha

$$F_Y(y) = P(g(X) \leq y) = P(X \leq g^{-1}(y)) = F_X(g^{-1}(y)) ,$$

mentre se g è decrescente si ha

$$F_Y(y) = P(g(X) \leq y) = P(X \geq g^{-1}(y)) = 1 - F_X(g^{-1}(y)) ,$$

in quanto l'immagine inversa dell'insieme $]-\infty, a]$ risulta $[a, \infty[$. La seconda parte della Proposizione si ottiene immediatamente derivando le precedenti espressioni della f.r. F_Y. □

La precedente Proposizione può essere estesa al caso in cui g non sia una funzione biunivoca, ma risulti biunivoca su ogni elemento di una partizione finita $B_1, \ldots, B_n$ di $\mathbb{R}$, ovvero $\bigcup_{k=1}^n B_k = \mathbb{R}$. Sotto questa ipotesi si ha

$$g(x) = \sum_{k=1}^n g_k(x) ,$$

dove $g_k(x) = g(x)\mathbf{1}_{B_k}(x)$ e dunque, applicando opportunamente la Proposizione 3.7.1, si verifica che la d.p. di Y è data da

$$f_Y(y) = \sum_{k=1}^n f_X(g_k^{-1}(y)) \left| \frac{d}{dy} g_k^{-1}(y) \right| .$$

• **Esempio 3.7.3.** Data la v.a. X assolutamente continua, si consideri la trasformata $Y = X^2$. Dal momento che la funzione $y : x \mapsto x^2$ è decrescente in $B_1 =]-\infty, 0]$ e crescente in $B_2 =]0, \infty[$, allora

$$f_Y(y) = (f_X(\sqrt{y}) + f_X(-\sqrt{y})) \frac{1}{2\sqrt{y}} \mathbf{1}_{[0,\infty[}(y) .$$

Nel caso particolare in cui

$$f_X(x) = \frac{1}{2} \mathbf{1}_{]-1,1[}(x) ,$$

tenendo presente che la v.a. X è simmetrica rispetto all'origine, si ha

$$f_Y(y) = \frac{1}{2\sqrt{y}} \mathbf{1}_{]0,1[}(y) . \qquad \square$$

Proposizione 3.7.2. *Sia dato il v.v.a. X assolutamente continuo definito sullo spazio probabilizzato $(\Omega, \mathcal{F}, P)$ e sia $g : \mathbb{R}^n \to \mathbb{R}^n$ un diffeomorfismo. La d.p.c. della trasformata data dal v.v.a. $Y = g(X)$ risulta*

$$f_Y(y) = f_X(g^{-1}(y)) \, |J(g^{-1}(y))| ,$$

dove $J(g^{-1}(y))$ è lo Jacobiano relativo alla funzione g^{-1} nel punto y.

Dimostrazione. Per ogni $B \in \mathcal{B}(\mathbb{R}^n)$ si ha

$$P_Y(B) = P_X(g^{-1}(B)) = \int_{g^{-1}(B)} f_X(x_1, \ldots, x_n)\, dx_1 \ldots dx_n\,.$$

Tenendo presente il commento successivo al Teorema A.8, allora risulta

$$P_Y(B) = \int_B f_X(g^{-1}(y_1, \ldots, y_n))\, |J(g^{-1}(y_1, \ldots, y_n))|\, dy_1 \ldots dy_n\,,$$

da cui segue la tesi. □

In modo simile a quanto visto per una singola v.a., la Proposizione 3.7.2 può essere estesa al caso in cui g sia un diffeomorfismo su ogni elemento di una partizione $B_1, \ldots, B_n$ di $\mathbb{R}^n$, ovvero in questo caso la d.p.c. del v.v.a. Y è data da

$$f_Y(y) = \sum_{k=1}^{n} f_X(g_k^{-1}(y))\, |J(g_k^{-1}(y))|\,,$$

dove $g_k(x) = g(x)\mathbf{1}_{B_k}(x)$.

• **Esempio 3.7.4.** Dato il v.v.a. $X = (X_1, X_2)^{\mathrm{T}}$ assolutamente continuo, si vuole determinare la d.p. della v.a. somma $(X_1 + X_2)$. A questo fine è conveniente considerare la trasformata $Y = (Y_1, Y_2)^{\mathrm{T}} = g(X)$, dove $g(x) = (x_1, x_1 + x_2)^{\mathrm{T}}$. Dal momento che $g^{-1}(y) = (y_1, y_2 - y_1)^{\mathrm{T}}$, allora $|J(g^{-1}(y))| = 1$, e quindi

$$f_Y(y) = f_Y(y_1, y_2) = f_X(y_1, y_2 - y_1)\,.$$

La d.p.m. della componente $Y_2 = X_1 + X_2$ è dunque data da

$$f_{Y_2}(y_2) = \int_{-\infty}^{\infty} f_X(y_1, y_2 - y_1)\, dy_1\,.$$

Ovviamente questo risultato coincide con quello ottenuto nell'Esempio 3.7.2. Si noti che, nel caso in cui le v.a. X_1 e X_2 siano indipendenti, si ha

$$f_{Y_2}(y_2) = \int_{-\infty}^{\infty} f_{X_1}(y_1) f_{X_2}(y_2 - y_1)\, dy_1\,.$$

Se si vuole determinare la d.p. della v.a. differenza $(X_1 - X_2)$, in modo simile si considera la trasformata $Y = (Y_1, Y_2)^{\mathrm{T}} = g(X)$, dove $g(x) = (x_1, x_1 - x_2)^{\mathrm{T}}$. In questo caso, la d.p.m. della componente $Y_2 = X_1 - X_2$ è data da

$$f_{Y_2}(y_2) = \int_{-\infty}^{\infty} f_X(y_1, y_1 + y_2)\, dy_1\,,$$

mentre se le v.a. X_1 e X_2 sono indipendenti risulta

$$f_{Y_2}(y_2) = \int_{-\infty}^{\infty} f_{X_1}(y_1) f_{X_2}(y_1 + y_2)\, dy_1\,.$$ □

• **Esempio 3.7.5.** Dato il v.v.a. assolutamente continuo $X = (X_1, X_2)^{\mathrm{T}}$, si vuole determinare la d.p. della v.a. prodotto $X_1 X_2$. A questo fine è conveniente considerare la trasformata $Y = (Y_1, Y_2)^{\mathrm{T}} = g(X)$, dove $g(x) = (x_1, x_1 x_2)^{\mathrm{T}}$. Dal momento che si ha $g^{-1}(y) = (y_1, \frac{y_2}{y_1})^{\mathrm{T}}$, allora $|J(g^{-1}(y))| = |y_1|^{-1}$, e quindi

$$f_Y(y) = f_X\left(y_1, \frac{y_2}{y_1}\right) \frac{1}{|y_1|} \, .$$

La d.p.m. della componente $Y_2 = X_1 X_2$ è dunque data da

$$f_{Y_2}(y_2) = \int_{-\infty}^{\infty} f_X\left(y_1, \frac{y_2}{y_1}\right) \frac{1}{|y_1|} \, dy_1 \, ,$$

che nel caso in cui le v.a. X_1 e X_2 siano indipendenti, si riduce a

$$f_{Y_2}(y_2) = \int_{-\infty}^{\infty} f_{X_1}(y_1) f_{X_2}\left(\frac{y_2}{y_1}\right) \frac{1}{|y_1|} \, dy_1 \, .$$

In modo simile, la d.p. della v.a. rapporto $Y_2 = \frac{X_2}{X_1}$ è data da

$$f_{Y_2}(y_2) = \int_{-\infty}^{\infty} f_X(y_1, y_1 y_2) |y_1| \, dy_1 \, ,$$

mentre se le v.a. X_1 e X_2 sono indipendenti risulta

$$f_{Y_2}(y_2) = \int_{-\infty}^{\infty} f_{X_1}(y_1) f_{X_2}(y_1 y_2) |y_1| \, dy_1 \, .$$

Nel caso particolare in cui si considera la d.p.c. dell'Esempio 3.7.2, allora la d.p.m. della trasformata $Y_2 = X_1 X_2$ risulta

$$f_{Y_2}(y_2) = \int_0^1 \mathbf{1}_{]0,1[}\left(\frac{y_2}{y_1}\right) \frac{1}{y_1} \, dy_1 = \, -\log(y_2) \, \mathbf{1}_{]0,1[}(y_2) \, ,$$

mentre la d.p.m. della trasformata $Y_2 = \frac{X_2}{X_1}$ risulta

$$f_{Y_2}(y_2) = \int_0^1 \mathbf{1}_{[0,1]}(y_1 y_2) y_1 \, dy_1 = \frac{1}{2} \, \mathbf{1}_{]0,1]}(y_2) + \frac{1}{2y_2^2} \, \mathbf{1}_{]1,\infty[}(y_2) \, . \qquad \square$$

• **Esempio 3.7.6.** (Paradosso delle corde di Bertrand) Sia data la legge Uniforme sul cerchio unitario introdotta nell'Esempio 3.6.3, e si consideri la trasformata in coordinate polari $Y = (Y_1, Y_2)^{\mathrm{T}} = g(X)$, dove $g^{-1}(y) = (y_1 \cos(y_2), y_1 \sin(y_2))^{\mathrm{T}}$. Poichè $|J(g^{-1}(y))| = y_1$, si ha

$$f_Y(y) = \frac{1}{\pi} \, y_1 \, \mathbf{1}_{]0,1[}(y_1) \mathbf{1}_{]0,2\pi[}(y_2) \, .$$

Dunque, è immediato verificare che le v.a. Y_1 e Y_2 sono indipendenti e che la d.p.m. di Y_1 è data da

$$f_{Y_1}(y_1) = 2 y_1 \, \mathbf{1}_{]0,1[}(y_1) \, ,$$

mentre la d.p.m. di Y_2 è data da

$$f_{Y_2}(y_2) = \frac{1}{2\pi} \, \mathbf{1}_{]0,2\pi[}(y_2) \, .$$

La legge associata alla v.a. Y_1 si ottiene per una particolare parametrizzazione della legge Beta (si veda Sezione 6.9), mentre la legge associata alla v.a. Y_2 si ottiene per una particolare parametrizzazione della legge Uniforme. Questi risultati possono essere applicati per illustrare il cosiddetto paradosso delle corde considerato da Joseph Bertrand nel suo testo *Calcul des probabilités*. Il problema consiste nello scegliere "casualmente" una corda nel cerchio unitario e nel determinare la probabilità che la corda sia più lunga del lato del triangolo equilatero inscritto nel cerchio (ovvero di

$\sqrt{3}$). Il paradosso deriva dal fatto che nella formulazione del problema non è ben definita la v.a. d'interesse, ovvero si hanno differenti risultati a secondo del modo in cui viene selezionata la corda. Ad esempio, se la corda viene scelta generando uniformemente un punto nel cerchio e in modo tale che il punto corrisponda al punto medio della corda, allora la probabilità richiesta è data da

$$P(2\sqrt{1-(X_1^2+X_2^2)} \geq \sqrt{3}) = P\left(1 - Y_1^2 \geq \frac{3}{4}\right) = P\left(Y_1 \leq \frac{1}{2}\right) = 2\int_0^{\frac{1}{2}} y_1 \, dy_1 = \frac{1}{4} \, .$$

Se invece si genera un punto casuale sul diametro perpendicolare al lato del triangolo equilatero e si sceglie la corda perpendicolare a questo punto, allora se la legge della v.a. Z è Uniforme su $[-1, 1]$, ovvero ammette densità data da $f_Z(z) = \frac{1}{2}\mathbf{1}_{[-1,1]}(z)$, la probabilità richiesta è data da

$$P(2\sqrt{1-Z^2} \geq \sqrt{3}) = P\left(1 - Z^2 \geq \frac{3}{4}\right) = P\left(|Z| \leq \frac{1}{2}\right) = \frac{1}{2}\int_{-\frac{1}{2}}^{\frac{1}{2}} dz = \frac{1}{2} \, .$$

Se infine la corda viene scelta in modo che uno dei suoi estremi coincida con un estremo del lato del triangolo e l'altro estremo abbia una distribuzione uniforme sulla circonferenza unitaria, allora la probabilità richiesta è data da

$$P(\sqrt{2-2\cos(Y_2)} \geq \sqrt{3}) = P\left(\cos(Y_2) \leq \frac{1}{2}\right) = P\left(\frac{\pi}{6} \leq Y_2 \leq \frac{5\pi}{6}\right) = \frac{1}{2\pi}\int_{\frac{\pi}{6}}^{\frac{5\pi}{6}} dy_2 = \frac{1}{3} \, . \quad \square$$

Proposizione 3.7.3. *Sia dato il v.v.a. $X = (X_1, \ldots, X_n)^{\mathrm{T}}$ con componenti indipendenti definito sullo spazio probabilizzato $(\Omega, \mathcal{F}, P)$. Se $Y = (Y_1, \ldots, Y_n)^{\mathrm{T}} = g(X)$ è una trasformata misurabile tale che $Y_k = g_k(X_k)$, le componenti del v.v.a. Y sono indipendenti.*

Dimostrazione. Dal momento che risulta

$$\begin{aligned} P\left(\bigcap_{k=1}^{n} \{g_k(X_k) \in B_k\}\right) &= P\left(\bigcap_{k=1}^{n} \{X_k \in g_k^{-1}(B_k)\}\right) = \prod_{k=1}^{n} P(X_k \in g_k^{-1}(B_k)) \\ &= \prod_{k=1}^{n} P(g_k(X_k) \in B_k) \end{aligned}$$

per ogni $B_k \in \mathcal{B}(\mathbb{R})$ con $k = 1, \ldots, n$, la tesi segue immediatamente dalla Definizione 3.6.1. $\square$

3.8. Riferimenti bibliografici

Una trattazione esauriente ed elegante delle proprietà della funzione di ripartizione è contenuta nel classico testo di Chung (2001). Numerosi esempi di equivalenze in legge e su trasformate di variabili aleatorie (o vettori aleatori) sono contenuti nei testi con esercizi svolti di Chaumont e Yor (2005), Dorogovtsev *et al.* (1997), Grimmett e Stirzaker (2001) e Hamming (1991). Infine, Devroye (1986) è un testo dedicato alla generazione di variabili pseudo-aleatorie con molteplici esempi di equivalenze in legge e proprietà delle trasformate.

3.9. Esercizi svolti

• **Nota.** Al fine di evitare notazioni ridondanti e ripetizioni, nei seguenti esercizi si assume l'esistenza di uno spazio probabilizzato $(\Omega, \mathcal{F}, P)$, a meno che non venga specificato diversamente.

Sezione 3.1

• **Esercizio 3.1.1.** Si consideri la funzione indicatrice $\mathbf{1}_E$ dell'evento $E \in \mathcal{F}$, ovvero $\mathbf{1}_E(\omega) = 1$ se $\omega \in E$ e $\mathbf{1}_E(\omega) = 0$ altrimenti. Si verifichi che $X = \mathbf{1}_E$ è in effetti una v.a.

Soluzione. Dal momento che la σ-algebra di Borel può essere costruita a partire dalla classe con elementi del tipo $]-\infty, x]$ è sufficiente verificare la definizione di v.a. per questo tipo di insiemi. Si osservi che risulta

$$X^{-1}(]-\infty, x]) = \begin{cases} \emptyset & x \in]-\infty, 0[\\ E^c & x \in [0, 1[\\ \Omega & x \in [1, \infty[\, , \end{cases}$$

ovvero $X^{-1}(]-\infty, x]) \in \mathcal{F}$ per ogni $x \in \mathbb{R}$. Dunque, X è una v.a. Inoltre, risulta immediato verificare che $\mathbf{1}_E$ è una v.a. discreta con legge essenziale data da $(P(E^c), P(E))$ con $S = \{0, 1\}$. Quindi, la v.a. X si distribuisce con legge di Bernoulli di parametro pari a $P(E)$. □

• **Esercizio 3.1.2.** Si ottengano le espressioni delle funzioni indicatrici $\mathbf{1}_{E^c}$, $\mathbf{1}_{E_1 \cap E_2}$ e $\mathbf{1}_{E_1 \cup E_2}$ con $E, E_1, E_2 \in \mathcal{F}$.

Soluzione. Evidentemente, risulta $\mathbf{1}_\Omega = 1$ per ogni $\omega \in \Omega$, mentre $\mathbf{1}_{E_1 \cup E_2} = \mathbf{1}_{E_1} + \mathbf{1}_{E_2}$ se E_1 e E_2 sono incompatibili. Dunque, si ha $\mathbf{1}_\Omega = \mathbf{1}_E + \mathbf{1}_{E^c}$, da cui $\mathbf{1}_{E^c} = 1 - \mathbf{1}_E$. Dalla precedente espressione si ottiene anche $\mathbf{1}_\emptyset = 0$ per ogni $\omega \in \Omega$. Inoltre, è immediato verificare che $\mathbf{1}_{E_1 \cap E_2} = \mathbf{1}_{E_1} \mathbf{1}_{E_2}$. Infine, esprimendo $E_1 \cup E_2$ come unione di eventi incompatibili, si ha

$$\mathbf{1}_{E_1 \cup E_2} = \mathbf{1}_{E_1 \setminus E_2} + \mathbf{1}_{E_2 \setminus E_1} + \mathbf{1}_{E_1 \cap E_2} \, ,$$

da cui, aggiungendo e sottraendo $\mathbf{1}_{E_1 \cap E_2}$, si ha infine

$$\mathbf{1}_{E_1 \cup E_2} = (\mathbf{1}_{E_1 \setminus E_2} - \mathbf{1}_{E_1 \cap E_2}) + (\mathbf{1}_{E_2 \setminus E_1} + \mathbf{1}_{E_1 \cap E_2}) - \mathbf{1}_{E_1 \cap E_2} = \mathbf{1}_{E_1} + \mathbf{1}_{E_2} - \mathbf{1}_{E_1} \mathbf{1}_{E_2} \, . \qquad \square$$

• **Esercizio 3.1.3.** Se $(E_k)_{k=1}^n$ è una classe di eventi di $\mathcal{F}$, si verifichi che

$$\mathbf{1}_{\bigcap_{k=1}^n E_k} = \prod_{k=1}^n \mathbf{1}_{E_k}$$

e

$$\mathbf{1}_{\bigcup_{k=1}^n E_k} = 1 - \prod_{k=1}^n (1 - \mathbf{1}_{E_k}) \, .$$

Soluzione. La prima relazione è valida per $n = 2$ sulla base dell'Esercizio 3.1.2. Inoltre, si ha

$$\mathbf{1}_{\bigcap_{k=1}^n E_k} = \mathbf{1}_{(\bigcap_{k=1}^{n-1} E_k) \cap E_n} = \mathbf{1}_{\bigcap_{k=1}^{n-1} E_k} \mathbf{1}_{E_n} \, ,$$

ovvero si ottiene la prima relazione per induzione. Sulla base dell'Esercizio 3.1.2, si ha

$$\mathbf{1}_{E_1 \cup E_2} = \mathbf{1}_{E_1} + \mathbf{1}_{E_2} - \mathbf{1}_{E_1} \mathbf{1}_{E_2} = 1 - (1 - \mathbf{1}_{E_1})(1 - \mathbf{1}_{E_2})$$

e quindi la seconda relazione è valida per $n = 2$. Inoltre, risulta

$$\mathbf{1}_{\bigcup_{k=1}^n E_k} = \mathbf{1}_{(\bigcup_{k=1}^{n-1} E_k) \cup E_n} = \mathbf{1}_{\bigcup_{k=1}^{n-1} E_k} + \mathbf{1}_{E_n} - \mathbf{1}_{\bigcup_{k=1}^{n-1} E_k} \mathbf{1}_{E_n} = 1 - (1 - \mathbf{1}_{\bigcup_{k=1}^{n-1} E_k})(1 - \mathbf{1}_{E_n})$$

da cui si ottiene anche la seconda relazione per induzione. □

Sezione 3.2

• **Esercizio 3.2.1.** (Variabili aleatorie simmetriche) Si consideri una v.a. X con f.r. F_X e tale che $X \stackrel{\mathcal{L}}{=} -X$. Si verifichi che si ha

$$F_X(x) = 1 - F_X(-x) + P(X = x) .$$

In questo caso, la v.a. X è detta simmetrica rispetto all'origine.

Soluzione. Dalla definizione di f.r. e sulla base delle assunzioni fatte si ha

$$\begin{aligned} F_X(x) &= P(X \le x) = P(-X \ge -x) = P(X \ge -x) \\ &= 1 - P(X < -x) = 1 - P(X \le -x) + P(X = -x) \\ &= 1 - P(X \le -x) + P(-X = x) = 1 - F_X(-x) + P(X = x) . \end{aligned}$$

Evidentemente, se X è una v.a. discreta con f.p. data da p_X, si ha

$$p_X(x) = P(X = x) = P(-X = x) = P(X = -x) = p_X(-x) .$$

Inoltre, se X è una v.a. assolutamente continua che ammette d.p. f_X, la relazione considerata si riduce a $F_X(x) = 1 - F_X(-x)$ e risulta

$$f_X(x) = \frac{d}{dx} F_X(x) = \frac{d}{dx} (1 - F_X(-x)) = f_X(-x) . \qquad \square$$

• **Esercizio 3.2.2.** (Mistura di leggi) Si assuma che $w_k \in [0, 1]$ per $k = 1, \ldots, n$ e $\sum_{k=1}^n w_k = 1$. Se $F_{X_1}, \ldots, F_{X_n}$ sono f.r. si verifichi che

$$F_X(x) = \sum_{k=1}^n w_k F_{X_k}(x)$$

è a sua volta una f.r.

Soluzione. Dal momento che $0 \le F_{X_k}(x) \le 1$ per ogni $x \in \mathbb{R}$ e $k = 1, \ldots, n$, si ha $F_X(x) \ge 0$ e $F_X(x) \le \sum_{k=1}^n w_k = 1$. Inoltre, si osservi che risulta

$$\lim_{x \to -\infty} F_X(x) = \sum_{k=1}^n w_k \lim_{x \to -\infty} F_{X_k}(x) = 0$$

e

$$\lim_{x \to \infty} F_X(x) = \sum_{k=1}^n w_k \lim_{x \to \infty} F_{X_k}(x) = \sum_{k=1}^n w_k = 1 ,$$

essendo $\lim_{x \to -\infty} F_{X_k}(x) = 0$ e $\lim_{x \to \infty} F_{X_k}(x) = 1$ per ogni $k = 1, \ldots, n$. Inoltre, se $x < y$ si ha $F_{X_k}(x) \le F_{X_k}(y)$ e dunque

$$F_X(x) = \sum_{k=1}^n w_k F_{X_k}(x) \le \sum_{k=1}^n w_k F_{X_k}(y) = F_X(y) .$$

Quindi, $F_X(x)$ è effettivamente una f.r. La legge considerata è in effetti una combinazione lineare convessa di leggi che viene usualmente denominata "mistura" di leggi. $\square$

Sezione 3.3

• **Esercizio 3.3.1.** (Legge Logaritmica) Si consideri la v.a. discreta X con f.p. data da

$$p_X(x) = -\frac{1}{\log(1-p)}\,\frac{p^x}{x}\,\mathbf{1}_{\{1,2,\dots\}}(x)\,,$$

dove $p \in]0,1[$. Si verifichi che p_X è in effetti una f.p.

Soluzione. Considerando la serie logaritmica, per $a \in]0,1[$ si verifica che

$$\sum_{n=1}^{\infty} \frac{a^n}{n} = -\log(1-a)\,.$$

Dunque, dal momento che $p_X(x) > 0$ per $x \in \{1,2,\dots\}$ e

$$\sum_{x=1}^{\infty} p_X(x) = -\frac{1}{\log(1-p)} \sum_{x=1}^{\infty} \frac{p^x}{x} = 1\,,$$

p_X è una f.p. La legge associata alla v.a. X è detta Logaritmica. □

• **Esercizio 3.3.2.** (Valori record, seconda parte) Si consideri la v.a. discreta X con f.p. data da

$$p_X(x) = \frac{1}{x(x+1)}\,\mathbf{1}_{\{1,2,\dots\}}(x)\,.$$

Si verifichi che p_X è effettivamente una f.p. e si determini un'espressione per la f.r. F_X.

Soluzione. Tenendo presente le proprietà delle serie telescopiche, si ha

$$\sum_{k=1}^{n} \frac{1}{k(k+1)} = \sum_{k=1}^{n} \left(\frac{1}{k} - \frac{1}{k+1}\right) = 1 - \frac{1}{n+1}$$

e quindi

$$\sum_{k=1}^{\infty} \frac{1}{k(k+1)} = \lim_n \left(1 - \frac{1}{n+1}\right) = 1\,.$$

Dunque, dal momento che $p_X(x) > 0$ per $x \in \{1,2,\dots\}$ e

$$\sum_{x=1}^{\infty} p_X(x) = \sum_{x=1}^{\infty} \frac{1}{x(x+1)} = 1\,,$$

p_X è effettivamente una f.p. Inoltre, risulta

$$F_X(x) = \sum_{u=1}^{\infty} \mathbf{1}_{]-\infty,x]}(u)\,\frac{1}{u(u+1)} = \mathbf{1}_{[1,\infty[}(x) \sum_{u=1}^{\lfloor x \rfloor} \frac{1}{u(u+1)} = \left(1 - \frac{1}{\lfloor x \rfloor + 1}\right)\mathbf{1}_{[1,\infty[}(x)\,.$$

Si osservi che un'assegnazione di probabilità di questo tipo è stata in effetti considerata nell'Esempio 2.7.3 relativo ai valori record. □

• **Esercizio 3.3.3.** (Legge di Benford) Si consideri la v.a. discreta X con f.p. data da

$$p_X(x) = \log_{10}\left(\frac{x+1}{x}\right)\mathbf{1}_{\{1,\dots,9\}}(x)\,.$$

Si verifichi che p_X è effettivamente una f.p. e si determini un'espressione per la f.r. F_X.

Soluzione. Tenendo presente le proprietà delle serie telescopiche, si ha

$$\sum_{k=1}^{n} \log_{10}\left(\frac{k+1}{k}\right) = \sum_{k=1}^{n} (\log_{10}(k+1) - \log_{10}(k)) = \log_{10}(n+1)\,.$$

Dunque, dal momento che $p_X(x) > 0$ per $x \in \{1, \ldots, 9\}$ e

$$\sum_{x=1}^{9} p_X(x) = \log_{10}(10) = 1\,,$$

p_X è effettivamente una f.p. Inoltre, risulta

$$\begin{aligned} F_X(x) &= \sum_{u=1}^{9} \mathbf{1}_{]-\infty,x]}(u) \log_{10}\left(\frac{u+1}{u}\right) = \mathbf{1}_{[1,9[}(x) \sum_{u=1}^{\lfloor x \rfloor} \log_{10}\left(\frac{u+1}{u}\right) + \mathbf{1}_{[9,\infty[}(x) \\ &= \log_{10}(\lfloor x \rfloor + 1)\, \mathbf{1}_{[1,9[}(x) + \mathbf{1}_{[9,\infty[}(x)\,. \end{aligned}$$

La legge prende il nome dall'ingegnere statunitense Frank Albert Benford (1883-1948), anche se introdotta originariamente dall'astronomo e matematico canadese Simon Newcomb (1835-1909). La legge è fondamentale nello studio della distribuzione della prima cifra significativa di un numero aleatorio. □

• **Esercizio 3.3.4.** Si consideri la v.a. discreta X con f.p. data da

$$p_X(x) = \binom{n+x}{n} 2^{-n-x}\ \mathbf{1}_{\{0,1,\ldots,n\}}(x)\,,$$

dove $n \in \mathbb{Z}^+$, e si verifichi che p_X è effettivamente una f.p.

Soluzione. Posto

$$C_n = \sum_{k=0}^{n} \binom{n+k}{n} 2^{-k}\,,$$

si deve verificare l'identità $C_n = 2^n$. Per le proprietà dei coefficienti binomiali si ha

$$\begin{aligned} C_n &= \sum_{k=0}^{n} \binom{n+k-1}{n} 2^{-k} + \sum_{k=0}^{n} \binom{n+k-1}{n-1} 2^{-k} \\ &= \frac{1}{2} \sum_{k=0}^{n-1} \binom{n+k}{n} 2^{-k} + \sum_{k=0}^{n-1} \binom{n+k-1}{n-1} 2^{-k} + \binom{2n-1}{n-1} 2^{-n} \\ &= \frac{1}{2}\, C_n - \frac{1}{2} \binom{2n}{n} 2^{-n} + C_{n-1} + \binom{2n-1}{n-1} 2^{-n} = \frac{1}{2}\, C_n + C_{n-1}\,, \end{aligned}$$

da cui si ha l'equazione ricorrente $C_n = 2C_{n-1}$. L'equazione è evidentemente soddisfatta per $C_n = 2^n$ e quindi l'identità richiesta è verificata. Dunque, dal momento che $p_X(x) > 0$ per $x \in \{0, 1, \ldots, n\}$ e

$$\sum_{x=0}^{n} p_X(x) = \sum_{x=0}^{n} \binom{n+x}{n} 2^{-n-x} = 1\,,$$

p_X è effettivamente una f.p. □

Sezione 3.4

• **Esercizio 3.4.1.** (Legge di Wigner) Si consideri la v.a. X che ammette d.p. data da

$$f_X(x) = \frac{2}{\pi r^2}\,\sqrt{r^2 - x^2}\,\mathbf{1}_{[-r,r]}(x)\,,$$

dove $r \in \,]0,\infty[$ e se ne determini la f.r.

Soluzione. Tenendo presente che

$$\int \sqrt{r^2 - x^2}\,dx = \frac{1}{2}\,x\sqrt{r^2 - x^2} + \frac{r^2}{2}\,\arcsin\Big(\frac{x}{r}\Big)\,,$$

si ha

$$\begin{aligned} F_X(x) &= \int_{-\infty}^{x} f_X(u)\,du = \frac{2}{\pi r^2}\,\mathbf{1}_{[-r,r[}(x)\int_{-r}^{x}\sqrt{r^2 - u^2}\,du + \mathbf{1}_{[r,\infty[}(x) \\ &= \left(\frac{1}{2} + \frac{1}{\pi r^2}\,x\sqrt{r^2 - x^2} + \frac{1}{\pi}\,\arcsin\Big(\frac{x}{r}\Big)\right)\mathbf{1}_{[-r,r[}(x) + \mathbf{1}_{[r,\infty[}(x)\,. \end{aligned}$$

La legge prende il nome dal fisico ungherese Eugene Wigner (1902-1995) e ha un ruolo fondamentale nella teoria delle matrici simmetriche a coefficienti aleatori. □

• **Esercizio 3.4.2.** (Legge di Cauchy circolare) Si consideri la v.a. X con f.r. data da

$$F_X(x) = \frac{1}{2} + \frac{1}{\pi}\,\arctan\left(\frac{1+\rho}{1-\rho}\,\tan\Big(\frac{x}{2}\Big)\right)\mathbf{1}_{]-\pi,\pi[}(x) + \mathbf{1}_{[\pi,\infty[}(x)\,,$$

dove $\rho \in \,]0,1]$. Si verifichi che F_X è effettivamente una f.r.

Soluzione. Evidentemente, si ha $\lim_{x\to-\infty}F_X(x) = 0$ e $\lim_{x\to\infty}F_X(x) = 1$. Inoltre, F_X è continua e risulta

$$f_X(x) = \frac{d}{dx}\,F_X(x) = \frac{1}{2\pi}\,\frac{1-\rho^2}{1+\rho^2-2\rho\cos(x)}\,\mathbf{1}_{]-\pi,\pi[}(x)\,.$$

Dal momento che per le assunzioni fatte su ρ si ha $1-\rho^2 \geq 0$ e

$$1+\rho^2-2\rho\cos(x) \geq 1+\rho^2-2\rho = (1-\rho)^2 \geq 0$$

per $x \in \,]-\pi,\pi[$, f_X è una funzione a valori non negativi e dunque F_X è una f.r. Questa legge ha importanti applicazioni nell'ambito della statistica direzionale e ha questa denominazione a causa del suo legame con la legge di Cauchy considerata nell'Esempio 4.1.4. □

Sezione 3.5

• **Esercizio 3.5.1.** Si consideri il v.v.a. discreto $X = (X_1, X_2)^{\mathrm{T}}$ con f.p.c.

$$p_X(x_1,x_2) = \frac{4}{n^2(n+1)^2}\,x_1x_2\,\mathbf{1}_{\{1,\ldots,n\}}(x_1)\mathbf{1}_{\{1,\ldots,n\}}(x_2)\,,$$

dove $n \in \mathbb{Z}^+$. Si determini le f.p.m. delle componenti marginali X_1 e X_2.

Soluzione. Tenendo presente che

$$\sum_{k=1}^{n} k = \frac{n(n+1)}{2}\,,$$

la f.p.m. di X_1 è data da

$$p_{X_1}(x_1) = \sum_{x_2=1}^{n} p_X(x_1, x_2) = \frac{4}{n^2(n+1)^2}\, x_1\, \mathbf{1}_{\{1,\ldots,n\}}(x_1) \sum_{x_2=1}^{n} x_2 = \frac{2}{n(n+1)}\, x_1\, \mathbf{1}_{\{1,\ldots,n\}}(x_1)\,.$$

Inoltre, dal momento che $p_X(x_2, x_1) = p_X(x_1, x_2)$ per simmetria, la f.p.m. di X_2 risulta

$$p_{X_2}(x_2) = \frac{2}{n(n+1)}\, x_2\, \mathbf{1}_{\{1,\ldots,n\}}(x_2)\,. \qquad \square$$

• **Esercizio 3.5.2.** (Legge di Farlie-Gumbel-Morgenstern) Si consideri il v.v.a. discreto $X = (X_1, X_2)^{\mathrm{T}}$ che ammette d.p.c.

$$f_X(x_1, x_2) = f_{X_1}(x_1) f_{X_2}(x_2)(1 + \alpha(2F_{X_1}(x_1) - 1)(2F_{X_2}(x_2) - 1))\,,$$

dove $\alpha \in [-1, 1]$, mentre F_{X_1}, F_{X_2} e f_{X_1}, f_{X_2}sono rispettivamente le f.r. e le d.p. delle v.a. X_1 e X_2. Si verifichi che f_X è una d.p.c.

Soluzione. Tenendo presente che

$$\int_{-\infty}^{\infty} f_{X_2}(x_2)(2F_{X_2}(x_2) - 1)\, dx_2 = 0\,,$$

si ha

$$\int_{-\infty}^{\infty} f_X(x_1, x_2)\, dx_2 = f_{X_1}(x_1)\,,$$

ovvero

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_X(x_1, x_2)\, dx_1 dx_2 = \int_{-\infty}^{\infty} f_{X_1}(x_1)\, dx_1 = 1\,.$$

Dal momento che f_X è anche una funzione non negativa, allora è effettivamente una d.p.c. $\square$

• **Esercizio 3.5.3.** Si consideri il v.v.a. assolutamente continuo $X = (X_1, X_2)^{\mathrm{T}}$ che ammette d.p.c.

$$f_X(x_1, x_2) = (x_1 + x_2 + x_1 x_2)\, e^{-x_1 - x_2 - x_1 x_2}\, \mathbf{1}_{[0,\infty[}(x_1) \mathbf{1}_{[0,\infty[}(x_2)\,.$$

Si determini la f.r.c. del v.v.a. X e le d.p.m. delle componenti X_1 e X_2.

Soluzione. La f.r.c. è data da

$$F_X(x) = \int_{-\infty}^{x_1}\int_{-\infty}^{x_2} f_X(u_1, u_2)\, du_1 du_2 = (1 - e^{-x_1} - e^{-x_2} + e^{-x_1 - x_2 - x_1 x_2})\, \mathbf{1}_{[0,\infty[}(x_1) \mathbf{1}_{[0,\infty[}(x_2)\,.$$

La d.p.m. della prima componente marginale è data da

$$f_{X_1}(x_1) = \int_{0}^{\infty} f_X(x_1, x_2)\, dx_2 = e^{-x_1}\, \mathbf{1}_{[0,\infty[}(x_1)$$

e quindi per simmetria risulta

$$f_{X_2}(x_2) = e^{-x_2}\, \mathbf{1}_{[0,\infty[}(x_2)\,. \qquad \square$$

Sezione 3.6

• **Esercizio 3.6.1.** Si consideri il v.v.a. assolutamente continuo $X = (X_1, X_2)^{\mathrm{T}}$ che ammette d.p.c.

$$f_X(x_1, x_2) = \frac{2}{\pi}\,(1 - x_1^2 - x_2^2)\,\mathbf{1}_{[0,1[}(x_1^2 + x_2^2)\,.$$

Si verifichi che le v.a. X_1 e X_2 non sono indipendenti. Si consideri inoltre la trasformata in coordinate polari $Y = (Y_1, Y_2)^{\mathrm{T}} = g(X)$, ovvero $g^{-1}(y) = (y_1\cos(y_2), y_1\sin(y_2))^{\mathrm{T}}$, e si verifichi che le v.a. Y_1 e Y_2 sono indipendenti.

Soluzione. La v.a. X_1 ammette d.p.m. data da

$$f_{X_1}(x_1) = \int_{-\sqrt{1-x_1^2}}^{\sqrt{1-x_1^2}} \frac{2}{\pi}\,(1 - x_1^2 - x_2^2)\,dx_2 = \frac{8}{3\pi}\,\sqrt{(1-x_1^2)^3}\,\mathbf{1}_{]-1,1[}(x_1)$$

e quindi per simmetria la v.a. X_2 ammette d.p.m.

$$f_{X_2}(x_2) = \frac{8}{3\pi}\,\sqrt{(1-x_2^2)^3}\,\mathbf{1}_{]-1,1[}(x_2)\,.$$

Le v.a. X_1 e X_2 non sono indipendenti in quanto

$$f_X(x_1, x_2) \neq f_{X_1}(x_1) f_{X_2}(x_2)\,.$$

Per quanto riguarda il v.v.a. Y, dal momento che $|J(g^{-1}(y))| = y_1$, la d.p.c. è data da

$$f_Y(y_1, y_2) = \frac{2}{\pi}\,y_1(1 - y_1^2)\,\mathbf{1}_{[0,1]}(y_1)\mathbf{1}_{[0,2\pi[}(y_2)\,.$$

Dunque, la d.p.m. di Y_1 è data da

$$f_{Y_1}(y_1) = \int_0^{2\pi} \frac{2}{\pi}\,y_1(1 - y_1^2)\,\mathbf{1}_{[0,1]}(y_1)\,dy_2 = 4y_1(1 - y_1^2)\,\mathbf{1}_{[0,1]}(y_1)\,,$$

mentre la d.p.m. di Y_2 è data da

$$f_{Y_2}(y_2) = \int_0^1 \frac{2}{\pi}\,y_1(1 - y_1^2)\,\mathbf{1}_{[0,2\pi[}(y_2)\,dy_1 = \frac{1}{2\pi}\,\mathbf{1}_{[0,2\pi[}(y_2)\,.$$

Quindi, le v.a. Y_1 e Y_2 sono indipendenti in quanto

$$f_Y(y_1, y_2) = f_{Y_1}(y_1) f_{Y_2}(y_2)\,.$$

Si deve concludere che trasformate di v.a. non indipendenti possono essere indipendenti. □

• **Esercizio 3.6.2.** Se $X = (X_1, X_2, X_3)^{\mathrm{T}}$ è un v.v.a. assolutamente continuo che ammette d.p.c.

$$f_X(x_1, x_2, x_3) = \frac{1}{8\pi^3}\,(1 - \sin(x_1)\sin(x_2)\sin(x_3))\,\mathbf{1}_{[0,2\pi[}(x_1)\mathbf{1}_{[0,2\pi[}(x_2)\mathbf{1}_{[0,2\pi[}(x_3)\,,$$

si verifichi che le v.a. X_1, X_2 e X_3 non sono indipendenti, anche se risultano reciprocamente indipendenti.

Soluzione. La d.p.c. del v.v.a. $(X_1, X_2)^{\mathrm{T}}$ è data da

$$\begin{aligned} f_{(X_1,X_2)}(x_1, x_2) &= \int_0^{2\pi} \frac{1}{8\pi^3}\,(1 - \sin(x_1)\sin(x_2)\sin(x_3))\,\mathbf{1}_{[0,2\pi[}(x_1)\mathbf{1}_{[0,2\pi[}(x_2)\,dx_3 \\ &= \frac{1}{4\pi^2}\,\mathbf{1}_{[0,2\pi[}(x_1)\mathbf{1}_{[0,2\pi[}(x_2)\,. \end{aligned}$$

Per simmetria le d.p.c. $f_{(X_1,X_2)}$, $f_{(X_1,X_3)}$ e $f_{(X_2,X_3)}$ hanno la medesima struttura. Inoltre, si ha

$$f_{X_1}(x_1) = \int_0^{2\pi} \frac{1}{4\pi^2}\, \mathbf{1}_{[0,2\pi[}(x_1)\, dx_2 = \frac{1}{2\pi}\, \mathbf{1}_{[0,2\pi[}(x_1)\ .$$

Di nuovo, per simmetria le d.p. f_{X_1}, f_{X_2} e f_{X_3} possiedono la medesima struttura. Dunque, il v.v.a. $(X_1, X_2)^{\mathrm{T}}$ ha componenti indipendenti in quanto

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

Data la simmetria, una simile considerazione vale anche per il v.v.a. $(X_1, X_3)^{\mathrm{T}}$ e il v.v.a. $(X_2, X_3)^{\mathrm{T}}$. Tuttavia, il v.v.a. X non ha componenti indipendenti in quanto

$$f_X(x_1, x_2, x_3) \neq f_{X_1}(x_1) f_{X_2}(x_2) f_{X_3}(x_3)\ ,$$

ovvero le componenti di X sono reciprocamente indipendenti, ma non indipendenti. □

Sezione 3.7

• **Esercizio 3.7.1.** (Legge Triangolare) Si consideri il v.v.a. $X = (X_1, X_2)^{\mathrm{T}}$ che ammette d.p.c. data da

$$f_X(x_1, x_2) = \mathbf{1}_{]0,1[}(x_1) \mathbf{1}_{]0,1[}(x_2)\ ,$$

e si determini la d.p. della v.a. $Y = X_1 - X_2$.

Soluzione. Risulta immediato verificare che le v.a. X_1 e X_2 sono indipendenti e ugualmente distribuite con legge Uniforme su $]0, 1[$. Dunque, sulla base dell'espressione della d.p. per la v.a. differenza di v.a. assolutamente continue, si ottiene

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X_1}(x) f_{X_2}(x + y)\, dx = \int_{-\infty}^{\infty} \mathbf{1}_{]0,1[}(x)\, \mathbf{1}_{]0,1[}(x + y)\, \mathbf{1}_{]-1,1[}(y)\, dx\ ,$$

ovvero, dal momento che risulta $\mathbf{1}_{]0,1[}(x)\mathbf{1}_{]0,1[}(x + y) = \mathbf{1}_{]-y,1[}(x)$ se $y \in]-1, 0[$, mentre si ha $\mathbf{1}_{]0,1[}(x)\mathbf{1}_{]0,1[}(x + y) = \mathbf{1}_{]0,1-y[}(x)$ se $y \in]0, 1[$, allora

$$\begin{aligned} f_Y(y) &= \int_{-y}^{1} \mathbf{1}_{]-1,0[}(y)\, dx + \int_0^{1-y} \mathbf{1}_{]0,1[}(y)\, dx \\ &= (1 + y)\, \mathbf{1}_{]-1,0[}(y) + (1 - y)\, \mathbf{1}_{]0,1[}(y) = (1 - |y|)\, \mathbf{1}_{]-1,1[}(y)\ . \end{aligned}$$

Evidentemente, la legge della v.a. Y è detta Triangolare a causa della morfologia della d.p. Tenendo presente l'Esempio 3.7.2, si ha quindi $X_1 - X_2 \stackrel{\mathcal{L}}{=} X_1 + X_2 - 1$. □

• **Esercizio 3.7.2.** (Legge Triangolare, seconda parte) Si consideri il v.v.a. $X = (X_1, X_2)^{\mathrm{T}}$ che ammette d.p.c. data da

$$f_X(x_1, x_2) = \frac{1}{4}\, \mathbf{1}_{]-1,1[}(x_1) \mathbf{1}_{]-1,1[}(x_2)\ ,$$

e si determini la d.p. della v.a. $Y = \frac{X_1 + X_2}{2}$.

Soluzione. Le v.a. X_1 e X_2 sono indipendenti e ugualmente distribuite con legge Uniforme su $]-1, 1[$. Se la v.a. Z è tale che $Z = X_1 + X_2$, si ottiene

$$f_Z(z) = \int_{-\infty}^{\infty} f_{X_1}(x) f_{X_2}(z - x)\, dx = \frac{1}{4} \int_{-\infty}^{\infty} \mathbf{1}_{]-1,1[}(x)\, \mathbf{1}_{]-1,1[}(z - x)\, \mathbf{1}_{]-2,2[}(z)\, dx\ ,$$

ovvero, dal momento che risulta $\mathbf{1}_{]-1,1[}(x)\mathbf{1}_{]-1,1[}(z - x) = \mathbf{1}_{]-1,1+z[}(x)$ se $z \in]-2, 0[$, mentre si ha $\mathbf{1}_{]-1,1[}(x)\mathbf{1}_{]-1,1[}(z - x) = \mathbf{1}_{]1-z,1[}(x)$ se $z \in]0, 2[$, allora

$$f_Z(z) = \frac{1}{4}\int_{-1}^{1+z} \mathbf{1}_{]-2,0[}(z)\,dx + \frac{1}{4}\int_{1-z}^{1} \mathbf{1}_{]0,2[}(z)\,dx$$
$$= \frac{2+z}{4}\,\mathbf{1}_{]-2,0[}(z) + \frac{2-z}{4}\mathbf{1}_{]0,2[}(z) = \frac{2-|z|}{4}\,\mathbf{1}_{]-2,2[}(z)\,.$$

Inoltre, essendo $Y = \frac{Z}{2}$, si ottiene

$$f_Y(y) = 2f_Z(2y) = (1-|y|)\,\mathbf{1}_{]-1,1[}(y)\,,$$

ovvero, tenendo presente l'Esercizio 7.3.1, la legge della v.a. Y è Triangolare. □

• **Esercizio 3.7.3.** Si consideri la v.a. X che ammette d.p. data da $f_X(x) = \mathbf{1}_{]0,1[}(x)$ e si determini la d.p. della v.a. $Y = \cos(2\pi X)$.

Soluzione. La trasformata $g(x) = \cos(2\pi x)$ non è biunivoca per $x \in]0,1[$, ma può essere espressa come somma di funzioni biunivoche su una partizione, ovvero

$$g(x) = g_1(x)\,\mathbf{1}_{]0,\frac{1}{2}[}(x) + g_2(x)\,\mathbf{1}_{]\frac{1}{2},1[}(x)\,,$$

dove $g_1(x) = \cos(2\pi x)$ e $g_2(x) = \cos(2\pi(1-x))$. Dunque, si ha

$$f_Y(y) = f_X(g_1^{-1}(y))\left|\frac{d}{dy}\,g_1^{-1}(y)\right| + f_X(g_2^{-1}(y))\left|\frac{d}{dy}\,g_2^{-1}(y)\right|,$$

ovvero, essendo $g_1^{-1}(y) = \frac{\arccos(y)}{2\pi}$ e $g_2^{-1}(y) = 1 - \frac{\arccos(y)}{2\pi}$, si ottiene

$$f_Y(y) = \frac{1}{\pi\sqrt{1-y^2}}\,\mathbf{1}_{[-1,1[}(y)\,.$$ □

• **Esercizio 3.7.4.** Dato il v.v.a. assolutamente continuo $X = (X_1, X_2, X_3)^{\mathrm{T}}$ che ammette d.p.c. data da

$$f_X(x_1, x_2, x_3) = h(x_1^2 + x_2^2 + x_3^2)$$

per una funzione opportuna h, si determini la d.p.c. della trasformata in coordinate sferiche, ovvero $Y = (Y_1, Y_2, Y_3)^{\mathrm{T}} = g(X)$, con

$$g^{-1}(y) = (y_1\,\mathrm{sin}(y_2)\mathrm{cos}(y_3), y_1\,\mathrm{sin}(y_2)\mathrm{sin}(y_3), y_1\,\mathrm{cos}(y_2))^{\mathrm{T}}$$

e dove $y_1 \in]0,\infty[$, $y_2 \in [0,\pi[$ e $y_3 \in [0,2\pi[$.

Soluzione. Dal momento che $|J(g^{-1}(y))| = y_1^2\mathrm{sin}(y_2)$, la d.p.c. è data da

$$f_Y(y) = y_1^2 h(y_1^2)\,\mathrm{sin}(y_2)\,\mathbf{1}_{]0,\infty[}(y_1)\mathbf{1}_{[0,\pi[}(y_2)\mathbf{1}_{[0,2\pi[}(y_3)\,.$$

Inoltre, si può verificare che le v.a. Y_1, Y_2 e Y_3 sono indipendenti. In effetti, la d.p.m. di Y_1 è data da

$$f_{Y_1}(y_1) = y_1^2 h(y_1^2)\,\mathbf{1}_{]0,\infty[}(y_1)\int_0^{\pi}\int_0^{2\pi} \mathrm{sin}(y_2)\,dy_2 dy_3 = 4\pi y_1^2 h(y_1^2)\,\mathbf{1}_{]0,\infty[}(y_1)$$

e, dovendo risultare dalla precedente espressione $4\pi\int_0^{\infty} y_1^2 h(y_1^2) dy_1 = 1$, la d.p.m. di Y_2 è data da

$$f_{Y_2}(y_2) = \mathrm{sin}(y_2)\,\mathbf{1}_{[0,\pi[}(y_2)\int_0^{\infty}\int_0^{2\pi} y_1^2 h(y_1^2)\,dy_1 dy_3 = \frac{1}{2}\,\mathrm{sin}(y_2)\,\mathbf{1}_{[0,\pi[}(y_2)\,.$$

Infine, si ha

$$f_{Y_3}(y_3) = \mathbf{1}_{[0,2\pi[}(y_3) \int_0^\infty \int_0^\pi y_1^2 h(y_1^2) \sin(y_2)\, dy_1 dy_2 = \frac{1}{2\pi}\, \mathbf{1}_{[0,2\pi[}(y_3)\,,$$

per cui la d.p.c. può essere fattorizzata come

$$f_Y(y) = f_{Y_1}(y_1) f_{Y_2}(y_2) f_{Y_3}(y_3)\,,$$

ovvero le componenti marginale del v.v.a. Y sono indipendenti. □

• **Esercizio 3.7.5.** (Convoluzione di v.a. a valori interi) Si consideri il v.v.a. $X = (X_1, X_2)^{\mathrm{T}}$ con f.p.c. p_X e legge essenziale definita su $\mathbb{N} \times \mathbb{N}$. Si determini la f.p. della v.a. $Y = X_1 + X_2$.

Soluzione. La legge essenziale di Y è evidentemente definita su $S = \mathbb{N}$. La f.p. della v.a. Y è data da

$$\begin{aligned} p_Y(y) &= P(X_1 + X_2 = y) = P(X_2 = y - X_1) \\ &= \sum_{x=0}^{y} P(\{X_1 = x\} \cap \{X_2 = y - x\}) = \sum_{x=0}^{y} p_X(x, y - x)\,. \end{aligned}$$

Inoltre, nel caso in cui le v.a. X_1 e X_2 siano indipendenti, risulta

$$p_Y(y) = \sum_{x=0}^{y} p_{X_1}(x) p_{X_2}(y - x)\,.$$

La precedente espressione è anche detta formula di convoluzione per v.a. a valori interi. □

• **Esercizio 3.7.6.** Si consideri il v.v.a. discreto $X = (X_1, X_2)^{\mathrm{T}}$ con f.p.c.

$$p_X(x_1, x_2) = 2^{-x_1 - x_2}\, \mathbf{1}_{\{1,2,\ldots\}}(x_1) \mathbf{1}_{\{1,2,\ldots\}}(x_2)\,.$$

Si determini la f.p. della v.a. $Y = \frac{X_1}{X_1 + X_2}$.

Soluzione. La v.a. Y è discreta e può assumere solamente valori frazionari. Dunque, la legge essenziale è definita su $S = \mathbb{Q} \cap]0, 1[$, ovvero S è un insieme denso su $]0, 1[$. Si assuma che $y \in S$ sia una frazione irriducibile, ovvero che risulti $y = \frac{m}{n}$, dove $m, n \in \mathbb{Z}^+$, con $m < n$, sono coprimi. Assumendo che $k \in \mathbb{Z}^+$, si ha

$$\begin{aligned} P(Y = y) &= P\left(\frac{X_1}{X_1 + X_2} = \frac{m}{n}\right) = P\left(\frac{X_1}{X_1 + X_2} = \frac{km}{kn}\right) \\ &= \sum_{k=1}^{\infty} P(\{X_1 = km\} \cap \{X_1 + X_2 = kn\}) \\ &= \sum_{k=1}^{\infty} P(\{X_1 = km\} \cap \{X_2 = k(n - m)\})\,. \end{aligned}$$

Dunque, se $n = 2, 3, \ldots$ e m è coprimo con n, la f.p. della v.a. Y è data da

$$p_Y\left(\frac{m}{n}\right) = \sum_{k=1}^{\infty} p_X(km, k(n - m)) = \sum_{k=1}^{\infty} 2^{-kn} = \frac{1}{2^n - 1}\,,$$

In modo notevole, la precedente probabilità non dipende da m, ovvero per tutte le frazioni $\frac{m}{n}$ con lo stesso denominatore si ottiene la medesima probabilità. Inoltre, si osservi che nella Teoria dei Numeri la cardinalità dei numeri coprimi con n è data dalla funzione toziente di Eulero $\varphi(n)$. Quindi, dal momento che p_Y è una f.p. deve risultare

$$\sum_{n=2}^{\infty} \frac{\varphi(n)}{2^n - 1} = 1 \,,$$

che in effetti costituisce un caso particolare di una nota identità relativa alla serie di Lambert per la funzione toziente. □

Capitolo 4

Valori attesi

4.1. Valore atteso di una variabile aleatoria

Il concetto di valore atteso è fondamentale nella Teoria della Probabilità. Formalmente, si ha la seguente definizione.

Definizione 4.1.1. Data la v.a. X definita sullo spazio probabilizzato $(\Omega, \mathcal{F}, P)$, si dice *valore atteso* l'integrale

$$\mathrm{E}[X] = \int X \, dP$$

se esiste finito. □

La simbologia adottata nella precedente definizione è basata sul termine inglese *expectation* o sul termine francese *espérance*. Il valore atteso è anche detto *valor medio* o *speranza matematica*. Inoltre, il valore atteso della v.a. X è spesso indicato con il simbolo μ_X, o semplicemente con μ quando il contesto non si presta a fraintendimenti.

Si osservi che se $X = \mathbf{1}_E$ dove $E \in \mathcal{F}$, allora risulta

$$\mathrm{E}[\mathbf{1}_E] = \int_E dP = P(E) \, ,$$

ovvero la probabilità di un evento può essere interpretata come un valore atteso. Tenendo presente il Teorema A.8, il valore atteso può anche essere espresso come

$$\mathrm{E}[X] = \int_{\mathbb{R}} x \, dP_X(x) \, .$$

Dalla precedente espressione e dal Teorema A.10, se la v.a. X è discreta a valori su insieme numerabile S con f.p. p_X, il valore atteso si riduce a

$$\mathrm{E}[X] = \int_{\mathbb{R}} x p_X(x) \, d\nu(x) = \sum_{x \in S} x p_X(x) \, ,$$

dove ν è la misura definita nella Sezione 3.3. Analogamente, se la v.a. X è assolutamente continua e ammette d.p. f_X, il valore atteso si riduce a

$$\mathrm{E}[X] = \int_{-\infty}^{\infty} x f_X(x) \, dx \, .$$

Infine, se X e Y sono v.a. tali che $X \stackrel{\mathcal{L}}{=} Y$, allora si ottiene immediatamente che $\mathrm{E}[X] = \mathrm{E}[Y]$.

• **Esempio 4.1.1.** Si consideri la v.a. discreta X analizzata nell'Esempio 3.3.1. In questo caso, il valore atteso della v.a. X è dato da

$$\mathrm{E}[X] = \sum_{x=0}^{\infty} x p_X(x) = \sum_{x=0}^{\infty} x 2^{-x-1} \,.$$

Tenendo presente che derivando ambo i membri della serie geometrica (si veda l'Esempio 3.3.1) rispetto ad a si ottiene

$$\sum_{n=1}^{\infty} n a^n = \sum_{n=0}^{\infty} n a^n = \frac{a}{(1-a)^2} \,,$$

allora, ponendo $a = \frac{1}{2}$ nella precedente relazione, il valore atteso risulta

$$\mathrm{E}[X] = \frac{1}{2} \sum_{x=0}^{\infty} x 2^{-x} = 1 \,. \qquad \square$$

• **Esempio 4.1.2.** Si consideri la v.a. assolutamente continua X analizzata nell'Esempio 3.4.1. In questo caso il valore atteso della v.a. X è dato da

$$\mathrm{E}[X] = \int_0^1 x f_X(x)\, dx = 3 \int_0^1 x^3\, dx = \frac{3}{4} \,. \qquad \square$$

Denotando con $X^+ = \max(X, 0)$ e $X^- = -\min(X, 0)$, il valore atteso può essere scritto come

$$\mathrm{E}[X] = \mathrm{E}[X^+] - \mathrm{E}[X^-] \,.$$

Dunque, il valore atteso esiste finito se si ha $\mathrm{E}[X^+] < \infty$ e $\mathrm{E}[X^-] < \infty$. Nel caso in cui uno dei valori attesi $\mathrm{E}[X^+]$ o $\mathrm{E}[X^-]$ non sia finito, allora il valore atteso $\mathrm{E}[X]$ non è finito. Infine, se entrambi i valori attesi $\mathrm{E}[X^+]$ e $\mathrm{E}[X^-]$ non sono finiti, allora il valore atteso $\mathrm{E}[X]$ non è definito, in quanto si ha la forma del tipo $+\infty - \infty$.

• **Esempio 4.1.3.** (Legge di Lévy) Si consideri la v.a. assolutamente continua X che ammette d.p. data da

$$f_X(x) = \frac{1}{\sqrt{2\pi x^3}}\, e^{-\frac{1}{2x}}\, \mathbf{1}_{]0,\infty[}(x) \,.$$

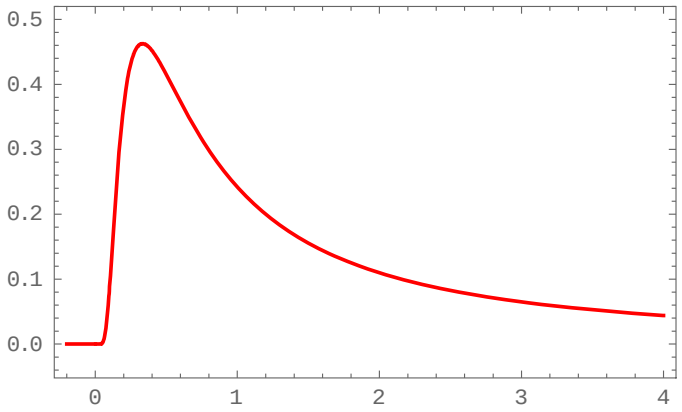


Figura 4.1.1. Densità di probabilità relativa alla legge ridotta di Lévy.

La legge associata a questa v.a. prende nome dal probabilista Paul Pierre Lévy (1886-1971) e in questo caso è considerata nella versione ridotta (si veda la Sezione 6.6). La legge ha interessanti

applicazioni nell'ambito finanziario e in fisica (per la genesi della legge, si veda la Sezione 9.4). Risulta immediato verificare che $\mathrm{E}[X^-] = 0$, mentre

$$\mathrm{E}[X^+] = \frac{1}{\sqrt{2\pi}} \int_0^\infty \frac{1}{\sqrt{x}}\, e^{-\frac{1}{2x}}\, dx \geq \frac{1}{\sqrt{2\pi}} \int_1^\infty \frac{1}{\sqrt{x}}\, e^{-\frac{1}{2x}}\, dx \geq \frac{e^{-\frac{1}{2}}}{\sqrt{2\pi}} \int_1^\infty \frac{1}{\sqrt{x}}\, dx = \infty$$

e quindi $\mathrm{E}[X]$ non è finito. □

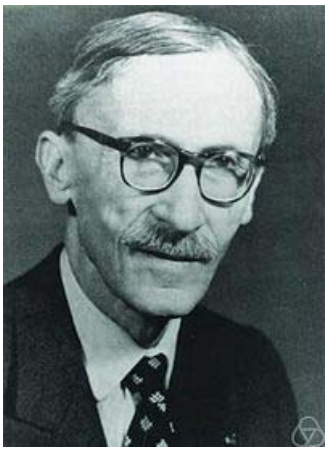


Figura 4.1.2. Paul Pierre Lévy (1886-1971).

• **Esempio 4.1.4.** (Legge di Cauchy) Si consideri la v.a. assolutamente continua X che ammette d.p. data da

$$f_X(x) = \frac{1}{\pi(1+x^2)}\,.$$

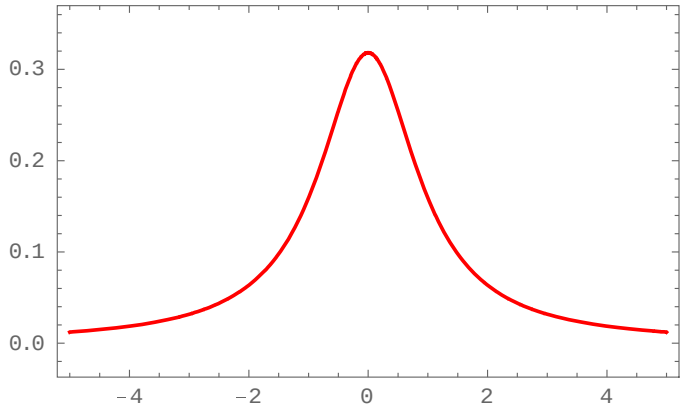


Figura 4.1.3. Densità di probabilità relativa alla legge ridotta di Cauchy.

La legge associata a questa v.a. è legata al nome del matematico Augustin-Louis Cauchy (1789-1857). Anche in questo caso la legge è considerata nella versione ridotta (si veda la Sezione 6.6). La legge ha importanti applicazioni in fisica e si ottiene per una particolare parametrizzazione della legge t di Student (si veda la Sezione 6.10). Data la simmetria di f_X risulta

$$\mathrm{E}[X^+] = \mathrm{E}[X^-] = \int_0^\infty \frac{x}{\pi(1+x^2)}\, dx = \infty$$

e quindi $\mathrm{E}[X]$ non è definito. □

Figura 4.1.4. Augustin-Louis Cauchy (1789-1857).

4.2. Proprietà del valore atteso

Le seguenti Proposizioni e Teoremi forniscono un insieme di proprietà, che riguardano alcune disuguaglianze e le condizioni di esistenza sul valore atteso.

Proposizione 4.2.1. *Si consideri la v.a. X definita sullo spazio probabilizzato $(\Omega, \mathcal{F}, P)$. Si ha*

$$|\mathrm{E}[X]| \leq \mathrm{E}[|X|]$$

e quindi se $\mathrm{E}[|X|]$ *esiste finito, allora anche* $\mathrm{E}[X]$ *esiste finito.*

Dimostrazione. Dal momento che $\{X \leq |X|\}$ $q.c.$, allora dal Teorema A.6 si ha

$$\mathrm{E}[X] \leq |\mathrm{E}[X]| = \left| \int X \, dP \right| \leq \int |X| \, dP = \mathrm{E}[|X|] \, ,$$

da cui segue immediatamente anche la seconda parte. □

Proposizione 4.2.2. *Si consideri la v.a. X definita sullo spazio probabilizzato $(\Omega, \mathcal{F}, P)$. Se $P(X \in [a, b]) = 1$ dove a e b sono costanti, allora si ha $a \leq \mathrm{E}[X] \leq b$. Quindi, se $a, b < \infty$, allora* $\mathrm{E}[X]$ *esiste finito.*

Dimostrazione. Dal momento che $\{X \geq a\}$ $q.c.$, dal Teorema A.6 si ha

$$\mathrm{E}[X] = \int X \, dP \geq a \int dP = a \, ,$$

mentre, dal momento che $\{X \leq b\}$ $q.c.$, allora

$$\mathrm{E}[X] = \int X \, dP \leq b \int dP = b \, ,$$

da cui segue anche la seconda parte. □

Se $g : \mathbb{R}^n \to \mathbb{R}$ è una funzione misurabile, risulta semplice generalizzare le Proposizioni 4.2.1 e 4.2.2 alla trasformata $Y = g(X)$ del v.v.a. $X = (X_1, \ldots, X_n)^{\mathrm{T}}$. Più esattamente, si ha che

$$|\mathrm{E}[g(X)]| \leq \mathrm{E}[|g(X)|]$$

e quindi se $\mathrm{E}[|g(X)|]$ esiste finito, allora anche $\mathrm{E}[g(X)]$ esiste finito. Inoltre, se $P(g(X) \in [a, b]) = 1$, si ha $a \leq \mathrm{E}[g(X)] \leq b$. Infine, la seguente Proposizione è fondamentale in quanto permette di ottenere

il valore atteso della trasformata $Y = g(X)$ senza dover prima determinare la legge della v.a. trasformata, come in effetti sembrerebbe necessario dalla definizione.

Proposizione 4.2.3. *Data la funzione misurabile $g : \mathbb{R}^n \to \mathbb{R}$, si consideri la trasformata $Y = g(X)$ del v.v.a. $X = (X_1, \ldots, X_n)^{\mathrm{T}}$ definito sullo spazio probabilizzato $(\Omega, \mathcal{F}, P)$. Si ha*

$$\mathrm{E}[Y] = \mathrm{E}[g(X)] = \int_{\mathbb{R}^n} g(x_1, \ldots, x_n)\, dP_X(x_1, \ldots, x_n)$$

se l'integrale esiste finito.

Dimostrazione. Dal Teorema A.8 si ha

$$\mathrm{E}[Y] = \int g(X)\, dP = \int_{\mathbb{R}^n} g(x_1, \ldots, x_n)\, dP_X(x_1, \ldots, x_n)$$

che è quanto si voleva dimostrare. □

In particolare, se X è un v.v.a. discreto a valori su un insieme S e con f.p.c. p_X, dalla Proposizione 4.2.3 e dal Teorema A.10 si ottiene che

$$\mathrm{E}[Y] = \mathrm{E}[g(X)] = \sum_{(x_1,\ldots,x_n)\in S} g(x_1, \ldots, x_n) p_X(x_1, \ldots, x_n)\,,$$

mentre se X è un v.v.a. assolutamente continuo che ammette d.p.c. f_X allora si ha

$$\mathrm{E}[Y] = \mathrm{E}[g(X)] = \int_{\mathbb{R}^n} g(x_1, \ldots, x_n) f_X(x_1, \ldots, x_n)\, dx_1 \ldots dx_n\,.$$

• **Esempio 4.2.1.** Si consideri la v.a. assolutamente continua X con legge Uniforme che ammette d.p.

$$f_X(x) = \mathbf{1}_{]0,1[}(x)$$

e si consideri la trasformata $Y = X^2$. Dal momento che la funzione $y = x^2$ è crescente in $]0, 1[$, allora dalla Proposizione 3.7.1 risulta

$$f_Y(y) = \frac{1}{2\sqrt{y}}\, \mathbf{1}_{]0,1[}(y)$$

e quindi

$$\mathrm{E}[Y] = \int_0^1 y f_Y(y)\, dy = \frac{1}{2} \int_0^1 \sqrt{y}\, dy = \frac{1}{3}\,.$$

Il medesimo risultato può essere ottenuto in modo immediato applicando direttamente la Proposizione 4.2.3, in quanto si ha

$$\mathrm{E}[Y] = \mathrm{E}[X^2] = \int_0^1 x^2 f_X(x)\, dx = \int_0^1 x^2\, dx = \frac{1}{3}\,.$$ □

• **Esempio 4.2.2.** Si consideri il v.v.a. assolutamente continuo $X = (X_1, X_2)^{\mathrm{T}}$ che ammette la d.p.c. introdotta nell'Esempio 3.7.5. Nel medesimo Esempio è stato verificato che la d.p. della trasformata $Y = X_1 X_2$ risulta

$$f_Y(y) = -\log(y)\, \mathbf{1}_{]0,1[}(y)$$

e quindi

$$\mathrm{E}[Y] = -\int_0^1 y \log(y)\, dy = \frac{1}{4}\,.$$

Tuttavia, applicando la Proposizione 4.2.3, si ha immediatamente

$$\mathrm{E}[Y] = \mathrm{E}[X_1 X_2] = \int_0^1 \int_0^1 x_1 x_2\, dx_1 dx_2 = \frac{1}{4}\,. \qquad \square$$

La seguente Proposizione fornisce un importante risultato sul valore atteso di una combinazione lineare di v.a., nel senso che il valore atteso di una combinazione lineare di v.a. è dato dalla combinazione lineare dei valori attesi.

Proposizione 4.2.4. *Se $X = (X_1, \ldots, X_n)^{\mathrm{T}}$ è un v.v.a. definito sullo spazio probabilizzato $(\Omega, \mathcal{F}, P)$, si consideri la trasformata $Y = a + \sum_{k=1}^n b_k X_k$ dove $a \in \mathbb{R}$ e $(b_1, \ldots, b_n)^{\mathrm{T}} \in \mathbb{R}^n$ sono costanti. Se $\mathrm{E}[X_k]$ esiste finito per ogni $k = 1, \ldots, n$, si ha*

$$\mathrm{E}[Y] = a + \sum_{k=1}^n b_k\, \mathrm{E}[X_k]\,.$$

Dimostrazione. Tenendo presente il Teorema A.6 si ha

$$\mathrm{E}[Y] = \int \left(a + \sum_{k=1}^n b_k X_k \right) dP = a \int dP + b_k \sum_{k=1}^n \int X_k\, dP = a + \sum_{k=1}^n b_k\, \mathrm{E}[X_k]\,,$$

da cui segue la tesi. $\square$

In particolare, quando si considera la trasformata $Y = a + bX$, dove $a, b \in \mathbb{R}$, dalla Proposizione 4.2.4 risulta

$$\mathrm{E}[Y] = a + b\, \mathrm{E}[X]\,.$$

Inoltre, se si ha la trasformata $Y = X - \mu$, con $\mu = \mathrm{E}[X]$, allora segue che

$$\mathrm{E}[Y] = \mathrm{E}[X] - \mu = 0\,.$$

Infine, se si considera la trasformata $Y = \sum_{k=1}^n X_k$, nelle ipotesi della Proposizione 4.2.4 risulta

$$\mathrm{E}[Y] = \sum_{k=1}^n \mathrm{E}[X_k]\,.$$

• **Esempio 4.2.3.** Si consideri il v.v.a. assolutamente continuo $X = (X_1, X_2)^{\mathrm{T}}$ introdotto nell'Esempio 3.5.2. Si ha

$$\mathrm{E}[X_1] = 24 \int_0^1 \int_0^{1-x_1} x_1^2 (1 - x_1 - x_2)\, dx_1 dx_2 = \frac{2}{5}$$

e

$$\mathrm{E}[X_2] = 24 \int_0^1 \int_0^{1-x_1} x_1 x_2 (1 - x_1 - x_2)\, dx_1 dx_2 = \frac{1}{5}\,.$$

Dunque, la media della v.a. $Y = X_1 + X_2$ è data da

$$\mathrm{E}[Y] = \mathrm{E}[X_1 + X_2] = \mathrm{E}[X_1] + \mathrm{E}[X_2] = \frac{3}{5}\,. \qquad \square$$

La seguente Proposizione permette di ottenere il valore atteso di un prodotto di v.a. indipendenti, nel senso che il valore atteso del prodotto di v.a. indipendenti è dato dal prodotto dei valori attesi.

Proposizione 4.2.5. *Se $X = (X_1, \ldots, X_n)^{\mathrm{T}}$ è un v.v.a. definito sullo spazio probabilizzato $(\Omega, \mathcal{F}, P)$ a componenti indipendenti, si consideri la trasformata $Y = \prod_{k=1}^{n} X_k$. Se $\mathrm{E}[X_k]$ esiste finito per ogni $k = 1, \ldots, n$, si ha*

$$\mathrm{E}[Y] = \prod_{k=1}^{n} \mathrm{E}[X_k]\,.$$

Dimostrazione. Tenendo presente la Proposizione 4.2.3 e il Teorema di Fubini (Teorema A.11) si ha

$$\mathrm{E}[Y] = \int \prod_{k=1}^{n} X_k \, dP = \int_{\mathbb{R}^n} \prod_{k=1}^{n} x_k \, d(P_{X_1} \otimes \cdots \otimes P_{X_n}) = \prod_{k=1}^{n} \int_{\mathbb{R}} x_k \, dP_{X_k} = \prod_{k=1}^{n} \mathrm{E}[X_k]\,,$$

che è quanto si voleva dimostrare. $\square$

Si osservi che non vale ovviamente l'inverso della Proposizione 4.2.5, ovvero, se si verifica che il valore atteso di un prodotto di v.a. equivale al prodotto dei singoli valori attesi, questo non ci permette di concludere che le v.a. sono indipendenti.

• **Esempio 4.2.4.** Si consideri il v.v.a. assolutamente continuo $X = (X_1, X_2)^{\mathrm{T}}$ dell'Esempio 3.7.2. Inoltre, nell'Esempio 4.2.2 è stato ottenuto il valore atteso della trasformata $Y = X_1 X_2$. Il calcolo di questa quantità può essere ulteriormente semplificato. In effetti, dal momento che

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

dove $f_{X_1}(x_1) = \mathbf{1}_{]0,1[}(x_1)$ e $f_{X_2}(x_2) = \mathbf{1}_{]0,1[}(x_2)$, allora le componenti marginali del v.v.a. sono indipendenti. Inoltre, si ha $\mathrm{E}[X_1] = \mathrm{E}[X_2] = \frac{1}{2}$ e quindi la media della v.a. $Y = X_1 X_2$ è data da

$$\mathrm{E}[Y] = \mathrm{E}[X_1 X_2] = \mathrm{E}[X_1]\mathrm{E}[X_2] = \frac{1}{4}\,. \qquad \square$$

• **Esempio 4.2.5.** Si consideri il v.v.a. assolutamente continuo $X = (X_1, X_2)^{\mathrm{T}}$ dell'Esempio 3.6.3. Risulta

$$\mathrm{E}[X_1 X_2] = \frac{1}{\pi} \int_{-1}^{1} \int_{-\sqrt{1-x_1^2}}^{\sqrt{1-x_1^2}} x_1 x_2 \, dx_1 dx_2 = 0\,,$$

mentre, data la simmetria rispetto all'origine della d.p. $f_{X_1}(x_1)$, si ha

$$\mathrm{E}[X_1] = \frac{2}{\pi} \int_{-1}^{1} x_1 \sqrt{1 - x_1^2} \, dx_1 = 0\,.$$

Inoltre, dal momento che X_1 e X_2 posseggono la medesima legge, allora si ha anche $\mathrm{E}[X_2] = 0$. Risulta quindi $\mathrm{E}[X_1 X_2] = \mathrm{E}[X_1]\mathrm{E}[X_2]$ anche se le v.a. X_1 e X_2 non sono indipendenti. $\square$

Il seguente Teorema introduce una importante disuguaglianza per il valore atteso di una trasformata $Y = g(X)$ di un v.v.a. X quando la funzione g risulta convessa, ovvero se g è definita su un insieme aperto C e

$$g(\alpha u + (1-\alpha)v) \leq \alpha g(u) + (1-\alpha)g(v)$$

per ogni $u, v \in C$ e dove $\alpha \in [0,1]$. La disuguaglianza prende nome dal matematico danese Johan Jensen (1859-1925).

Teorema 4.2.6. (Disuguaglianza di Jensen) *Data la funzione convessa $g : \mathbb{R}^n \to \mathbb{R}$, si consideri la trasformata $Y = g(X)$ del v.v.a. $X = (X_1, \ldots, X_n)^{\mathrm{T}}$ definito sullo spazio probabilizzato $(\Omega, \mathcal{F}, P)$. Si ha*

$$g(\mathrm{E}[X]) \leq \mathrm{E}[g(X)] \, .$$

Dimostrazione. Se g è una funzione convessa, per ogni v esiste un $c = c(v)$ tale che

$$g(u) \geq g(v) + c(u - v)$$

(si veda Billingsley, 1995, p.545). Tenendo presente questa disuguaglianza con $v = \mathrm{E}[X]$ e le proprietà del valore atteso, si ha

$$\mathrm{E}[g(X)] \geq \mathrm{E}[g(\mathrm{E}[X]) + c(X - \mathrm{E}[X])] = g(\mathrm{E}[X]) + c\,\mathrm{E}[X - \mathrm{E}[X]] = g(\mathrm{E}[X]) \, ,$$

che è quanto si voleva dimostrare. □

• **Esempio 4.2.6.** Si consideri la v.a. X e la trasformata $g(X) = |X|$. Dal momento che g è una funzione convessa, allora risulta $|\mathrm{E}[X]| \leq \mathrm{E}[|X|]$, ovvero si ha una ulteriore dimostrazione della Proposizione 4.2.1, in questo caso basata sulla Disuguaglianza di Jensen. □

4.3. Momenti di variabili aleatorie

I momenti sono medie di particolari trasformate della v.a. in oggetto di studio, ovvero medie di potenze, e che servono a caratterizzare la distribuzione della v.a. Più esattamente si ha la seguente definizione.

Definizione 4.3.1. Data la v.a. X definita sullo spazio probabilizzato $(\Omega, \mathcal{F}, P)$, si dice *momento di ordine r*, dove $r = 0, 1, \ldots$, l'integrale

$$\mathrm{E}[X^r] = \int X^r \, dP$$

se esiste finito. Si dice inoltre *momento assoluto di ordine r* l'integrale

$$\mathrm{E}[|X|^r] = \int |X|^r \, dP$$

se esiste finito. Infine, si dicono rispettivamente *momento centrale di ordine r* l'integrale

$$\mathrm{E}[(X - \mu)^r] = \int (X - \mu)^r \, dP$$

e *momento centrale assoluto di ordine r* l'integrale

$$\mathrm{E}[|X-\mu|^r] = \int |X-\mu|^r \, dP$$

se esistono finiti. □

Il momento di ordine r della v.a. X è spesso indicato con il simbolo $\mu_{X,r}$, o semplicemente con μ_r quando il contesto è chiaro. Dalla Definizione 4.3.1 risulta evidente che $\mu_0 = 1$, mentre si ha $\mu_1 = \mu = \mathrm{E}[X]$. Ovviamente, i momenti e i momenti assoluti (così come i momenti centrali e i momenti centrali assoluti) coincidono quando r è pari. Risulta inoltre immediato verificare che il primo momento centrale è nullo (vedi Sezione 4.2). Il secondo momento centrale è detto *varianza* e viene indicato con la notazione

$$\mathrm{Var}[X] = \mathrm{E}[(X-\mu)^2] \,,$$

o anche con la notazione σ_X^2, che al solito viene ulteriormente semplificata con σ^2 quando non vi è possibilità di fraintendimento. Infine, la radice quadrata della varianza σ è detta *scarto quadratico medio*.

Tenendo presente la Proposizione 4.2.4, è immediato verificare che fra i momenti centrali e i momenti esiste la seguente relazione

$$\mathrm{E}[(X-\mu)^r] = \mathrm{E}[\sum_{k=0}^{r} \binom{r}{k} (-1)^{r-k} \mu^{r-k} X^k] = \sum_{k=0}^{r} \binom{r}{k} (-1)^{r-k} \mu^{r-k} \mu_k \,.$$

In particolare, si ha

$$\mathrm{Var}[X] = \mathrm{E}[X^2] - \mathrm{E}[X]^2 = \mu_2 - \mu^2 \,.$$

Nel linguaggio della Teoria della Misura, i momenti sono evidentemente funzioni di norme, ovvero $\mathrm{E}[|X|^r] = \|X\|_r^r$ e $\mathrm{E}[|X-\mu|^r] = \|X-\mu\|_r^r$.

• **Esempio 4.3.1.** Si consideri la v.a. assolutamente continua X analizzata nell'Esempio 4.2.1. In questo caso il momento di ordine r della v.a. X è dato da

$$\mathrm{E}[X^r] = \int_0^1 x^r \, dx = \frac{1}{r+1} \,.$$

Si ha dunque $\mu = \frac{1}{2}$, mentre il momento di ordine r coincide con il momento assoluto di ordine r essendo $\{X > 0\}$ $q.c.$ Per quanto riguarda il momento centrale di ordine r della v.a., risulta

$$\mathrm{E}[(X-\mu)^r] = \int_0^1 \left(x - \frac{1}{2}\right)^r dx = \frac{2^{-(r+1)}}{r+1} \left(1 + (-1)^r\right)$$

e quindi $\mathrm{E}[(X-\mu)^r] = 0$ se r è dispari, mentre

$$\mathrm{E}[(X-\mu)^r] = \frac{2^{-r}}{r+1}$$

se r è pari. Dunque, si ha $\mathrm{Var}[X] = \frac{1}{12}$. Infine, risulta

$$\mathrm{E}[|X-\mu|^r] = \int_0^1 \left|x - \frac{1}{2}\right|^r dx = \int_0^{\frac{1}{2}} \left(\frac{1}{2} - x\right)^r dx + \int_{\frac{1}{2}}^1 \left(x - \frac{1}{2}\right)^r dx = \frac{2^{-r}}{r+1} \,. \qquad \square$$

• **Esempio 4.3.2.** Si noti che v.a. con leggi differenti possono avere tutti i momenti coincidenti. Come caso specifico si consideri la v.a. assolutamente continua X che ammette densità

$$f_X(x) = \frac{1}{6}\, e^{-x^{1/3}}\, \mathbf{1}_{[0,\infty[}(x)$$

e la v.a. assolutamente continua Y che ammette densità

$$f_Y(y) = \frac{1}{6}\, e^{-y^{1/3}}\, (1 + \sin(\sqrt{3}\, y^{1/3}))\, \mathbf{1}_{[0,\infty[}(y)\ .$$

Incidentalmente, si osservi che la legge associata alla v.a. $Z = X^{1/3}$ si ottiene per una particolare parametrizzazione della legge Gamma (si veda la Sezione 6.8). Dal momento che risulta

$$\int_0^\infty y^r e^{-y^{1/3}} \sin(\sqrt{3}\, y^{1/3})\, dy = 0\ ,$$

allora segue immediatamente che $\mathrm{E}[X^r] = \mathrm{E}[Y^r]$ per ogni $r = 0, 1, \ldots$. □

Il seguente Teorema fornisce una importante disuguaglianza, che nella successiva Proposizione permette di ottenere le condizioni di esistenza per i momenti. La disuguaglianza è comunemente attribuita al matematico e probabilista russo Aleksandr Mikhailovich Lyapunov (1857-1918).

Figura 4.3.1. Aleksandr Mikhailovich Lyapunov (1857-1918).

Teorema 4.3.2. (Disuguaglianza di Lyapunov) *Si consideri la v.a. X definita sullo spazio probabilizzato $(\Omega, \mathcal{F}, P)$. Se $0 < s \leq r$, allora si ha*

$$\mathrm{E}[|X|^s]^{1/s} \leq \mathrm{E}[|X|^r]^{1/r}\ .$$

Dimostrazione. Se $\{Y \geq 0\}$ *q.c.*, dalla disuguaglianza di Jensen (Teorema 4.2.6) con $g(y) = y^a$ e $a \geq 1$, si ha $\mathrm{E}[Y]^a \leq \mathrm{E}[Y^a]$. Dal momento che $\frac{r}{s} \geq 1$, ponendo $a = \frac{r}{s}$ e $Y = |X|^s$ nella precedente disuguaglianza, si ottiene

$$\mathrm{E}[|X|^s]^{r/s} \leq \mathrm{E}[|X|^r]\ ,$$

da cui si ha immediatamente la tesi. □

Nel linguaggio della Teoria della Misura, la disuguaglianza di Lyapunov stabilisce in effetti che $\|X\|_s \leq \|X\|_r$ se $0 < s \leq r$.

Proposizione 4.3.3. *Si consideri la v.a. X definita sullo spazio probabilizzato $(\Omega, \mathcal{F}, P)$. Se si ha $\mathrm{E}[|X|^r] < \infty$, allora $\mathrm{E}[X^k]$ esiste finito per ogni $k = 1, \ldots, r$.*

Dimostrazione. Dal momento che $k \leq r$, dalla disuguaglianza di Lyapunov (Teorema 4.3.2) si ha

$$\mathrm{E}[|X|^k]^{1/k} \leq \mathrm{E}[|X|^r]^{1/r} .$$

Dunque, se $\mathrm{E}[|X|^r] < \infty$, allora si ha $\mathrm{E}[|X|^k] < \infty$. Tuttavia, dalla Proposizione 4.2.1 risulta che se $\mathrm{E}[|X|^k] < \infty$, allora si ha anche $\mathrm{E}[X^k] < \infty$. □

Tenendo presente la dimostrazione della precedente Proposizione, è inoltre facile verificare che se il momento di ordine r non esiste finito, allora non esistono finiti neppure i momenti di ordine superiore. Inoltre, dal momento che $\mathrm{Var}[X] \leq \mathrm{E}[X^2]$, la varianza esiste finita se esiste finito il secondo momento. In generale, è immediato verificare che il momento centrale di ordine r esiste finito se esiste finito il momento di ordine r.

Si considerano di seguito alcune Proposizioni sulle proprietà della varianza.

Proposizione 4.3.4. *Si consideri la v.a. X definita sullo spazio probabilizzato $(\Omega, \mathcal{F}, P)$. Se $\mathrm{E}[X^2]$ esiste finito, si ha*

$$\mathrm{Var}[X] \leq \min_{a \in \mathbb{R}} \mathrm{E}[(X-a)^2] .$$

Dimostrazione. Tenendo presente le proprietà del valore atteso, si ha

$$\begin{aligned} \mathrm{E}[(X-a)^2] &= \mathrm{E}[(X-\mu+\mu-a)^2] = \mathrm{E}[(X-\mu)^2 + 2(X-\mu)(\mu-a) + (\mu-a)^2] \\ &= \mathrm{Var}[X] + (\mu-a)^2 \end{aligned}$$

e, dal momento che $\mathrm{Var}[X] \geq 0$ non dipende da a, il minimo della precedente quantità si ottiene per $a = \mu$. □

Proposizione 4.3.5. *Si consideri la v.a. X definita sullo spazio probabilizzato $(\Omega, \mathcal{F}, P)$. Se $\mathrm{E}[X^2]$ esiste finito, la varianza della trasformata $Y = a + bX$, dove $a, b \in \mathbb{R}$, è data da*

$$\mathrm{Var}[Y] = \mathrm{Var}[a + bX] = b^2 \, \mathrm{Var}[X] .$$

Dimostrazione. Tenendo presente la Proposizione 4.2.4 si ha $\mathrm{E}[Y] = a + b\mu$, e quindi

$$\mathrm{Var}[Y] = \mathrm{E}[(a + bX - a - b\mu)^2] = b^2 \, \mathrm{E}[(X-\mu)^2] = b^2 \, \mathrm{Var}[X] ,$$

che è quanto si voleva dimostrare. □

In particolare, se $Y = a + X$, dalla Proposizione 4.3.5 si ha

$$\mathrm{Var}[Y] = \mathrm{Var}[a + X] = \mathrm{Var}[X] .$$

Inoltre, se si considera la trasformazione

$$Z = \frac{X - \mu}{\sigma} ,$$

allora risulta

$$\mathrm{E}[Z] = \frac{1}{\sigma} \mathrm{E}[X - \mu] = 0 ,$$

mentre

$$\mathrm{Var}[Z] = \frac{1}{\sigma^2} \mathrm{Var}[X - \mu] = \frac{1}{\sigma^2} \mathrm{Var}[X] = 1 .$$

Per questi motivi, la precedente è detta *trasformazione di standardizzazione*.

Il seguente Teorema fornisce una importante e utile disuguaglianza, che prende nome dal matematico e probabilista russo Andrey Andreyevich Markov (1856-1922).

Teorema 4.3.6. (Disuguaglianza di Markov) *Si consideri la v.a. X definita sullo spazio probabilizzato $(\Omega, \mathcal{F}, P)$. Se $c > 0$, allora si ha*

$$P(|X| \geq c) \leq \frac{1}{c} \mathrm{E}[|X|] \, .$$

Dimostrazione. Si ha

$$\mathrm{E}[|X|] = \int |X| \, dP \geq \int c \mathbf{1}_{\{|X| \geq c\}} \, dP = cP(|X| \geq c) \, ,$$

da cui segue immediatamente la tesi. □

Figura 4.3.2. Andrey Andreyevich Markov (1856-1922).

Il seguente Teorema fornisce una famosa e celebrata disuguaglianza introdotta dal matematico russo Pafnuty Lvovich Chebyshev (1821-1894), padre fondatore della scuola matematica russa e in particolare mentore di Lyapunov e Markov.

Teorema 4.3.7. (Disuguaglianza di Chebyshev) *Si consideri la v.a. X definita sullo spazio probabilizzato $(\Omega, \mathcal{F}, P)$. Se $c > 0$, allora si ha*

$$P(|X - \mu| \geq c) \leq \frac{\sigma^2}{c^2} \, .$$

Equivalentemente, se $\sigma^2 < \infty$ si ha

$$P\left(\left| \frac{X - \mu}{\sigma} \right| \geq c \right) \leq \frac{1}{c^2} \, .$$

Dimostrazione. Applicando opportunamente la disuguaglianza di Markov (Teorema 4.3.6) alla v.a. trasformata $(X - \mu)^2$, si ha

$$P(|X - \mu| \geq c) = P((X - \mu)^2 \geq c^2) \leq \frac{1}{c^2} \mathrm{E}[(X - \mu)^2] \, ,$$

da cui segue la prima parte del Teorema. La seconda parte si ottiene in modo simile considerando la v.a. trasformata $\frac{(X-\mu)^2}{\sigma^2}$. □

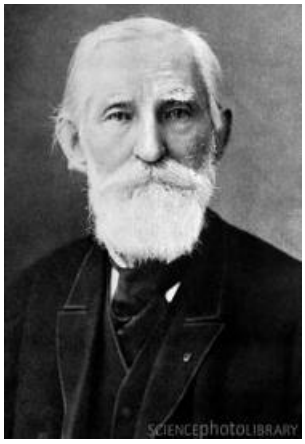


Figura 4.3.3. Pafnuty Lvovich Chebyshev (1821-1894).

La precedente disuguaglianza è anche detta di Bienaymé-Chebyshev dal momento che in effetti fu congiuntamente formulata da Chebyshev insieme allo statistico francese Irénée-Jules Bienaymé (1796-1878). Si noti inoltre che la disuguaglianza può essere facilmente generalizzata con i momenti centrali assoluti, ovvero risulta

$$P(|X-\mu| \geq c) \leq \frac{1}{c^r}\,\mathrm{E}[|X-\mu|^r]\,.$$

Proposizione 4.3.8. *Se X è una v.a. definita sullo spazio probabilizzato $(\Omega, \mathcal{F}, P)$ e se* $\mathrm{E}[X^2]$ *esiste finito, si ha* $\mathrm{Var}[X] = 0$ *se e solo se X è degenere.*

Dimostrazione. Se X è degenere, allora risulta $\{X = a\}$ $q.c.$ per una data costante $a \in \mathbb{R}$. In questo caso, si ha

$$\mathrm{E}[X] = aP(X=a) = a\,,$$

mentre

$$\mathrm{Var}[X] = (a - \mathrm{E}[X])^2 P(X=a) = 0\,.$$

Inversamente, se $\mathrm{Var}[X] = 0$, applicando la disuguaglianza di Chebyshev, per ogni $c > 0$ si ha

$$P(|X-a| \geq c) \leq \frac{\sigma^2}{c^2} = 0\,,$$

ovvero $\{X = a\}$ $q.c.$ □

4.4. Covarianza e matrice di varianza-covarianza

Il concetto di momento può essere esteso in modo generale a un v.v.a. In effetti, si ha la seguente definizione formale.

Definizione 4.4.1. Dato il v.v.a. $X = (X_1, \ldots, X_n)^{\mathrm{T}}$ definito sullo spazio probabilizzato $(\Omega, \mathcal{F}, P)$, si dice *momento misto di ordine* $(r_1, \ldots, r_n)$, dove $r_k = 0, 1, \ldots$ e $k = 1, \ldots, n$, l'integrale

$$\mathrm{E}[X_1^{r_1} \ldots X_n^{r_n}] = \int X_1^{r_1} \ldots X_n^{r_n}\, dP$$

se esiste finito. Inoltre, si dice *momento centrale misto di ordine* $(r_1, \ldots, r_n)$ l'integrale

$$\mathrm{E}[(X_1 - \mu_{X_1})^{r_1} \dots (X_n - \mu_{X_n})^{r_n}] = \int (X_1 - \mu_{X_1})^{r_1} \dots (X_n - \mu_{X_n})^{r_n} \, dP$$

se esiste finito. □

Si osservi che, scegliendo opportunamente gli indici $(r_1, \dots, r_n)$, si possono ottenere i momenti di qualsiasi ordine di tutte le possibili scelte delle componenti marginali. Ad esempio, ponendo $r_k = 1$ e $r_l = 0$ per ogni $l \neq k = 1, \dots, n$, si ottiene $\mathrm{E}[X_k]$. In particolare, risulta fondamentale considerare i momenti per vettori bivariati di v.a. Più esattamente, dato il v.v.a. $X = (X_1, X_2)^{\mathrm{T}}$, il *momento misto* è dato dall'integrale

$$\mathrm{E}[X_1 X_2] = \int X_1 X_2 \, dP$$

se esiste finito. Analogamente, il *momento misto centrale*, detto usualmente *covarianza*, è dato dall'integrale

$$\mathrm{Cov}[X_1, X_2] = \mathrm{E}[(X_1 - \mu_{X_1})(X_2 - \mu_{X_2})] = \int (X_1 - \mu_{X_1})(X_2 - \mu_{X_2}) \, dP$$

se esiste finito. La covarianza viene anche usualmente denotata con il simbolo σ_{X_1,X_2}. Tenendo presente il Teorema A.6, risulta immediato verificare che

$$\mathrm{Cov}[X_1, X_2] = \mathrm{E}[X_1 X_2] - \mathrm{E}[X_1]\mathrm{E}[X_2] \, ,$$

mentre ovviamente $\mathrm{Cov}[X_1, X_1] = \mathrm{Var}[X_1]$. In Teoria della Misura, il momento misto e la covarianza sono prodotti interni, ovvero $\mathrm{E}[X_1 X_2] = \langle X_1, X_2 \rangle$ e $\mathrm{Cov}[X_1 X_2] = \langle X_1 - \mu_{X_1}, X_2 - \mu_{X_2} \rangle$.

Il seguente Teorema introduce una famosa disuguaglianza che permette fra l'altro di ottenere le condizioni di esistenza della covarianza e che prende nome dal matematico tedesco Otto Ludwig Hölder (1859-1937). Tuttavia, la disuguaglianza dovrebbe più correttamente essere denominata di Rogers-Hölder, dal momento che è stata introdotta indipendentemente e contemporaneamente anche dal matematico inglese Leonard James Rogers (1862-1933).

Teorema 4.4.2. (Disuguaglianza di Hölder) *Sia dato un v.v.a.* $X = (X_1, X_2)^{\mathrm{T}}$ *definito sullo spazio probabilizzato* $(\Omega, \mathcal{F}, P)$. *Se* $r, s \in]1, \infty[$ *sono costanti tali che* $\frac{1}{r} + \frac{1}{s} = 1$ *e se* $\mathrm{E}[|X_1|^r] < \infty$ *e* $\mathrm{E}[|X_2|^s] < \infty$, *allora*

$$|\mathrm{E}[X_1 X_2]| \leq \mathrm{E}[|X_1 X_2|] \leq \mathrm{E}[|X_1|^r]^{1/r} \mathrm{E}[|X_2|^s]^{1/s} \, .$$

Inoltre, l'uguaglianza si ottiene se e solo se $\{b_1 X_1 = b_2 X_2\}$ *q.c.*

Dimostrazione. La prima disuguaglianza segue dalla Proposizione 4.2.1. Al fine di dimostrare la seconda disuguaglianza, si noti che la funzione $x \mapsto -\log(x)$ è convessa. Dunque, se $\alpha \in [0, 1]$, dalla definizione di funzione convessa si ha

$$-\log(\alpha u + (1 - \alpha)v) \leq -\alpha \log(u) - (1 - \alpha) \log(v) = -\log(u^{\alpha} v^{1-\alpha})$$

per ogni $u, v \in]0, \infty[$. Quindi, dalla precedente disuguaglianza si ottiene

$$u^{\alpha} v^{1-\alpha} \leq \alpha u + (1 - \alpha)v \, .$$

Infine, dal momento che la funzione $x \mapsto -\log(x)$ è strettamente convessa, l'uguaglianza si ha se e solo se $u = v$. Adoperando dunque la precedente disuguaglianza con $\alpha = \frac{1}{r}$ e $1 - \alpha = \frac{1}{s}$, posto

$$Y_1 = \frac{|X_1|^r}{\mathrm{E}[|X_1|^r]}$$

e

$$Y_2 = \frac{|X_2|^s}{\mathrm{E}[|X_2|^s]} ,$$

tenendo presente che $\mathrm{E}[Y_1] = \mathrm{E}[Y_2] = 1$, allora si ha

$$\mathrm{E}[Y_1^{1/r} Y_2^{1/s}] \leq \frac{1}{r}\,\mathrm{E}[Y_1] + \frac{1}{s}\,\mathrm{E}[Y_2] = \frac{1}{r} + \frac{1}{s} = 1 .$$

Infine, sostituendo opportunamente, dalle proprietà del valore atteso si ottiene immediatamente la prima parte del Teorema. Inoltre, l'uguaglianza si ottiene se e solo se $\{Y_1 = Y_2\}$ $q.c.$, ovvero, tenendo presente che

$$\frac{|X_1|^r}{\mathrm{E}[|X_1|^r]} = \frac{|b_1 X_1|^r}{\mathrm{E}[|b_1 X_1|^r]}$$

e

$$\frac{|X_2|^s}{\mathrm{E}[|X_2|^s]} = \frac{|b_2 X_2|^s}{\mathrm{E}[|b_2 X_2|^s]} ,$$

quando si ha $\{b_1 X_1 = b_2 X_2\}$ $q.c.$ □

Nel caso particolare in cui $r = s = 2$, la disuguaglianza si riduce a

$$\mathrm{E}[X_1 X_2]^2 \leq \mathrm{E}[X_1^2]\mathrm{E}[X_2^2] ,$$

e viene detta disuguaglianza di Schwarz in onore del matematico tedesco Karl Hermann Amandus Schwarz (1843-1921). Anche in questo caso si dovrebbe parlare di disuguaglianza di Cauchy-Bunyakovsky-Schwarz, in quanto la versione elementare della disuguaglianza è stata introdotta da Cauchy e successivamente estesa dal matematico russo Viktor Yakovlevich Bunyakovsky (1804-1889). La disuguaglianza di Schwarz implica inoltre che $\mathrm{Cov}[X_1, X_2]$ esiste finita se esistono finiti i momenti $\mathrm{E}[X_1^2]$ e $\mathrm{E}[X_2^2]$. Infine, si noti che in Teoria della Misura la disuguaglianza di Hölder stabilisce che $|\langle X_1, X_2 \rangle| \leq \|X_1\|_s \|X_2\|_r$ se $\frac{1}{r} + \frac{1}{s} = 1$.

Il seguente Teorema introduce un'altra celebrata disuguaglianza che discende dalla disuguaglianza di Hölder e che prende nome dal matematico tedesco Hermann Minkowski (1864-1909).

Teorema 4.4.3. (Disuguaglianza di Minkowski) *Sia dato un v.v.a.* $X = (X_1, X_2)^{\mathrm{T}}$ *definito sullo spazio probabilizzato* $(\Omega, \mathcal{F}, P)$. *Se* $r \geq 1$ *e se* $\mathrm{E}[|X_1|^r] < \infty$ *e* $\mathrm{E}[|X_2|^r] < \infty$, *allora*

$$\mathrm{E}[|X_1 + X_2|^r]^{1/r} \leq \mathrm{E}[|X_1|^r]^{1/r} + \mathrm{E}[|X_2|^r]^{1/r} .$$

Dimostrazione. Se $\mathrm{E}[|X_1 + X_2|^r] = 0$ la disuguaglianza è immediatamente verificata. Si consideri quindi il caso in cui $\mathrm{E}[|X_1 + X_2|^r] > 0$. Tenendo presente che per ogni $u, v \in \mathbb{R}$ vale la disuguaglianza $|u + v| \leq |u| + |v|$, si ha

$$\begin{aligned} \mathrm{E}[|X_1 + X_2|^r] &= \mathrm{E}[|X_1 + X_2| \cdot |X_1 + X_2|^{r-1}] \leq \mathrm{E}[(|X_1| + |X_2|) \cdot |X_1 + X_2|^{r-1}] \\ &= \mathrm{E}[|X_1| \cdot |X_1 + X_2|^{r-1}] + \mathrm{E}[|X_2| \cdot |X_1 + X_2|^{r-1}] . \end{aligned}$$

Inoltre, dalla disuguaglianza di Hölder applicata alle v.a. $|X_1|$ e $|X_1 + X_2|^{r-1}$ risulta

$$\mathrm{E}[|X_1| \cdot |X_1 + X_2|^{r-1}] \leq \mathrm{E}[|X_1|^r]^{1/r} \mathrm{E}[|X_1 + X_2|^{(r-1)s}]^{1/s} = \mathrm{E}[|X_1|^r]^{1/r} \mathrm{E}[|X_1 + X_2|^r]^{1-1/r} ,$$

dal momento che, essendo $\frac{1}{r} + \frac{1}{s} = 1$, si ha anche $(r-1)s = r$ e $\frac{1}{s} = 1 - \frac{1}{r}$. In modo simile risulta

$$\mathrm{E}[|X_2| \cdot |X_1 + X_2|^{r-1}] \leq \mathrm{E}[|X_2|^r]^{1/r} \mathrm{E}[|X_1 + X_2|^r]^{1-1/r} .$$

Dunque, si ha

$$\mathrm{E}[|X_1 + X_2|^r] \leq (\mathrm{E}[|X_1|^r]^{1/r} + \mathrm{E}[|X_2|^r]^{1/r}) \mathrm{E}[|X_1 + X_2|^r]^{1-1/r} ,$$

da cui discende immediatamente la tesi. □

Nella Teoria della Misura, la disuguaglianza di Minkowski stabilisce in effetti la cosiddetta disuguaglianza triangolare, ovvero $\|X_1 + X_2\|_r \leq \|X_1\|_r + \|X_2\|_r$.

Si considerano di seguito alcune Proposizioni sulle proprietà della covarianza.

Proposizione 4.4.4. *Dato un v.v.a.* $X = (X_1, X_2)^{\mathrm{T}}$ *definito sullo spazio probabilizzato* $(\Omega, \mathcal{F}, P)$*, se* X_1 *e* X_2 *sono indipendenti e se* $\mathrm{E}[X_1]$ *e* $\mathrm{E}[X_2]$ *esistono finiti, allora si ha* $\mathrm{Cov}[X_1, X_2] = 0$.

Dimostrazione. Tenendo presente la Proposizione 4.2.5 si ha

$$\mathrm{E}[X_1 X_2] = \mathrm{E}[X_1]\mathrm{E}[X_2] ,$$

da cui segue immediatamente la tesi. □

Si deve notare che non è in generale valido l'inverso della Proposizione 4.4.4, ovvero se si verifica $\mathrm{Cov}[X_1, X_2] = 0$ non si può concludere che le due componenti del v.v.a. $X = (X_1, X_2)^{\mathrm{T}}$ sono indipendenti. In effetti, come sarà visto in seguito, la covarianza dipende dal'intensità del legame lineare tra le due componenti del v.v.a. X. Se tali componenti sono indipendenti, non esistendo tra loro nessun tipo di legame, non esiste nemmeno un legame di tipo lineare e quindi la covarianza risulta nulla. Al contrario, se le due componenti hanno covarianza nulla, si può solamente affermare che non esiste tra loro nessun legame di tipo lineare, ma questo non esclude che esista un legame di tipo differente.

• **Esempio 4.4.1.** Si consideri il v.v.a. assolutamente continuo $X = (X_1, X_2)^{\mathrm{T}}$. Si supponga che la componente marginale X_1 ammetta d.p. data da

$$f_{X_1}(x_1) = \frac{1}{2} \mathbf{1}_{]-1,1[}(x_1)$$

e che $X_2 = X_1^2$. Dal momento che

$$\mathrm{E}[X_1^r] = \frac{1}{2} \int_{-1}^{1} x^r \, dx = \frac{1}{2(r+1)} \left(1 + (-1)^r\right)$$

e quindi $\mathrm{E}[X_1^r] = 0$ se r è dispari, mentre

$$\mathrm{E}[X_1^r] = \frac{1}{r+1}$$

se r è pari. Dunque, si ha

$$\mathrm{Cov}[X_1, X_2] = \mathrm{E}[X_1^3] - \mathrm{E}[X_1]\mathrm{E}[X_1^2] = 0$$

anche se X_2 è una trasformata di X_1 e dunque X_1 e X_2 non risultano ovviamente indipendenti. □

Proposizione 4.4.5. *Dato un v.v.a.* $X = (X_1, X_2)^{\mathrm{T}}$ *definito sullo spazio probabilizzato* $(\Omega, \mathcal{F}, P)$*, se* $Y_1 = a_1 + b_1 X_1$ *e* $Y_2 = a_2 + b_2 X_2$*, dove* $a_1, a_2, b_1, b_2 \in \mathbb{R}$ *sono costanti, e se* $\mathrm{Cov}[X_1, X_2]$ *esiste finita, allora*

$$\mathrm{Cov}[Y_1, Y_2] = \mathrm{Cov}[a_1 + b_1 X_1, a_2 + b_2 X_2] = b_1 b_2 \,\mathrm{Cov}[X_1, X_2] \,.$$

Dimostrazione. Tenendo presente la Proposizione 4.2.4 si ha

$$\begin{aligned}\mathrm{Cov}[Y_1, Y_2] &= \mathrm{E}[(a_1 + b_1 X_1 - \mathrm{E}[a_1 + b_1 X_1])(a_2 + b_2 X_2 - \mathrm{E}[a_2 + b_2 X_2])] \\ &= b_1 b_2 \,\mathrm{E}[(X_1 - \mathrm{E}[X_1])(X_2 - \mathrm{E}[X_2])] = b_1 b_2 \,\mathrm{Cov}[X_1, X_2] \,,\end{aligned}$$

che è quanto si voleva dimostrare. □

In particolare, per $Y_1 = a_1 + X_1$ e $Y_2 = a_2 + X_2$, dalla Proposizione 4.4.5 si ha

$$\mathrm{Cov}[Y_1, Y_2] = \mathrm{Cov}[a_1 + X_1, a_2 + X_2] = \mathrm{Cov}[X_1, X_2] \,.$$

Viene introdotto di seguito un importante indice che risulta fondamentale nel valutare il grado di dipendenza lineare tra due v.a. Più esattamente, dato il v.v.a. $X = (X_1, X_2)^{\mathrm{T}}$ definito sullo spazio probabilizzato $(\Omega, \mathcal{F}, P)$, se X_1 e X_2 sono v.a. non degeneri tali che $\mathrm{E}[X_1^2]$ e $\mathrm{E}[X_2^2]$ esistono finiti, si dice *coefficiente di correlazione* tra X_1 e X_2 la quantità

$$\rho_{X_1,X_2} = \frac{\sigma_{X_1,X_2}}{\sigma_{X_1}\sigma_{X_2}} \,.$$

Tenendo presente la Proposizione 4.4.5 e la Proposizione 4.3.5, se si considera le v.a. standardizzate $Z_1 = \frac{X_1 - \mu_{X_1}}{\sigma_{X_1}}$ e $Z_2 = \frac{X_2 - \mu_{X_2}}{\sigma_{X_2}}$, è immediato verificare che

$$\mathrm{E}[Z_1 Z_2] = \sigma_{Z_1,Z_2} = \rho_{Z_1,Z_2} = \rho_{X_1,X_2} \,,$$

ovvero il coefficiente di correlazione rappresenta in pratica la covarianza tra le v.a. standardizzate. Inoltre, quando $\rho_{X_1,X_2} = 0$ le v.a. sono dette *non correlate*. Ovviamente, questo si verifica se e solo se $\sigma_{X_1,X_2} = 0$. Dalle seguenti Proposizioni risultano le principali proprietà del coefficiente di correlazione.

Proposizione 4.4.6. *Dato un v.v.a. $X = (X_1, X_2)^{\mathrm{T}}$ definito sullo spazio probabilizzato $(\Omega, \mathcal{F}, P)$, se $Y_1 = a_1 + b_1 X_1$ e $Y_2 = a_2 + b_2 X_2$, dove $a_1, a_2, b_1, b_2 \in \mathbb{R}$, e se X_1 e X_2 sono v.a. non degeneri tali che $\mathrm{E}[X_1^2]$ e $\mathrm{E}[X_2^2]$ esistono finiti, allora*

$$\rho_{Y_1,Y_2} = \mathrm{sgn}(b_1 b_2)\rho_{X_1,X_2} \,.$$

Dimostrazione. Tenendo presente la Proposizione 4.4.5 e la Proposizione 4.3.5 si ha

$$\rho_{Y_1,Y_2} = \frac{b_1 b_2 \sigma_{X_1,X_2}}{|b_1|\sigma_{X_1}|b_2|\sigma_{X_1}} = \mathrm{sgn}(b_1 b_2)\rho_{X_1,X_2} \,.$$ □

Proposizione 4.4.7. *Dato un v.v.a. $X = (X_1, X_2)^{\mathrm{T}}$ definito sullo spazio probabilizzato $(\Omega, \mathcal{F}, P)$, se X_1 e X_2 sono v.a. non degeneri tali che $\mathrm{E}[X_1^2]$ e $\mathrm{E}[X_2^2]$ esistono finiti, si ha $|\rho_{X_1,X_2}| \leq 1$. Inoltre, $|\rho_{X_1,X_2}| = 1$ se e solo se $\{X_2 = a + bX_1\}$ q.c. o $\{X_2 = -a - bX_1\}$ q.c. dove*

$$a = \mu_{X_2} - \mu_{X_1}\frac{\sigma_{X_2}}{\sigma_{X_1}}$$

e

$$b = \frac{\sigma_{X_2}}{\sigma_{X_1}} \,.$$

Dimostrazione. Se si considera le v.a. $Z_1 = \frac{X_1 - \mu_{X_1}}{\sigma_{X_1}}$ e $Z_2 = \frac{X_2 - \mu_{X_2}}{\sigma_{X_2}}$, dal momento che $\mathrm{E}[Z_1^2] = \sigma_{Y_1}^2 = 1$ e $\mathrm{E}[Z_2^2] = \sigma_{Y_2}^2 = 1$, dalla disuguaglianza di Schwarz si ha $\mathrm{E}[Z_1 Z_2]^2 \leq 1$. La prima parte della Proposizione è verificata dal momento che $\mathrm{E}[Z_1 Z_2] = \rho_{X_1,X_2}$. Tenendo presente la dimostrazione del Teorema 4.4.2, l'uguaglianza si ha se e solo se $\{Z_1 = Z_2\}$ $q.c.$ o $\{Z_1 = -Z_2\}$ $q.c$, da cui segue la seconda parte della Proposizione. □

• **Esempio 4.4.2.** Si consideri il v.v.a. assolutamente continuo $X = (X_1, X_2)^{\mathrm{T}}$ dell'Esempio 3.5.2. Si ha

$$\mathrm{E}[X_1 X_2] = 24 \int_0^1 \int_0^{1-x_1} x_1^2 x_2 (1 - x_1 - x_2)\, dx_1 dx_2 = \frac{1}{15}\,.$$

Inoltre, è stato visto nell'Esempio 4.2.3 che $\mathrm{E}[X_1] = \frac{2}{5}$ e $\mathrm{E}[X_2] = \frac{1}{5}$, mentre risulta

$$\mathrm{E}[X_1^2] = 24 \int_0^1 \int_0^{1-x_1} x_1^3 (1 - x_1 - x_2)\, dx_1 dx_2 = \frac{1}{5}\,,$$

e

$$\mathrm{E}[X_2^2] = 24 \int_0^1 \int_0^{1-x_1} x_1 x_2^2 (1 - x_1 - x_2)\, dx_1 dx_2 = \frac{1}{15}\,,$$

Dunque, si ottiene $\sigma_{X_1,X_2} = -\frac{1}{75}$, $\sigma_{X_1}^2 = \frac{1}{25}$ e $\sigma_{X_2}^2 = \frac{2}{75}$, per cui risulta $\rho_{X_1,X_2} = \frac{\sqrt{6}}{6}$. □

In generale, considerato il v.v.a. $X = (X_1, \ldots, X_n)^{\mathrm{T}}$ definito sullo spazio probabilizzato $(\Omega, \mathcal{F}, P)$, se $\mathrm{E}[X_k]$ esiste finito per ogni $k = 1, \ldots, n$, il *vettore delle medie* è definito come $\mu_X = (\mu_{X_1}, \ldots, \mu_{X_n})^{\mathrm{T}}$, mentre se $\mathrm{Cov}[X_k, X_j]$ esiste finito per ogni $k = j = 1, \ldots, n$, la *matrice di varianza-covarianza* è definita come

$$\Sigma_X = \begin{pmatrix} \sigma_{X_1}^2 & \sigma_{X_1,X_2} & \cdots & \sigma_{X_1,X_n} \\ \sigma_{X_2,X_1} & \sigma_{X_2}^2 & \cdots & \sigma_{X_2,X_n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{X_n,X_1} & \sigma_{X_n,X_2} & \cdots & \sigma_{X_n}^2 \end{pmatrix}.$$

Il determinante $\det(\Sigma_X)$ è detto *varianza generalizzata* del v.v.a. X. Al solito, queste notazioni vengono ulteriormente semplificate con μ e Σ quando non vi è possibilità di fraintendimento.

• **Esempio 4.4.3.** Si consideri il v.v.a. assolutamente continuo $X = (X_1, X_2, X_3)^{\mathrm{T}}$ che ammette la seguente d.p.c.

$$f_X(x_1, x_2, x_3) = (1 + (1 - 2x_1)(1 - 2x_2))\, \mathbf{1}_{]0,1[}(x_1) \mathbf{1}_{]0,1[}(x_2) \mathbf{1}_{]0,1[}(x_3)\,.$$

Si ha

$$\mathrm{E}[X_1 X_2] = \int_0^1 \int_0^1 \int_0^1 x_1 x_2 (1 + (1 - 2x_1)(1 - 2x_2))\, dx_1 dx_2 dx_3 = \frac{5}{18}\,,$$

mentre

$$\mathrm{E}[X_1] = \int_0^1 \int_0^1 \int_0^1 x_1 (1 + (1 - 2x_1)(1 - 2x_2))\, dx_1 dx_2 dx_3 = \frac{1}{2}$$

e

$$\mathrm{E}[X_1^2] = \int_0^1 \int_0^1 \int_0^1 x_1^2(1 + (1-2x_1)(1-2x_2))\, dx_1 dx_2 dx_3 = \frac{1}{3}\,.$$

Inoltre, per simmetria si ha $\mathrm{E}[X_2] = \mathrm{E}[X_1]$ e $\mathrm{E}[X_2^2] = \mathrm{E}[X_1^2]$. Si noti che

$$f_X(x_1, x_2, x_3) = f_{X_1,X_2}(x_1, x_2) f_{X_3}(x_3)\,,$$

dove

$$f_{X_1,X_2}(x_1, x_2) = (1 + (1-2x_1)(1-2x_2))\, \mathbf{1}_{]0,1[}(x_1)\mathbf{1}_{]0,1[}(x_2)$$

e

$$f_{X_3}(x_3) = \mathbf{1}_{]0,1[}(x_3)\,.$$

Quindi, il v.v.a. $(X_1, X_2)^{\mathrm{T}}$ e la v.a. X_3 sono indipendenti. Inoltre, si ha $\mathrm{E}[X_3] = \frac{1}{2}$ e $\mathrm{E}[X_3^2] = \frac{1}{3}$. Dunque, si ha infine che $\mu = (\frac{1}{2}, \frac{1}{2}, \frac{1}{2})^{\mathrm{T}}$, mentre

$$\Sigma = \begin{pmatrix} \sigma_{X_1}^2 & \sigma_{X_1,X_2} & \sigma_{X_1,X_3} \\ \sigma_{X_2,X_1} & \sigma_{X_2}^2 & \sigma_{X_2,X_3} \\ \sigma_{X_3,X_1} & \sigma_{X_3,X_2} & \sigma_{X_3}^2 \end{pmatrix} = \begin{pmatrix} \frac{1}{12} & \frac{1}{36} & 0 \\ \frac{1}{36} & \frac{1}{12} & 0 \\ 0 & 0 & \frac{1}{12} \end{pmatrix}. \qquad \square$$

La seguente Proposizione fornisce un importante risultato sulla varianza di una combinazione lineare di v.a.

Proposizione 4.4.8. *Se $X = (X_1, \ldots, X_n)^{\mathrm{T}}$ è un v.v.a. definito sullo spazio probabilizzato $(\Omega, \mathcal{F}, P)$ con matrice di varianza-covarianza data da Σ, si consideri la trasformata $Y = a + \sum_{k=1}^n b_k X_k$ dove $a \in \mathbb{R}$ e $b = (b_1, \ldots, b_n)^{\mathrm{T}} \in \mathbb{R}^n$ sono costanti. Se $\mathrm{Cov}[X_k, X_j]$ esiste finita per ogni $k = 1, \ldots, n$, si ha*

$$\mathrm{Var}[Y] = b^{\mathrm{T}} \Sigma b\,.$$

Dimostrazione. Tenendo presente il Teorema A.6 e la Proposizione 4.2.4, risulta

$$\begin{aligned} \mathrm{Var}[Y] &= \mathrm{E}[(a + \sum_{k=1}^n b_k X_k - a - \sum_{k=1}^n b_k \mathrm{E}[X_k])^2] = \mathrm{E}[(\sum_{k=1}^n b_k (X_k - \mathrm{E}[X_k]))^2] \\ &= \mathrm{E}[\sum_{k=1}^n b_k^2 (X_k - \mathrm{E}[X_k])^2 + \sum_{k=1}^n \sum_{j \neq k=1}^n b_k b_j (X_k - \mathrm{E}[X_k])(X_j - \mathrm{E}[X_j])] \\ &= \sum_{k=1}^n b_k^2 \mathrm{Var}[X_k] + \sum_{k=1}^n \sum_{j \neq k=1}^n b_k b_j \mathrm{Cov}[X_k, X_j] = b^{\mathrm{T}} \Sigma b\,, \end{aligned}$$

che è quanto si voleva dimostrare. $\square$

Dalla Proposizione 4.4.8 si deduce che la matrice di varianza-covarianza è semidefinita positiva. In effetti, dal momento che $b^{\mathrm{T}} \Sigma b$ è una varianza di una v.a., per ogni b diverso dal vettore a componenti nulle si ha $b^{\mathrm{T}} \Sigma b \geq 0$, ovvero la condizione che definisce una matrice semidefinita positiva. Inoltre, dalla Proposizione 4.4.8 si ha che la varianza di una somma di v.a., ovvero $Y = \sum_{k=1}^n X_k$, risulta

$$\mathrm{Var}[Y] = \sum_{k=1}^n \sigma_{X_k}^2 + \sum_{k=1}^n \sum_{j \neq k=1}^n \sigma_{X_k, X_j}$$

e quindi, se le v.a. sono non correlate, si ha

$$\text{Var}[Y] = \sum_{k=1}^{n} \sigma_{X_k}^2 \, .$$

In particolare, per $n = 2$ si ottiene

$$\text{Var}[X_1 + X_2] = \text{Var}[X_1] + \text{Var}[X_2] + 2\,\text{Cov}[X_1, X_2] \, ,$$

mentre

$$\text{Var}[X_1 - X_2] = \text{Var}[X_1] + \text{Var}[X_2] - 2\,\text{Cov}[X_1, X_2] \, .$$

• **Esempio 4.4.4.** Si consideri di nuovo il v.v.a. analizzato nell'Esempio 4.2.3 e nell'Esempio 4.4.2. I precedenti risultati consento di determinare la varianza della v.a. trasformata $Y = X_1 + X_2$ senza doverne determinare la relativa distribuzione. In effetti, tenendo presente i risultati dell'Esempio 4.4.2, si ottiene $\text{Var}[X_1 + X_2] = \frac{1}{25}$. Inoltre, risulta anche $\text{Var}[X_1 - X_2] = \frac{7}{75}$. □

Un v.v.a. $X = (X_1, \ldots, X_n)^{\text{T}}$ definito sullo spazio probabilizzato $(\Omega, \mathcal{F}, P)$, è detto *linearmente degenere* se la trasformata $Y = a + \sum_{k=1}^{n} b_k X_k$ è degenere. In pratica, se un v.v.a. è linearmente degenere, esiste una componente del v.v.a. che è una combinazione lineare delle rimanenti $(n-1)$ componenti. Da un punto di vista geometrico questo implica che la distribuzione di probabilità sia concentrata su un iperpiano a $(n-1)$ dimensioni. Se X è un v.v.a. linearmente degenere, allora dalla Proposizione 4.3.8 si ha

$$\text{Var}[Y] = b^{\text{T}} \Sigma b = 0 \, ,$$

ovvero Σ non è definita positiva. Di conseguenza, si ha che il rango di Σ è inferiore a n e quindi $\det(\Sigma) = 0$. Quanto detto appare evidente nel caso di due variabili. In effetti, quando $n = 2$, si ha

$$\det(\Sigma) = \sigma_{X_1}^2 \sigma_{X_2}^2 - \sigma_{X_1,X_2}^2 = \sigma_{X_1}^2 \sigma_{X_2}^2 (1 - \rho_{X_1,X_2}^2) \, .$$

Dunque, se X è un v.v.a. linearmente degenere, dalla Proposizione 4.4.6 si ha $|\rho_{X_1,X_2}| = 1$ e quindi $\det(\Sigma) = 0$. Al contrario, se X non è un v.v.a. linearmente degenere, dalla Proposizione 4.4.7 si ha $|\rho_{X_1,X_2}| < 1$ e quindi $\det(\Sigma) > 0$, ovvero Σ_X risulta definita positiva.

4.5. Riferimenti bibliografici

Un approccio non usuale alla Teoria della Probabilità basato su una assiomatizzazione del valore atteso è contenuto in Whittle (2000). I testi avanzati, quali Ash e Doléans-Dade (2000), Athreya e Lahiri (2006), Brémaud (2020), Gut (2005), Khoshnevisan (2007), Resnick (2014) e Shiryaev (2016) considerano estesamente il concetto di valore atteso. Per quanto riguarda una trattazione approfondita delle disuguaglianze nella Teoria della Probabilità, si dovrebbero consultare i testi di Boucheron *et al.* (2013) e Lin e Bai (2010).

4.6. Esercizi svolti

• **Nota.** Al fine di evitare notazioni ridondanti e ripetizioni, nei seguenti esercizi si assume l'esistenza di uno spazio probabilizzato $(\Omega, \mathcal{F}, P)$, a meno che non venga specificato diversamente.

Sezione 4.1

• **Esercizio 4.1.1.** Si consideri la v.a. X tale che $P(X \geq 0) = 1$. Si verifichi che

$$\mathrm{E}[X] = \int_0^\infty P(X \geq x)\, dx\,.$$

Si dimostri inoltre che, nel caso di una v.a. discreta X con legge essenziale definita su $\mathbb{N}$, la precedente espressione può essere riformulata come

$$\mathrm{E}[X] = \sum_{x=1}^{\infty} P(X \geq x)\,.$$

Soluzione. Dalla definizione di valore atteso si ha

$$\mathrm{E}[X] = \int X\, dP = \int\int_0^\infty \mathbf{1}_{]-\infty,X]}(x)\, dx\, dP = \int\int_0^\infty \mathbf{1}_{[x,\infty[}(X)\, dx\, dP\,.$$

Sulla base del Teorema di Fubini l'ordine di integrazione può essere invertito e quindi

$$\mathrm{E}[X] = \int_0^\infty \int \mathbf{1}_{[x,\infty[}(X)\, dP\, dx = \int_0^\infty P(X \geq x)\, dx\,.$$

Si osservi che, essendo $\mathrm{E}[X] = \mathrm{E}[X^+] - \mathrm{E}[X^-]$ e applicando il precedente risultato rispettivamente a $\mathrm{E}[X^+]$ e $\mathrm{E}[X^-]$, per una v.a. X si ha in generale

$$\mathrm{E}[X] = \int_0^\infty P(X \geq x)\, dx - \int_{-\infty}^0 P(X \leq x)\, dx\,.$$

Se si considera una v.a. discreta X con legge essenziale definita su $\mathbb{N}$, è semplice verificare che la funzione $P(X \geq x)$ è costante a tratti. In particolare, questa funzione assume valore $P(X \geq k+1)$ in ogni intervallo $]k, k+1]$, dove $k \in \mathbb{N}$. Dunque, dalla espressione ottenuta per $\mathrm{E}[X]$ si ha

$$\mathrm{E}[X] = \sum_{k=0}^{\infty} \int_k^{k+1} P(X \geq x)\, dx = \sum_{k=0}^{\infty} P(X \geq k+1) = \sum_{x=1}^{\infty} P(X \geq x)\,. \qquad \square$$

• **Esercizio 4.1.2.** (Legge Zeta) Si consideri la v.a. discreta X con legge Zeta (detta anche legge di Zipf, si veda l'Esempio 7.2.7), ovvero la f.p. di X è data da

$$p_X(x) = \frac{1}{\zeta(s)}\, x^{-s}\, \mathbf{1}_{\{1,2,\ldots\}}(x)\,,$$

dove $s \in]1, \infty[$. Si ricorda che

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s}\,,$$

dove $\Re(s) \in]1, \infty[$, rappresenta la funzione Zeta di Riemann. Si determini il valore atteso $\mathrm{E}[X]$.

Soluzione. Si ha

$$\mathrm{E}[X] = \sum_{x=1}^{\infty} x\, \frac{1}{\zeta(s)}\, x^{-s} = \frac{1}{\zeta(s)} \sum_{x=1}^{\infty} x^{-(s-1)} = \frac{\zeta(s-1)}{\zeta(s)}\,,$$

e dunque $\mathrm{E}[X]$ esiste finito solo se $s \in]2, \infty[$, mentre risulta $\mathrm{E}[X] = \infty$ se $s \in]1, 2]$. $\square$

• **Esercizio 4.1.3.** Se è $S = (x_n)_{n\geq 1}$ una enumerazione dei numeri razionali, si consideri la v.a. X con legge essenziale data dall'insieme delle probabilità $(2^{-n})_{n\geq 1}$ e con f.r.

$$F_X(x) = \sum_{n=1}^{\infty} \mathbf{1}_{]-\infty,x]}(x_n) 2^{-n} \, .$$

Si determini il valore atteso $\mathrm{E}[X]$.

Soluzione. Dal momento che l'enumerazione dei numeri razionali non è unica, il valore atteso dipende da questa enumerazione e quindi non è definito in modo univoco. □

Sezione 4.2

• **Esercizio 4.2.1.** Si consideri una v.a. X simmetrica rispetto all'origine e la funzione misurabile $g : \mathbb{R} \to \mathbb{R}$. Si assuma inoltre che $g(-x) = -g(x)$ per ogni $x \in \mathbb{R}$, ovvero che la funzione g sia dispari. Si verifichi che risulta $\mathrm{E}[g(X)] = 0$.

Soluzione. Tenendo presente che $X \stackrel{\mathcal{L}}{=} -X$, si ha

$$\mathrm{E}[g(X)] = \mathrm{E}[g(-X)] = \mathrm{E}[-g(X)] = -\mathrm{E}[g(X)] \, ,$$

da cui si ottiene immediatamente la relazione $\mathrm{E}[g(X)] = 0$. □

• **Esercizio 4.2.2.** (Relazione fra media geometrica e media aritmetica) Date le successioni $(x_n)_{n\geq 1}$ e $(p_n)_{n\geq 1}$, dove $x_n > 0$ e $p_n \geq 0$ per $n = 1, 2, \ldots$, mentre $\sum_{n=1}^{\infty} p_n = 1$, si verifichi che

$$\prod_{n=1}^{\infty} x_n^{p_n} \leq \sum_{n=1}^{\infty} x_n p_n$$

con metodi probabilistici.

Soluzione. Si consideri la v.a. discreta X con f.p. data da

$$p_X(x) = \sum_{n=1}^{\infty} p_n \mathbf{1}_{\{x_n\}}(e^x) = \sum_{n=1}^{\infty} p_n \mathbf{1}_{\{\log(x_n)\}}(x) \, .$$

Evidentemente, si ha

$$\mathrm{E}[X] = \sum_{n=1}^{\infty} \log(x_n) p_n$$

e

$$\mathrm{E}[e^X] = \sum_{n=1}^{\infty} x_n p_n \, .$$

Dal momento che la funzione esponenziale è convessa, dalla disuguaglianza di Jensen si ottiene

$$e^{\mathrm{E}[X]} = \prod_{n=1}^{\infty} x_n^{p_n} \leq \mathrm{E}[e^X] = \sum_{n=1}^{\infty} x_n p_n \, ,$$

ovvero si ha la nota disuguaglianza fra la media geometrica e la media aritmetica. □

• **Esercizio 4.2.3.** Si consideri il v.v.a. $X = (X_1, X_2)^{\mathrm{T}}$ tale che le v.a. X_1 e X_2 sono indipendenti e con la medesima legge. Inoltre, si assuma che $\{X_1 > 0\}$ $q.c.$ e $\{X_2 > 0\}$ $q.c.$, mentre $\mathrm{E}[\frac{1}{X_1}] < \infty$. Si verifichi che $\mathrm{E}[\frac{X_1}{X_2}] \geq 1$ e $\mathrm{E}[\frac{X_2}{X_1}] \geq 1$.

Soluzione. Dal momento che la funzione $g(x) = x^{-1}$ è convessa su $\mathbb{R}^+$, dalla disuguaglianza di Jensen si ottiene

$$\frac{1}{\mathrm{E}[X_2]} \leq \mathrm{E}\left[\frac{1}{X_2}\right] .$$

Dunque, dall'assunzione di indipendenza si ha

$$\mathrm{E}\left[\frac{X_1}{X_2}\right] = \mathrm{E}[X_1]\mathrm{E}\left[\frac{1}{X_2}\right] \geq \mathrm{E}[X_1]\,\frac{1}{\mathrm{E}[X_2]} = 1 .$$

Inoltre, dal momento che $X_1 \stackrel{\mathcal{L}}{=} X_2$, per simmetria si ha anche $\mathrm{E}[\frac{X_2}{X_1}] \geq 1$. Il risultato è solo apparentemente paradossale, in quanto si deve tenere presente che le relazioni ottenute sono riferite a valori attesi di rapporti di v.a. e non a rapporti numerici. □

• **Esercizio 4.2.4.** Si consideri la v.a. assolutamente continua X che ammette d.p. data da $f_X(x) = e^{-x}\mathbf{1}_{[0,\infty[}(x)$ e si determini il valore atteso $\mathrm{E}[\lfloor X \rfloor]$, dove $\lfloor x \rfloor$ denota il più grande numero intero minore o uguale a $x \in \mathbb{R}$.

Soluzione. Tenendo presente le proprietà della serie geometrica, si ha

$$\begin{aligned}\mathrm{E}[\lfloor X \rfloor] &= \int_0^\infty \lfloor x \rfloor\, e^{-x}\, dx = \sum_{n=0}^{\infty} \int_n^{n+1} n e^{-x}\, dx = (1 - e^{-1}) \sum_{n=0}^{\infty} n e^{-n} \\ &= (1 - e^{-1})\, \frac{e^{-1}}{(1 - e^{-1})^2} = \frac{1}{e - 1} .\end{aligned}$$

Si osservi che $\lfloor x \rfloor \leq x$ per $x \in \mathbb{R}$ e per le proprietà del valore atteso deve risultare $\mathrm{E}[\lfloor X \rfloor] \leq \mathrm{E}[X]$. In effetti, essendo $\mathrm{E}[X] = 1$ si ha $\frac{1}{e-1} < 1$. □

• **Esercizio 4.2.5.** (Formula di Inclusione ed Esclusione) Si verifichi la Formula di Inclusione ed Esclusione (Teorema 2.2.5) mediante le proprietà del valore atteso.

Soluzione. Sviluppando il prodotto considerato nella relazione per funzioni indicatrici ottenuta nell'Esercizio 3.1.3 per l'unione di eventi, si ha

$$\begin{aligned}\mathbf{1}_{\bigcup_{k=1}^n E_k} &= 1 - \prod_{k=1}^{n} (1 - \mathbf{1}_{E_k}) = \sum_{k=1}^{n} (-1)^{k+1} \sum_{1 \leq j_1 < \ldots < j_k \leq n} \mathbf{1}_{E_{j_1}} \cdots \mathbf{1}_{E_{j_k}} \\ &= \sum_{k=1}^{n} (-1)^{k+1} \sum_{1 \leq j_1 < \ldots < j_k \leq n} \mathbf{1}_{E_{j_1} \cap \ldots \cap E_{j_k}} ,\end{aligned}$$

dove si è tenuto presente anche l'espressione per l'intersezione di eventi ottenuta nel medesimo esercizio. Dunque, per le proprietà del valore atteso si ha

$$\mathrm{E}[\mathbf{1}_{\bigcup_{k=1}^n E_k}] = \sum_{k=1}^{n} (-1)^{k+1} \sum_{1 \leq j_1 < \ldots < j_k \leq n} \mathrm{E}[\mathbf{1}_{E_{j_1} \cap \ldots \cap E_{j_k}}] .$$

Dal momento che $\mathrm{E}[\mathbf{1}_E] = P(E)$, la Formula di Inclusione ed Esclusione risulta immediatamente dalla precedente relazione. □

• **Esercizio 4.2.6.** Se X è v.a. discreta con f.p. data da $p_X(x) = 2^{-x}\mathbf{1}_{\{1,2,\ldots\}}(x)$, si consideri la trasformata

$$Y = g(X) = (-1)^{X+1}\,\frac{2^X}{X}\,.$$

Si verifichi che $\mathrm{E}[Y]$ non è definito anche se

$$\sum_{x=1}^{\infty} g(x)p_X(x) = -\sum_{x=1}^{\infty}\frac{(-1)^x}{x} = \log(2) < \infty\,.$$

Soluzione. Si tenga presente che la v.a. Y è discreta e assume valori positivi se X assume valori interi dispari e viceversa. Dunque, per $x = 2k-1$ si ha

$$\mathrm{E}[Y^+] = \sum_{k=1}^{\infty}(-1)^{2k}\,\frac{2^{2k-1}}{2k-1}\,p_X(2k-1) = \sum_{k=1}^{\infty}\frac{1}{2k-1} > \frac{1}{2}\sum_{k=1}^{\infty}\frac{1}{k} = \infty\,,$$

mentre per $x = 2k$ si ha

$$\mathrm{E}[Y^-] = -\sum_{k=1}^{\infty}(-1)^{2k+1}\,\frac{2^{2k}}{2k}\,p_X(2k) = \frac{1}{2}\sum_{k=1}^{\infty}\frac{1}{k} = \infty\,,$$

ovvero $\mathrm{E}[Y]$ non è definito. □

Sezione 4.3

• **Esercizio 4.3.1.** (Disuguaglianza di Cantelli) Si consideri la v.a. X tale che $\mathrm{E}[X] = \mu$ e $\mathrm{Var}[X] = \sigma^2 < \infty$. Se $c > 0$, si verifichi che

$$P(|X-\mu| \geq c) \leq \frac{2\sigma^2}{\sigma^2 + c^2}\,.$$

Soluzione. Si ponga $Y = X - \mu$, per cui $\mathrm{E}[Y] = 0$ e $\mathrm{Var}[Y] = \sigma^2$. Tenendo presente la disuguaglianza di Markov, per $a > 0$ si ha

$$P(Y \geq c) = P(Y + a \geq c + a) \leq P((Y+a)^2 \geq (c+a)^2) \leq \frac{1}{(c+a)^2}\,\mathrm{E}[(Y+a)^2] = \frac{\sigma^2 + a^2}{(c+a)^2}\,.$$

Il secondo membro della disuguaglianza viene minimizzato per $a = \frac{\sigma^2}{c}$ e, sostituendo questo valore nella precedente espressione, si ha

$$P(Y \geq c) \leq \frac{\sigma^2}{\sigma^2 + c^2}\,.$$

In modo analogo, si ottiene che

$$P(-Y \geq c) \leq \frac{\sigma^2}{\sigma^2 + c^2}$$

e, combinando le due disuguaglianze, si ha

$$P(Y \geq c) + P(-Y \geq c) = P(|Y| \geq c) \leq \frac{2\sigma^2}{\sigma^2 + c^2}\,.$$

Infine, si noti che la disuguaglianza di Cantelli è preferibile a quella di Chebyshev quando

$$\frac{2\sigma^2}{\sigma^2 + c^2} \leq \frac{\sigma^2}{c^2}\,,$$

ovvero quando $c < \sigma$. □

• **Esercizio 4.3.2.** (Disuguaglianza di Chernoff) Considerata la v.a. X, se $c > 0$ si verifichi che

$$P(X \geq c) \leq e^{-\psi_X(c)} ,$$

dove

$$\psi_X(c) = \sup_{t \geq 0} (ct - \log(\mathrm{E}[e^{tX}]) .$$

Soluzione. Tenendo presente la disuguaglianza di Markov, per $t \geq 0$ si ha

$$P(X \geq c) = P(e^{tX} \geq e^{tc}) \leq \frac{\mathrm{E}[e^{tX}]}{e^{tc}} .$$

Dalla precedente espressione risulta dunque

$$P(X \geq c) \leq \inf_{t \geq 0} \frac{\mathrm{E}[e^{tX}]}{e^{tc}} = e^{-\psi_X(c)} .$$

Si noti che la disuguaglianza di Chernoff può fornire un estremo superiore di ordine esponenziale. Considerando un caso specifico, si assuma che la v.a. assolutamente continua X ammetta d.p. data da $f_X(x) = e^{-x}\mathbf{1}_{[0,\infty[}(x)$. In questo caso, se $t \in [0, 1[$ si ha

$$\mathrm{E}[e^{tX}] = \int_0^\infty e^{tx} e^{-x}\, dx = \frac{1}{1-t} ,$$

per cui $\psi_X(c) = c - 1 - \log(c)$ se $c > 1$, ovvero la disuguaglianza di Chernoff è data da

$$P(X \geq c) \leq c e^{-c+1} .$$

Si noti che il valore esatto è dato da $P(X \geq c) = e^{-c}$. La disuguaglianza prende il nome dallo statistico e fisico Herman Chernoff (1923-). □

Sezione 4.4

• **Esercizio 4.4.1.** Si consideri una v.a. discreta X tale che $\{X \geq 0\}$ $q.c.$ e $\mathrm{E}[X^2] < \infty$. Si verifichi che

$$P(X > 0) \geq \frac{\mathrm{E}[X]^2}{\mathrm{E}[X^2]} .$$

Soluzione. Considerando la v.a. $Y = \mathbf{1}_{]0,\infty[}(X)$, si ha

$$\mathrm{E}[Y^2] = \mathrm{E}[\mathbf{1}_{]0,\infty[}(X)^2] = \mathrm{E}[\mathbf{1}_{]0,\infty[}(X)] = P(X > 0)$$

e

$$\mathrm{E}[XY] = \mathrm{E}[X\,\mathbf{1}_{]0,\infty[}(X)] = \mathrm{E}[X\,\mathbf{1}_{[0,\infty[}(X)] = \mathrm{E}[X] .$$

Dunque, dalla disuguaglianza di Schwarz si ottiene

$$\mathrm{E}[X^2]\mathrm{E}[Y^2] = \mathrm{E}[X^2]P(X > 0) \geq \mathrm{E}[XY]^2 = \mathrm{E}[X]^2 ,$$

da cui segue la disuguaglianza richiesta. □

• **Esercizio 4.4.2.** (Disuguaglianza di Chebyshev bivariata) Si consideri il v.v.a. $X = (X_1, X_2)^{\mathrm{T}}$. Assumendo che $\sigma^2_{X_1}, \sigma^2_{X_2} < \infty$, si verifichi che per $c > 0$ risulta

$$P(\{|X_1 - \mu_{X_1}| \geq c\sigma_{X_1}\} \cup \{|X_2 - \mu_{X_2}| \geq c\sigma_{X_2}\}) \leq \frac{1}{c^2}\,(1 + \sqrt{1 - \rho^2_{X_1,X_2}})\,.$$

Soluzione. Posto $Z_1 = \frac{X_1 - \mu_{X_1}}{\sigma_{X_1}}$ e $Z_2 = \frac{X_2 - \mu_{X_2}}{\sigma_{X_2}}$, essendo $\rho_{Z_1,Z_2} = \rho_{X_1,X_2}$, la disuguaglianza da dimostrare si riduce a

$$P(\{|Z_1| \geq c\} \cup \{|Z_2| \geq c\}) \leq \frac{1}{c^2}\,(1 + \sqrt{1 - \rho^2_{Z_1,Z_2}})\,.$$

Tenendo presente la disuguaglianza di Markov, si ha

$$\begin{aligned} P(\{|Z_1| \geq c\} \cup \{|Z_2| \geq c\}) &= P(\{Z_1^2 \geq c^2\} \cup \{Z_2^2 \geq c^2\}) \\ &= P(\max(Z_1^2, Z_2^2) \geq c^2) \leq \frac{1}{c^2}\,\mathrm{E}[\max(Z_1^2, Z_2^2)]\,. \end{aligned}$$

Inoltre, dal momento che per $x, y \in \mathbb{R}$ si ha

$$\max(x^2, y^2) = \frac{1}{2}\,(x^2 + y^2 + |x^2 - y^2|) = \frac{1}{2}\,(x^2 + y^2 + |x + y| \cdot |x - y|)$$

ed essendo $\mathrm{E}[Z_1^2] = \mathrm{E}[Z_2^2] = 1$ e $\mathrm{E}[Z_1 Z_2] = \rho_{Z_1,Z_2}$, dalla disuguaglianza di Schwarz si ottiene

$$\begin{aligned} \mathrm{E}[\max(Z_1^2, Z_2^2)] &= 1 + \frac{1}{2}\,\mathrm{E}[|Z_1 + Z_2| \cdot |Z_1 - Z_2|] \leq 1 + \frac{1}{2}\,\sqrt{\mathrm{E}[(Z_1 + Z_2)^2]\mathrm{E}[(Z_1 - Z_2)^2]} \\ &= 1 + \sqrt{1 - \rho^2_{Z_1,Z_2}}\,, \end{aligned}$$

da cui segue immediatamente la disuguaglianza richiesta. □

• **Esercizio 4.4.3.** Si consideri il v.v.a. assolutamente continuo $X = (X_1, X_2)^{\mathrm{T}}$ che ammette d.p.c. data da

$$f_X(x_1, x_2) = ((1 + cx_1)(1 + cx_2) - c)\, e^{-x_1 - x_2 - cx_1x_2}\, \mathbf{1}_{[0,\infty[}(x_1)\mathbf{1}_{[0,\infty[}(x_2)\,,$$

dove $c \in [0, 1]$. Si determini il coefficiente di correlazione ρ_{X_1,X_2}.

Soluzione. Si ha

$$\mathrm{E}[X_1] = \int_0^\infty \int_0^\infty x_1((1 + cx_1)(1 + cx_2) - c)\, e^{-x_1 - x_2 - cx_1x_2}\, dx_1 dx_2 = \int_0^\infty x_1\, e^{-x_1}\, dx_1 = 1$$

e dunque per simmetria risulta anche $\mathrm{E}[X_2] = 1$. Inoltre, risulta

$$\mathrm{E}[X_1^2] = \int_0^\infty \int_0^\infty x_1^2((1 + cx_1)(1 + cx_2) - c)\, e^{-x_1 - x_2 - cx_1x_2}\, dx_1 dx_2 = \int_0^\infty x_1^2\, e^{-x_1}\, dx_1 = 2$$

e quindi $\mathrm{E}[X_2^2] = 1$. Dalle precedenti espressioni si ha $\mathrm{Var}[X_1] = \mathrm{Var}[X_2] = 1$. Inoltre, se $c \neq 0$ risulta

$$\begin{aligned} \mathrm{E}[X_1 X_2] &= \int_0^\infty \int_0^\infty x_1 x_2((1 + cx_1)(1 + cx_2) - c)\, e^{-x_1 - x_2 - cx_1x_2}\, dx_1 dx_2 \\ &= \int_0^\infty \frac{x_1(1 + cx_1 + c)}{(1 + cx_1)^2}\, e^{-x_1}\, dx_1 = -\frac{e^{1/c}}{c}\,\mathrm{Ei}\Big(-\frac{1}{c}\Big)\,, \end{aligned}$$

dove

$$\mathrm{Ei}(x) = \int_{-\infty}^{x} \frac{e^u}{u}\, du$$

rappresenta la funzione integrale esponenziale. In questo caso, si ha

$$\rho_{X_1,X_2} = -\frac{e^{1/c}}{c}\,\mathrm{Ei}\Big(-\frac{1}{c}\Big) - 1\,.$$

Se $c = 0$ risulta immediato verificare che le v.a. X_1 e X_2 sono indipendenti e quindi $\rho_{X_1,X_2} = 0$. Inoltre, si può verificare che ρ_{X_1,X_2} è una funzione decrescente di c e quindi si ha $\rho_{X_1,X_2} \in [-e\mathrm{Ei}(-1) - 1, 0]$. □

• **Esercizio 4.4.4.** (Identità di Hoeffding) Dato il v.v.a. $X = (X_1, X_2)^{\mathrm{T}}$, si verifichi che

$$\mathrm{Cov}[X] = \int_{-\infty}^{\infty}\int_{\infty}^{\infty} (F_X(x_1, x_2) - F_{X_1}(x_1)F_{X_2}(x_2))\, dx_1 dx_2\,,$$

assumendo che $\mathrm{E}[X_1] < \infty$ e $\mathrm{E}[X_2] < \infty$.

Soluzione. Si consideri l'ulteriore v.v.a. $Y = (Y_1, Y_2)^{\mathrm{T}}$ tale che X è indipendente da Y e $X \overset{\mathcal{L}}{=} Y$. Risulta immediato verificare che

$$\mathrm{E}[(X_1 - Y_1)(X_2 - Y_2)] = 2\,\mathrm{Cov}[X]\,.$$

Inoltre, si ha

$$X_1 - Y_1 = \int_{-\infty}^{\infty} (\mathbf{1}_{[Y_1,\infty[}(x_1) - \mathbf{1}_{[X_1,\infty[}(x_1))\, dx_1 = \int_{-\infty}^{\infty} (\mathbf{1}_{]-\infty,x_1]}(Y_1) - \mathbf{1}_{]-\infty,x_1]}(X_1))\, dx_1$$

e, in modo simile,

$$X_2 - Y_2 = \int_{-\infty}^{\infty} (\mathbf{1}_{]-\infty,x_2]}(Y_2) - \mathbf{1}_{]-\infty,x_2]}(X_2))\, dx_2\,.$$

Infine, si noti che

$$\begin{aligned}\mathrm{E}[(\mathbf{1}_{]-\infty,x_1]}(Y_1) - \mathbf{1}_{]-\infty,x_1]}(X_1))(\mathbf{1}_{]-\infty,x_2]}(Y_2) - \mathbf{1}_{]-\infty,x_2]}(X_2))] &= \\ &= 2\,\mathrm{E}[(\mathbf{1}_{]-\infty,x_1]}(X_1)\mathbf{1}_{]-\infty,x_2]}(X_2) - \mathbf{1}_{]-\infty,x_1]}(X_1)\mathbf{1}_{]-\infty,x_2]}(X_2))] \\ &= 2\,(F_X(x_1, x_2) - F_{X_1}(x_1)F_{X_2}(x_2))\,.\end{aligned}$$

Tenendo presente tutte le precedenti relazioni, sulla base del Teorema di Fubini si ottiene l'identità richiesta. La relazione prende il nome dallo statistico probabilista finlandese Wassily Hoeffding (1914-1991). □

• **Esercizio 4.4.5.** (Coefficienti di asimmetria e curtosi) Si osservi che il coefficiente di asimmetria di una v.a. X tale che $\mathrm{E}[X^3] < \infty$ è definito come

$$\beta_{1,X} = \frac{\mathrm{E}[(X - \mu_X)^3]}{\sigma_X^3}\,,$$

mentre se $\mathrm{E}[X^4] < \infty$ il coefficiente di curtosi è definito come

$$\beta_{2,X} = \frac{\mathrm{E}[(X - \mu_X)^4]}{\sigma_X^4}\,.$$

Si consideri il v.v.a. $(X_1, \ldots, X_n)^{\mathrm{T}}$ a componenti indipendenti con la medesima legge della v.a. X. Data la trasformata $Y = \sum_{k=1}^{n} X_k$, si determini $\beta_{1,Y}$ e $\beta_{2,Y}$.

Soluzione. Risulta immediato verificare che $\mu_Y = n\mu_X$ e $\sigma_Y^2 = n\sigma_X^2$. Per quanto riguarda il coefficiente di asimmetria, dall'assunzione di indipendenza e dal Teorema Multinomiale si ha

$$\begin{aligned}\mathrm{E}[(Y-\mu_Y)^3] &= \mathrm{E}[(\sum_{k=1}^{n} X_k - \sum_{k=1}^{n} \mu_X])^3] = \mathrm{E}[(\sum_{k=1}^{n} (X_k - \mu_X))^3] \\ &= \sum_{j_1+\ldots+j_n=3} \binom{3}{j_1 \ldots j_n} \prod_{k=1}^{n} \mathrm{E}[(X_k-\mu_X)^{j_k}] \,.\end{aligned}$$

Tenendo presente che gli addendi nella sommatoria che contengono i fattori $\mathrm{E}[X_k - \mu_X] = 0$ si annullano, mentre $\mathrm{E}[(X_k-\mu_X)^3] = \mathrm{E}[(X-\mu_X)^3]$ per $k = 1, \ldots, n$ dal momento che le componenti del v.v.a. X sono equivalenti in legge alla v.a. X, risulta

$$\mathrm{E}[(Y-\mu_Y)^3] = \sum_{k=1}^{n} \mathrm{E}[(X_k-\mu_X)^3] = n\mathrm{E}[(X-\mu_X)^3] \,,$$

da cui

$$\beta_{1,Y} = \frac{\mathrm{E}[(Y-\mu_Y)^3]}{\sigma_Y^3} = \frac{n\mathrm{E}[(X-\mu_X)^3]}{\sqrt{n^3}\sigma_X^3} = \frac{\beta_{1,X}}{\sqrt{n}} \,.$$

Per quanto riguarda il coefficiente di curtosi, si ha

$$\begin{aligned}\mathrm{E}[(Y-\mu_Y)^4] &= \mathrm{E}[(\sum_{k=1}^{n} X_k - \sum_{k=1}^{n} \mu_X])^4] = \mathrm{E}[(\sum_{k=1}^{n} (X_k - \mu_X))^4] \\ &= \sum_{j_1+\ldots+j_n=4} \binom{4}{j_1 \ldots j_n} \prod_{k=1}^{n} \mathrm{E}[(X_k-\mu_X)^{j_k}] \,.\end{aligned}$$

Dal momento che

$$\mathrm{E}[(X_k-\mu_X)^2(X_j-\mu_X)^2] = \mathrm{E}[(X_k-\mu_X)^2]\mathrm{E}[(X_j-\mu_X)^2] = \sigma_X^4$$

ed esistono $\binom{4}{2}\binom{n}{2} = 3n(n-1)$ addendi di questo tipo nella sommatoria, tenendo presente le osservazioni fatte per il coefficiente di asimmetria si ha

$$\mathrm{E}[(Y-\mu_Y)^4] = n\mathrm{E}[(X-\mu_X)^4] + 3n(n-1)\sigma_X^4 \,.$$

Si deve dunque concludere che

$$\beta_{2,Y} = \frac{\mathrm{E}[(Y-\mu_Y)^4]}{\sigma_Y^4} = \frac{n\mathrm{E}[(X-\mu_X)^4] + 3n(n-1)\sigma_X^4}{n^2\sigma_X^4} = 3 + \frac{\beta_{2,X}-3}{n} \,.$$

Risulta interessante osservare che $\lim_n \beta_{1,Y} = 0$ e $\lim_n \beta_{2,Y} = 3$. In effetti, questi valori del coefficiente di asimmetria e curtosi corrispondono a quelli di una v.a. con legge Normale ridotta (si veda l'Esempio 7.1.8). □

Capitolo 5

Valori attesi condizionati

5.1. Valore atteso condizionato

Prima di introdurre formalmente in modo generale il concetto di valore atteso condizionato è opportuno definire il valore atteso condizionato di una v.a. rispetto ad un evento. Tenendo presente la Proposizione 2.4.2, si ha dunque la seguente definizione.

Definizione 5.1.1. Data la v.a. X definita sullo spazio probabilizzato $(\Omega, \mathcal{F}, P)$ e l'evento $E_0 \in \mathcal{F}$ tale che $P(E_0) > 0$, si dice *valore atteso condizionato* all'evento E_0 l'integrale

$$\mathrm{E}[X \mid E_0] = \int X \, dP(\,\cdot \mid E_0) = \frac{1}{P(E_0)} \int_{E_0} X \, dP$$

se esiste finito. □

Si osservi che dalla Definizione 5.1.1 si ha in effetti

$$\mathrm{E}[X \mid E_0] = \frac{\mathrm{E}[X \, \mathbf{1}_{E_0}]}{P(E_0)} \, .$$

Se $X = \mathbf{1}_E$ dove $E \in \mathcal{F}$, allora risulta

$$\mathrm{E}[\mathbf{1}_E \mid E_0] = \frac{\mathrm{E}[\mathbf{1}_E \mathbf{1}_{E_0}]}{P(E_0)} = \frac{\mathrm{E}[\mathbf{1}_{E \cap E_0}]}{P(E_0)} = \frac{P(E \cap E_0)}{P(E_0)} = P(E \mid E_0) \, ,$$

ovvero la probabilità condizionata di E dato E_0 può essere interpretata come un valore atteso condizionato. Inoltre, è immediato verificare che il valore atteso condizionato all'evento E_0 esiste finito se il valore atteso $\mathrm{E}[X]$ esiste finito dal momento che $\mathrm{E}[X \, \mathbf{1}_{E_0}] \leq \mathrm{E}[X]$.

• **Esempio 5.1.1.** Si consideri la v.a. assolutamente continua X analizzata nell'Esempio 3.2.1 e si consideri l'evento $E_0 = X^{-1}(]0, \tfrac{1}{3}[)$. Dal momento che

$$P(E_0) = \int_0^{\frac{1}{3}} dx = \frac{1}{3}$$

e

$$\mathrm{E}[X \, \mathbf{1}_{E_0}] = \int_0^{\frac{1}{3}} x \, dx = \frac{1}{18} \, ,$$

si ha

$$\mathrm{E}[X \mid E_0] = \frac{1}{6} \, .$$ □

La successiva definizione introduce la definizione di valore atteso condizionato ad una σ-algebra, nel caso in cui la σ-algebra sia generata da una classe di eventi che costituiscono una partizione di Ω.

Definizione 5.1.2. Sia data la v.a. X definita sullo spazio probabilizzato $(\Omega, \mathcal{F}, P)$. Sia inoltre $(E_n)_{n\geq 1}$ una partizione di Ω tale che $P(E_n) > 0$ per ogni $n = 1, 2, \ldots$ e sia $\mathcal{F}_0$ la σ-algebra generata dalla partizione. Se $\mathrm{E}[X]$ esiste finito, si dice *valore atteso condizionato* alla σ-algebra $\mathcal{F}_0$ la v.a.

$$\mathrm{E}[X \mid \mathcal{F}_0] = \sum_{n=1}^{\infty} \mathrm{E}[X \mid E_n]\, \mathbf{1}_{E_n}\,. \qquad \square$$

Dalla precedente definizione, è immediato verificare che $\mathrm{E}[X \mid \mathcal{F}_0]$ è una v.a. che assume valore $\mathrm{E}[X \mid E_n]$ per ogni $\omega \in E_n$, dove $n = 1, 2, \ldots$. Dunque, $\mathrm{E}[X \mid \mathcal{F}_0]$ è una v.a. discreta che assume valori su $S = (\mathrm{E}[X \mid E_n])_{n\geq 1}$ con legge essenziale data dall'insieme delle probabilità $(P(E_n))_{n\geq 1}$.

• **Esempio 5.1.2.** Si consideri la v.a. assolutamente continua X analizzata nell'Esempio 5.1.1 e si consideri la partizione di Ω data da $\{E_1, E_2\}$, dove $E_1 = X^{-1}(]-\infty, \frac{1}{3}[)$ e $E_2 = X^{-1}([\frac{1}{3}, \infty[)$. Si ha $\mathcal{F}_0 = \{\emptyset, E_1, E_2, \Omega\}$. Inoltre, si ha $P(E_1) = \frac{1}{3}$ e $P(E_2) = \frac{2}{3}$, mentre $\mathrm{E}[X \mid E_1] = \frac{1}{6}$ e $\mathrm{E}[X \mid E_2] = \frac{2}{3}$. In questo caso, il valore atteso condizionato $\mathrm{E}[X \mid \mathcal{F}_0]$ è una v.a. discreta che assume valori su $S = \{\frac{1}{6}, \frac{2}{3}\}$ con legge essenziale data dall'insieme delle probabilità $\{\frac{1}{3}, \frac{2}{3}\}$. $\square$

Dalle Definizioni 5.1.1 e 5.1.2, tenendo presente che per un insieme opportuno di indici $I \subset \mathbb{N}$ ogni $E_0 \in \mathcal{F}_0$ può essere rappresentato come

$$E_0 = \bigcup_{n\in I} E_n\,,$$

si ha

$$\begin{aligned}\int_{E_0} \mathrm{E}[X \mid \mathcal{F}_0]\, dP &= \sum_{n=1}^{\infty} \mathrm{E}[X \mid E_n] \int_{E_0} \mathbf{1}_{E_n}\, dP = \sum_{n=1}^{\infty} \mathrm{E}[X \mid E_n] P(E_n \cap E_0) \\ &= \sum_{n\in I} \mathrm{E}[X \mid E_n] P(E_n) = \sum_{n\in I} \mathrm{E}[X\, \mathbf{1}_{E_n}] = \sum_{n\in I} \int_{E_n} X\, dP = \int_{E_0} X\, dP\,,\end{aligned}$$

ovvero

$$\mathrm{E}[\mathrm{E}[X \mid \mathcal{F}_0]\, \mathbf{1}_{E_0}] = \mathrm{E}[X\, \mathbf{1}_{E_0}]\,.$$

Si osservi che nel caso in cui $E_0 = \Omega$, la precedente relazione stabilisce che il valore atteso della v.a. $\mathrm{E}[X \mid \mathcal{F}_0]$ è pari al valore atteso della v.a. X, ovvero $\mathrm{E}[\mathrm{E}[X \mid \mathcal{F}_0]] = \mathrm{E}[X]$. Inoltre, se si ha anche $X = \mathbf{1}_E$ e tenendo presente che

$$\mathrm{E}[\mathbf{1}_E] = \int_{\Omega} \mathbf{1}_E\, dP = P(E)\,,$$

si ottiene la generalizzazione della Legge delle Probabilità Totali (Teorema 2.4.3), ovvero

$$P(E) = \int_{\Omega} \mathbf{1}_E\, dP = \sum_{n=1}^{\infty} \mathrm{E}[\mathbf{1}_E \mid E_n] P(E_n) = \sum_{n=1}^{\infty} P(E \mid E_n) P(E_n)\,.$$

Inoltre, la medesima relazione permette di introdurre la seguente definizione generale di valore atteso condizionato.

Definizione 5.1.3. Sia data la v.a. X definita sullo spazio probabilizzato $(\Omega, \mathcal{F}, P)$ e sia $\mathcal{F}_0 \subset \mathcal{F}$ una σ-algebra. Se $\mathrm{E}[X]$ esiste finito, si dice *valore atteso condizionato* alla σ-algebra $\mathcal{F}_0$ un membro della classe di v.a. che coincidono $q.c.$ rispetto a P e tali che per ogni $E_0 \in \mathcal{F}_0$ soddisfano la relazione

$$\int_{E_0} \mathrm{E}[X \mid \mathcal{F}_0]\, dP = \int_{E_0} X\, dP\,. \qquad \square$$

L'esistenza del valore atteso condizionato alla σ-algebra $\mathcal{F}_0$ è assicurata dal Teorema di Radon-Nikodym (Teorema A.9). Inoltre, se si ha la trasformata $Y = g(X)$ del v.v.a. $X = (X_1, \ldots, X_n)^{\mathrm{T}}$ definito sullo spazio probabilizzato $(\Omega, \mathcal{F}, P)$, il valore atteso condizionato alla σ-algebra $\mathcal{F}_0$ viene ovviamente definito mediante la relazione

$$\int_{E_0} \mathrm{E}[g(X) \mid \mathcal{F}_0]\, dP = \int_{E_0} g(X)\, dP$$

per ogni $E_0 \in \mathcal{F}_0$. Infine, nel caso in cui $E_0 = \Omega$ e $X = \mathbf{1}_E$ si ottiene una ulteriore generalizzazione della Legge delle Probabilità Totali (Teorema 2.4.3), ovvero

$$P(E) = \int \mathrm{E}[\mathbf{1}_E \mid \mathcal{F}_0]\, dP = \mathrm{E}[\mathrm{E}[\mathbf{1}_E \mid \mathcal{F}_0]]\,.$$

5.2. Proprietà del valore atteso condizionato

Le seguenti Proposizioni e Teoremi forniscono un insieme di proprietà che riguardano il valore atteso condizionato alla σ-algebra.

Proposizione 5.2.1. *Si consideri la v.a. X definita sullo spazio probabilizzato $(\Omega, \mathcal{F}, P)$ e sia $\mathcal{F}_0 \subset \mathcal{F}$ una σ-algebra. Se la v.a. X è tale che $X^{-1}(B) \in \mathcal{F}_0$ per ogni $B \in \mathcal{B}(\mathbb{R})$, allora si ha $\{\mathrm{E}[X \mid \mathcal{F}_0] = X\}$ $q.c.$*

Dimostrazione. Risulta immediata dalla relazione della Definizione 5.1.3. $\square$

Proposizione 5.2.2. *Se $X = (X_1, \ldots, X_n)^{\mathrm{T}}$ è un v.v.a. definito sullo spazio probabilizzato $(\Omega, \mathcal{F}, P)$, si consideri la trasformata $Y = a + \sum_{k=1}^n b_k X_k$ dove $a \in \mathbb{R}$ e $(b_1, \ldots, b_n)^{\mathrm{T}} \in \mathbb{R}^n$ sono costanti. Se $\mathcal{F}_0 \subset \mathcal{F}$ è una σ-algebra e se $\mathrm{E}[Y] < \infty$, si ha $\{\mathrm{E}[Y \mid \mathcal{F}_0] = a + \sum_{k=1}^n b_k \mathrm{E}[X_k \mid \mathcal{F}_0]\}$ $q.c.$*

Dimostrazione. Tenendo presente la Definizione 5.1.3, per ogni $E_0 \in \mathcal{F}_0$ si ha

$$\begin{aligned}
\int_{E_0} \mathrm{E}[Y \mid \mathcal{F}_0]\, dP &= \int_{E_0} \left(a + \sum_{k=1}^n b_k X_k \right) dP = a \int_{E_0} dP + b_k \sum_{k=1}^n \int_{E_0} X_k\, dP \\
&= a \int_{E_0} dP + b_k \sum_{k=1}^n \int_{E_0} \mathrm{E}[X_k \mid \mathcal{F}_0]\, dP \\
&= \int_{E_0} (a + \sum_{k=1}^n b_k\, \mathrm{E}[X_k \mid \mathcal{F}_0])\, dP\,,
\end{aligned}$$

da cui segue la tesi. $\square$

Il seguente Teorema fornisce le versioni condizionate delle principali disuguaglianze ottenute in precedenza per il valore atteso.

Teorema 5.2.3. *Si ha:*
i) **(Disuguaglianza di Jensen condizionata)** *Data la funzione convessa* $g : \mathbb{R}^n \to \mathbb{R}$*, si consideri la trasformata* $Y = g(X)$ *del v.v.a.* $X = (X_1, \ldots, X_n)^{\mathrm{T}}$ *definito sullo spazio probabilizzato* $(\Omega, \mathcal{F}, P)$ *e sia* $\mathcal{F}_0 \subset \mathcal{F}$ *una* σ*-algebra. Risulta* $\{g(\mathrm{E}[X \mid \mathcal{F}_0]) \leq \mathrm{E}[g(X) \mid \mathcal{F}_0]\}$ $q.c.$
ii) **(Disuguaglianza di Lyapunov condizionata)** *Si consideri la v.a.* X *definita sullo spazio probabilizzato* $(\Omega, \mathcal{F}, P)$ *e sia* $\mathcal{F}_0 \subset \mathcal{F}$ *una* σ*-algebra. Assumendo che* $0 < s \leq r$*, allora risulta* $\{\mathrm{E}[|X|^s \mid \mathcal{F}_0]^{1/s} \leq \mathrm{E}[|X|^r \mid \mathcal{F}_0]^{1/r}\}$ $q.c.$
iii) **(Disuguaglianza di Hölder condizionata)** *Si consideri il v.v.a.* $X = (X_1, X_2)^{\mathrm{T}}$ *definito sullo spazio probabilizzato* $(\Omega, \mathcal{F}, P)$ *e sia* $\mathcal{F}_0 \subset \mathcal{F}$ *una* σ*-algebra. Se* $r, s \in]1, \infty[$ *sono costanti tali che* $\frac{1}{r} + \frac{1}{s} = 1$*, allora risulta* $\{|\mathrm{E}[X_1 X_2 \mid \mathcal{F}_0]| \leq \mathrm{E}[|X_1|^r \mid \mathcal{F}_0]^{1/r} \mathrm{E}[|X_2|^s \mid \mathcal{F}_0]^{1/s}\}$ $q.c.$
Dimostrazione. Le dimostrazioni sono analoghe a quelle dei Teoremi 4.2.6, 4.3.2 e 4.4.2, dal momento che le dimostrazioni sono basate su disuguaglianze fra numeri reali. □

Nel caso in cui si consideri la v.a. X e la trasformata $g(X) = |X|$, dal momento che g è una funzione convessa, allora dalla parte *i*) del Teorema 5.2.3 risulta $\{|\mathrm{E}[X \mid \mathcal{F}_0]| \leq \mathrm{E}[|X| \mid \mathcal{F}_0]\}$ $q.c.$, ovvero si ottiene la disuguaglianza condizionata corrispondente a quella della Proposizione 4.2.1. Il seguente Teorema fornisce una classica scomposizione della varianza di una v.a.

Teorema 5.2.4. *Si consideri la v.a.* X *definita sullo spazio probabilizzato* $(\Omega, \mathcal{F}, P)$ *e sia* $\mathcal{F}_0 \subset \mathcal{F}$ *una* σ*-algebra. Se* $\mathrm{E}[X]$ *esiste finito, si ha*

$$\mathrm{E}[X] = \mathrm{E}[\mathrm{E}[X \mid \mathcal{F}_0]] \,,$$

mentre se $\mathrm{E}[X^2]$ *esiste finito, si ha*

$$\mathrm{Var}[X] = \mathrm{E}[\mathrm{Var}[X \mid \mathcal{F}_0]] + \mathrm{Var}[\mathrm{E}[X \mid \mathcal{F}_0]] \,,$$

dove $\mathrm{Var}[X \mid \mathcal{F}_0] = \mathrm{E}[X^2 \mid \mathcal{F}_0] - \mathrm{E}[X \mid \mathcal{F}_0]^2$.
Dimostrazione. La prima parte risulta immediata ponendo $E_0 = \Omega$ nella relazione data nella Definizione 5.1.3. Per quanto riguarda la seconda parte, posto $Y = X - \mathrm{E}[X \mid \mathcal{F}_0]$ e $Z = \mathrm{E}[X \mid \mathcal{F}_0] - \mathrm{E}[X]$, si osservi che

$$\mathrm{Var}[X] = \mathrm{E}[(Y + Z)^2] = \mathrm{E}[Y^2] + \mathrm{E}[Z^2] + 2\mathrm{E}[YZ] \,.$$

Si può verificare che $\{\mathrm{E}[YZ \mid \mathcal{F}_0] = Z\mathrm{E}[Y \mid \mathcal{F}_0]\}$ $q.c.$ se Z è una v.a. tale che $Z^{-1}(B) \in \mathcal{F}_0$ per ogni $B \in \mathcal{B}(\mathbb{R})$. Tenendo presente le Proposizioni 5.2.1 e 5.2.2 risulta inoltre $\{\mathrm{E}[Y \mid \mathcal{F}_0] = 0\}$ $q.c.$ e di conseguenza si ha $\{\mathrm{E}[YZ \mid \mathcal{F}_0] = 0\}$ $q.c.$, ovvero dalla prima parte del Teorema si ottiene

$$\mathrm{E}[YZ] = \mathrm{E}[\mathrm{E}[YZ \mid \mathcal{F}_0]] = 0 \,.$$

Essendo $\{\mathrm{E}[Y \mid \mathcal{F}_0] = 0\}$ $q.c.$, risulta

$$\mathrm{E}[Y^2] = \mathrm{E}[\mathrm{E}[Y^2 \mid \mathcal{F}_0]] = \mathrm{E}[\mathrm{Var}[Y \mid \mathcal{F}_0]] = \mathrm{E}[\mathrm{Var}[X \mid \mathcal{F}_0]] \,,$$

dal momento che

$$\mathrm{Var}[X \mid \mathcal{F}_0] = \mathrm{E}[X^2 \mid \mathcal{F}_0] - \mathrm{E}[X \mid \mathcal{F}_0]^2 = \mathrm{E}[(X - \mathrm{E}[X \mid \mathcal{F}_0])^2 \mid \mathcal{F}_0] \,.$$

Infine, dalla prima parte del Teorema si ha

$$\mathrm{E}[Z^2] = \mathrm{E}[(\mathrm{E}[X \mid \mathcal{F}_0] - \mathrm{E}[X])^2] = \mathrm{E}[(\mathrm{E}[X \mid \mathcal{F}_0] - \mathrm{E}[X])^2 \mid \mathcal{F}_0] = \mathrm{Var}[\mathrm{E}[X \mid \mathcal{F}_0]] \,,$$

da cui segue la tesi. □

Si osservi che dal Teorema 5.2.4 si ha

$$X = Y + \mathrm{E}[X \mid \mathcal{F}_0] \,,$$

con $Y = X - \mathrm{E}[X \mid \mathcal{F}_0]$ e dove $\mathrm{E}[Y\mathrm{E}[X \mid \mathcal{F}_0]] = 0$. In altri termini, la v.a. $\mathrm{E}[X \mid \mathcal{F}_0]$ può essere considerata come la "proiezione" della v.a. X su $\mathcal{F}_0$ con il relativo "complemento ortogonale" dato dalla v.a. Y.

Teorema 5.2.5. *Sia dato un v.v.a. $X = (X_1, X_2)^{\mathrm{T}}$ definito sullo spazio probabilizzato $(\Omega, \mathcal{F}, P)$ e sia $\mathcal{F}_0 \subset \mathcal{F}$ una σ-algebra. Se $\mathrm{E}[X_1^2]$ e $\mathrm{E}[X_2^2]$ esistono finiti, e se la v.a. X_1 è tale che $X_1^{-1}(B) \in \mathcal{F}_0$ per ogni $B \in \mathcal{B}(\mathbb{R})$, si ha*

$$\mathrm{E}[(X_2 - X_1)^2] = \mathrm{E}[\mathrm{Var}[X_2 \mid \mathcal{F}_0]] + \mathrm{E}[(\mathrm{E}[X_2 \mid \mathcal{F}_0] - X_1)^2] \,.$$

Inoltre, si ha

$$\mathrm{E}[(X_2 - X_1)^2] \geq \mathrm{E}[\mathrm{Var}[X_2 \mid \mathcal{F}_0]] \,,$$

dove l'uguaglianza si ottiene se $\{X_1 = \mathrm{E}[X_2 \mid \mathcal{F}_0]\}$ $q.c.$

Dimostrazione. Posto $Y = X_2 - \mathrm{E}[X_2 \mid \mathcal{F}_0]$ e $Z = \mathrm{E}[X_2 \mid \mathcal{F}_0] - X_1$, dal momento che $X_1^{-1}(B) \in \mathcal{F}_0$ per ogni $B \in \mathcal{B}(\mathbb{R})$, la prima parte si dimostra in modo simile al Teorema 5.2.4. La seconda parte è immediata, dal momento che $\mathrm{E}[(\mathrm{E}[X_2 \mid \mathcal{F}_0] - X_1)^2] \geq 0$ e l'uguaglianza è ottenuta quando $\{\mathrm{E}[X_2 \mid \mathcal{F}_0] - X_1 = 0\}$ $q.c.$ □

Si osservi che il Teorema 5.2.5 fornisce in effetti una versione generale del Teorema di Rao-Blackwell comunemente adottato nell'inferenza statistica. Il Teorema prende il nome da due famosi statistici, ovvero David Harold Blackwell (1919-2010) e Calyampudi Radhakrishna Rao (1920-). Per maggiori informazioni sul Teorema di Rao-Blackwell, si veda Lehmann e Casella (1998).

Teorema 5.2.6. *Si consideri la v.a. X definita sullo spazio probabilizzato $(\Omega, \mathcal{F}, P)$ e si consideri le σ-algebre $\mathcal{F}_0$ e $\mathcal{F}_1$ tali che $\mathcal{F}_0 \subset \mathcal{F}_1 \subset \mathcal{F}$. Se $\mathrm{E}[X]$ esiste finito, allora si ha $\{\mathrm{E}[\mathrm{E}[X \mid \mathcal{F}_1] \mid \mathcal{F}_0] = \mathrm{E}[X \mid \mathcal{F}_0]\}$ $q.c.$*

Dimostrazione. Posto $Y = \mathrm{E}[X \mid \mathcal{F}_1]$, tenendo presente la disuguaglianza di Jensen condizionata (Teorema 5.2.3, *i*), si ha

$$\mathrm{E}[Y] \leq \mathrm{E}[|Y|] = \mathrm{E}[|\mathrm{E}[X \mid \mathcal{F}_1]|] \leq \mathrm{E}[\mathrm{E}[|X| \mid \mathcal{F}_1]] = \mathrm{E}[|X|] < \infty \,.$$

Dunque, per ogni $E_0 \in \mathcal{F}_0$ dalla Definizione 5.1.3 si ha

$$\int_{E_0} \mathrm{E}[Y \mid \mathcal{F}_0] \, dP = \int_{E_0} Y \, dP \,.$$

Inoltre, di nuovo dalla Definizione 5.1.3 e tenendo presente che $E_0 \in \mathcal{F}_0 \subset \mathcal{F}_1$, si ha

$$\int_{E_0} \mathrm{E}[X \mid \mathcal{F}_1] \, dP = \int_{E_0} X \, dP \,.$$

Dalle precedenti espressioni si ha

$$\int_{E_0} \mathrm{E}[Y \mid \mathcal{F}_0] \, dP = \int_{E_0} X \, dP \,,$$

da cui segue la tesi. □

Il precedente Teorema è anche detto della "Proprietà della Torre" e generalizza in effetti la prima parte del Teorema 5.2.4.

5.3. Valore atteso condizionato ad una variabile aleatoria

Si osservi che la v.a. X_1 definita sullo spazio probabilizzato $(\Omega, \mathcal{F}, P)$ induce in effetti una σ-algebra $\mathcal{F}_{X_1} \subset \mathcal{F}$ data da

$$\mathcal{F}_{X_1} = \{E \in \mathcal{F} : E = X_1^{-1}(B), B \in \mathcal{B}(\mathbb{R})\} .$$

Di conseguenza, considerato il v.v.a. $X = (X_1, X_2)^{\mathrm{T}}$ definito sullo spazio probabilizzato $(\Omega, \mathcal{F}, P)$, si può introdurre il *valore atteso condizionato* della v.a. X_2 (alla v.a. X_1) come

$$\mathrm{E}[X_2 \mid X_1] = \mathrm{E}[X_2 \mid \mathcal{F}_{X_1}] .$$

Dunque, tenendo presente la Definizione 5.1.3, se $\mathrm{E}[X_2]$ esiste finito, $\mathrm{E}[X_2 \mid X_1]$ è un membro della classe di v.a. che coincidono $q.c.$ rispetto a P e tali che per ogni $B \in \mathcal{B}(\mathbb{R})$ soddisfano la relazione

$$\int_{X_1^{-1}(B)} \mathrm{E}[X_2 \mid X_1]\, dP = \int_{X_1^{-1}(B)} X_2\, dP .$$

Inoltre, dalla precedente relazione, si può verificare che esiste una funzione $h(x_1) = \mathrm{E}[X_2 \mid X_1 = x_1]$, detta *valore atteso condizionato* della v.a. X_2 all'evento $\{X_1 = x_1\}$, tale che se $P(X_1^{-1}(B)) > 0$ risulta

$$\int_{X_1^{-1}(B)} X_2\, dP = \int_{X_1^{-1}(B)} \mathrm{E}[X_2 \mid X_1 = x_1]\, dP = \int_B h(x_1)\, dP_{X_1}(x_1) .$$

Si assuma ora che il v.v.a. X sia discreto a valori su insieme numerabile $S \in \mathbb{R}^2$, con f.p.c. data da $p_X(x_1, x_2)$ e f.p.m. date da $p_{X_1}(x_1)$ e $p_{X_2}(x_2)$. Siano inoltre S_1 e S_2 le proiezioni di S sugli assi cartesiani. Per $B = \{x_1\}$, dove ovviamente $x_1 \in S_1$, si ha

$$\int_{X_1^{-1}(B)} X_2\, dP = \sum_{x_2 \in S_2} x_2 p_X(x_1, x_2) ,$$

mentre

$$\int_B h(x_1)\, dP_{X_1}(x_1) = \mathrm{E}[X_2 \mid X_1 = x_1] p_{X_1}(x_1) .$$

Dunque, sulla base delle precedenti relazioni il valore atteso condizionato della v.a. X_2 all'evento $\{X_1 = x_1\}$ con $x_1 \in S_1$, è in questo caso dato da

$$\mathrm{E}[X_2 \mid X_1 = x_1] = \sum_{x_2 \in S_2} x_2 \, \frac{p_X(x_1, x_2)}{p_{X_1}(x_1)} .$$

• **Esempio 5.3.1.** Si consideri il v.v.a. discreto $X = (X_1, X_2)^{\mathrm{T}}$ dell'Esempio 3.5.1. Si noti che $S = \{(0,0), (1,0), (0,1)\}$, mentre $S_1 = S_2 = \{0, 1\}$. Dunque, dal momento che $\mathrm{E}[X_2] = \frac{1}{2} < \infty$, si ha

$$\mathrm{E}[X_2 \mid X_1 = 0] = \sum_{x_2 \in S_2} x_2 \, \frac{p_X(0, x_2)}{p_{X_1}(0)} = \frac{2}{3}$$

e

$$\mathrm{E}[X_2 \mid X_1 = 1] = \sum_{x_2 \in S_2} x_2 \, \frac{p_X(1, x_2)}{p_{X_1}(1)} = 0 \, . \qquad \square$$

Si assuma invece che il v.v.a. X sia assolutamente continuo con d.p.c. data da $f_X(x_1, x_2)$ e d.p.m. date da $f_{X_1}(x_1)$ e $f_{X_2}(x_2)$. Inoltre, si consideri l'insieme

$$D = \{(x_1, x_2) \in \mathbb{R}^2 : f_{X_1}(x_1) = 0\} \, .$$

Risulta $P(X \in D) = 0$ e quindi per ogni $B \in \mathcal{B}(\mathbb{R})$, grazie al Teorema di Fubini (Teorema A.11), si ha

$$\begin{aligned} \int_{X_1^{-1}(B)} X_2 \, dP &= \int_{\mathbb{R} \times B} x_2 f_X(x_1, x_2) \, dx_1 dx_2 = \int_{B \setminus D} \left(\int_{\mathbb{R}} x_2 \, \frac{f_X(x_1, x_2)}{f_{X_1}(x_1)} \, dx_2 \right) f_{X_1}(x_1) \, dx_1 \\ &= \int_B \left(\int_{\mathbb{R}} x_2 \, \frac{f_X(x_1, x_2)}{f_{X_1}(x_1)} \, dx_2 \right) f_{X_1}(x_1) \, dx_1 = \int_B h(x_1) \, dP_{X_1}(x_1) \end{aligned}$$

e dunque

$$\mathrm{E}[X_2 \mid X_1 = x_1] = \int_{\mathbb{R}} x_2 \, \frac{f_X(x_1, x_2)}{f_{X_1}(x_1)} \, dx_2 \, .$$

• **Esempio 5.3.2.** Si consideri il v.v.a. assolutamente continuo $X = (X_1, X_2)^{\mathrm{T}}$ dell'Esempio 3.5.2. Dall'Esempio 4.2.3 risulta $\mathrm{E}[X_2] = \frac{1}{5} < \infty$ e dunque, se $x_1 \in \,]0, 1[$, si ha

$$\mathrm{E}[X_2 \mid X_1 = x_1] = \int_0^{1-x_1} \frac{2x_2}{1 - x_1} \left(1 - \frac{x_2}{1 - x_1} \right) dx_2 = \frac{1 - x_1}{3} \, ,$$

ovvero in questo caso il valore atteso condizionato della v.a. X_2 all'evento $\{X_1 = x_1\}$ è una funzione lineare di x_1. In modo simile, dal momento che dall'Esempio 4.2.3 risulta $\mathrm{E}[X_1] = \frac{2}{5} < \infty$, se $x_2 \in \,]0, 1[$, si ottiene

$$\mathrm{E}[X_1 \mid X_2 = x_2] = \int_0^{1-x_2} \frac{6x_1^2}{(1 - x_2)^2} \left(1 - \frac{x_1}{1 - x_2} \right) dx_1 = \frac{1 - x_2}{2} \, ,$$

ovvero il valore atteso condizionato della v.a. X_1 all'evento $\{X_2 = x_2\}$ è a sua volta una funzione lineare di x_2. $\square$

Nel caso in cui $B = \mathbb{R}$, dal Teorema 5.2.4 si ottiene

$$\mathrm{E}[X_2] = \mathrm{E}[\mathrm{E}[X_2 \mid X_1]]$$

se $\mathrm{E}[X_2]$ esiste finito. Inoltre, posto $\mathrm{Var}[X_2 \mid X_1] = \mathrm{E}[X_2^2 \mid X_1] - \mathrm{E}[X_2 \mid X_1]^2$, di nuovo dal Teorema 5.2.4 si ha

$$\mathrm{Var}[X_2] = \mathrm{E}[\mathrm{Var}[X_2 \mid X_1]] + \mathrm{Var}[\mathrm{E}[X_2 \mid X_1]] \, ,$$

se $\mathrm{E}[X_2^2]$ esiste finito. Infine, dal Teorema 5.2.5 si ha

$$\mathrm{E}[(X_2 - X_1)^2] = \mathrm{E}[\mathrm{Var}[X_2 \mid X_1]] + \mathrm{E}[(\mathrm{E}[X_2 \mid X_1] - X_1)^2]$$

e

$$\mathrm{E}[(X_2 - X_1)^2] \geq \mathrm{E}[\mathrm{Var}[X_2 \mid X_1]] \, ,$$

dove l'uguaglianza si ottiene se $\{X_1 = \mathrm{E}[X_2 \mid X_1]\}$ $q.c.$

• **Esempio 5.3.3.** Si consideri il v.v.a. assolutamente continuo $X = (X_1, X_2)^{\mathrm{T}}$ dell'Esempio 3.5.2. Tenendo presente l'Esempio 5.3.2. e l'Esempio 4.2.3, si ha

$$\mathrm{E}[X_2] = \mathrm{E}[\mathrm{E}[X_2 \mid X_1]] = \frac{1}{3}\,(1 - \mathrm{E}[X_1]) = \frac{1}{5}\,,$$

coerentemente con quanto visto nell'Esempio 4.2.3. In modo simile, si ottiene

$$\mathrm{E}[X_1] = \mathrm{E}[\mathrm{E}[X_1 \mid X_2]] = \frac{1}{2}\,(1 - \mathrm{E}[X_2]) = \frac{2}{5}\,,$$

in accordo con quanto visto nell'Esempio 4.2.3. Inoltre, si ha

$$\mathrm{E}[X_2^2 \mid X_1 = x_1] = \int_0^{1-x_1} \frac{2x_2^2}{1-x_1}\left(1 - \frac{x_2}{1-x_1}\right) dx_2 = \frac{(1-x_1)^2}{6}\,,$$

da cui

$$\mathrm{Var}[X_2 \mid X_1] = \frac{(1-X_1)^2}{6} - \frac{(1-X_1)^2}{9} = \frac{(1-X_1)^2}{18}\,.$$

Inoltre, tenendo presente l'Esempio 4.4.2, si ha

$$\mathrm{E}[\mathrm{Var}[X_2 \mid X_1]] = \frac{1}{18}\,\mathrm{E}[(1-X_1)^2] = \frac{1}{18}\,(1 - 2\,\mathrm{E}[X_1] + \mathrm{E}[X_1^2]) = \frac{1}{45}$$

e

$$\mathrm{Var}[\mathrm{E}[X_2 \mid X_1]] = \frac{1}{9}\,\mathrm{Var}[1 - X_1] = \frac{1}{9}\,\mathrm{Var}[X_1] = \frac{1}{225}\,,$$

da cui risulta

$$\mathrm{Var}[X_2] = \mathrm{E}[\mathrm{Var}[X_2 \mid X_1]] + \mathrm{Var}[\mathrm{E}[X_2 \mid X_1]] = \frac{2}{75}\,,$$

in accordo con i risultati ottenuti nell'Esempio 4.4.2. □

• **Esempio 5.3.4.** Si consideri il v.v.a. assolutamente continuo $X = (X_1, X_2)^{\mathrm{T}}$ che ammette la seguente d.p.c.

$$f_X(x_1, x_2) = \frac{1}{2\pi|x_1|}\, e^{-\frac{1}{2x_1^2}(x_2 - x_1)^2 - \frac{1}{2}(x_1 - 1)^2}\, \mathbf{1}_{\mathbb{R}\setminus\{0\}}(x_1)\,.$$

Sulla base dei risultati ottenuti, si può ottenere il valore atteso $\mathrm{E}[X_2]$ senza calcolare esplicitamente la d.p.m. della v.a. X_2 (che in questo caso è un'operazione abbastanza complessa). Tenendo presente le proprietà della legge Normale descritte nella Sezione 6.7, la d.p.m. della v.a. X_1 risulta

$$\begin{aligned} f_{X_1}(x_1) &= \int_{-\infty}^{\infty} f_X(x_1, x_2)\, dx_2 = \frac{1}{\sqrt{2\pi}}\, e^{-\frac{1}{2}(x_1 - 1)^2} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}|x_1|}\, e^{-\frac{1}{2x_1^2}(x_2 - x_1)^2}\, dx_2 \\ &= \frac{1}{\sqrt{2\pi}}\, e^{-\frac{1}{2}(x_1 - 1)^2}\,, \end{aligned}$$

ovvero la v.a. X_1 si distribuisce con legge Normale di parametri $\mu = 1$ e $\sigma = 1$ (si veda la Sezione 6.7). Si ha

$$\mathrm{E}[X_2 \mid X_1 = x_1] = \int_{-\infty}^{\infty} x_2 \, \frac{f_X(x_1, x_2)}{f_{X_1}(x_1)} \, dx_2 = \int_{-\infty}^{\infty} \frac{x_2}{\sqrt{2\pi}|x_1|} \, e^{-\frac{1}{2x_1^2}(x_2 - x_1)^2} \, dx_2 = x_1 \, ,$$

ovvero

$$\mathrm{E}[X_2] = \mathrm{E}[\mathrm{E}[X_2 \mid X_1]] = \mathrm{E}[X_1] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x_1 \, e^{-\frac{1}{2}(x_1 - 1)^2} \, dx_1 = 1 \, ,$$

e dunque il valore atteso della v.a. X_2 è stato ottenuto senza conoscere la legge corrispondente. Evidentemente, si può calcolare anche la varianza della v.a. X_2 senza determinarne la legge. □

5.4. Leggi condizionate

Si consideri il v.v.a. $X = (X_1, X_2)^{\mathrm{T}}$ definito sullo spazio probabilizzato $(\Omega, \mathcal{F}, P)$. Si supponga inizialmente che il v.v.a. X sia discreto a valori su insieme numerabile $S \in \mathbb{R}^2$, con f.p.c. data da $p_X(x_1, x_2)$ e f.p.m. date da $p_{X_1}(x_1)$ e $p_{X_2}(x_2)$. Siano inoltre S_1 e S_2 le proiezioni di S sugli assi cartesiani. Dato l'evento $\{X_1 = x_1\}$ con $x_1 \in S_1$, $P(\,\cdot \mid X_1 = x_1)$ è in effetti una misura di probabilità (vedi Sezione 2.4) e per $B \in \mathcal{B}(\mathbb{R})$

$$P_{X_2|X_1=x_1}(B) = P(X_2^{-1}(B) \mid X_1 = x_1) = \frac{P(X_2^{-1}(B) \cap \{X_1 = x_1\})}{P(X_1 = x_1)}$$

è detta *legge condizionata* della v.a. X_2 all'evento $\{X_1 = x_1\}$. Dunque, si può definire la *f.p. condizionata* della v.a. X_2 all'evento $\{X_1 = x_1\}$, con $x_1 \in S_1$, come

$$p_{X_2|X_1=x_1}(x_2) = P(X_2 = x_2 \mid X_1 = x_1) = \frac{p_X(x_1, x_2)}{p_{X_1}(x_1)} \, .$$

Risulta facile verificare che $p_{X_2|X_1=x_1}(x_2)$ è in effetti una f.p. e che la sua definizione è coerente con quella di valore atteso condizionato della v.a. X_2 all'evento $\{X_1 = x_1\}$, dal momento che, per quanto visto nella Sezione 5.4, risulta

$$\mathrm{E}[X_2 \mid X_1 = x_1] = \sum_{x_2 \in S_2} x_2 p_{X_2|X_1=x_1}(x_2) \, .$$

• **Esempio 5.4.1.** Si consideri il v.v.a. discreto $X = (X_1, X_2)^{\mathrm{T}}$ dell'Esempio 3.5.1. La f.p. condizionata della v.a. X_2 all'evento $\{X_1 = 0\}$ è data dunque da

$$p_{X_2|X_1=0}(x_2) = \frac{p_X(0, x_2)}{p_{X_1}(0)} = \left(\frac{2}{3}\right)^{x_2} \left(\frac{1}{3}\right)^{1-x_2} \mathbf{1}_{\{0,1\}}(x_2) \, ,$$

mentre la f.p. condizionata della v.a. X_2 all'evento $\{X_1 = 1\}$ risulta

$$p_{X_2|X_1=1}(x_1) = \frac{p_X(1, x_2)}{p_{X_1}(1)} = \mathbf{1}_{\{0\}}(x_2) \, ,$$

ovvero le leggi associate alle v.a. condizionate si ottengono per particolari parametrizzazioni della legge Binomiale (si veda Sezione 6.5). □

Il modo con cui è stata costruita la legge condizionata nel caso di un v.v.a. X discreto perde di senso quando l'evento $\{X_1 = x_1\}$ è di probabilità nulla. In generale, non sempre esiste una legge condizionata della v.a. X_2 all'evento $\{X_1 = x_1\}$, anche se gli esempi di non esistenza sono artificiosi

e non fondamentali nell'uso della teoria (per maggiori dettagli, si veda Wise e Hall, 1993, p.159). Nel caso in cui il v.v.a. X sia assolutamente continuo con d.p.c. data da $f_X(x_1, x_2)$ e d.p.m. date da $f_{X_1}(x_1)$ e $f_{X_2}(x_2)$, si può definire la ***d.p. condizionata*** della v.a. X_2 all'evento $\{X_1 = x_1\}$ come

$$f_{X_2|X_1=x_1}(x_2) = \frac{f_X(x_1, x_2)}{f_{X_1}(x_1)} \mathbf{1}_{\mathbb{R}^2 \setminus D}(x_1, x_2) ,$$

(si veda la Sezione 5.3 per la definizione dell'insieme D). Il Teorema di Radon-Nikodym (Teorema A.9) assicura in questo caso l'esistenza della *legge condizionata* della v.a. X_2 all'evento $\{X_1 = x_1\}$, ovvero per ogni $B \in \mathcal{B}(\mathbb{R})$ si ha

$$P_{X_2|X_1=x_1}(B) = \int_B f_{X_2|X_1=x_1}(x_2)\, dx_2 .$$

Si dovrebbe evidenziare che la d.p. condizionata non è unica e quindi, al solito, si dovrebbe parlare di una versione della densità. Inoltre, la sua definizione è coerente con quella di valore atteso condizionato della v.a. X_2 all'evento $\{X_1 = x_1\}$, dal momento che per quanto visto nella Sezione 5.4 risulta

$$\mathrm{E}[X_2 \mid X_1 = x_1] = \int_{\mathbb{R}} x_2 f_{X_2|X_1=x_1}(x_2)\, dx_2 .$$

• **Esempio 5.4.2.** Si consideri il v.v.a. assolutamente continuo $X = (X_1, X_2)^{\mathrm{T}}$ dell'Esempio 3.5.2. La d.p. condizionata della v.a. X_2 all'evento $\{X_1 = x_1\}$ è data dunque da

$$f_{X_2|X_1=x_1}(x_2) = \frac{f_X(x_1, x_2)}{f_{X_1}(x_1)} = \frac{2}{1 - x_1} \left(1 - \frac{x_2}{1 - x_1}\right) \mathbf{1}_{]0,1[}(x_1) \mathbf{1}_{]0,1-x_1[}(x_2) ,$$

mentre la d.p. condizionata della v.a. X_1 all'evento $\{X_2 = x_2\}$ risulta

$$f_{X_1|X_2=x_2}(x_1) = \frac{f_X(x_1, x_2)}{f_{X_2}(x_2)} = \frac{6x_1}{(1 - x_2)^2} \left(1 - \frac{x_1}{1 - x_2}\right) \mathbf{1}_{]0,1[}(x_2) \mathbf{1}_{]0,1-x_2[}(x_1) ,$$

ovvero le leggi associate alle v.a. condizionate si ottengono per particolari parametrizzazioni della legge Beta (si veda la Sezione 6.12). □

• **Esempio 5.4.3.** Sia data la legge Uniforme sul cerchio unitario introdotta nell'Esempio 3.6.3. La d.p. condizionata della v.a. X_1 all'evento $\{X_2 = x_2\}$ è data dunque da

$$f_{X_2|X_1=x_1}(x_2) = \frac{f_X(x_1, x_2)}{f_{X_1}(x_1)} = \frac{1}{2\sqrt{1 - x_1^2}} \mathbf{1}_{]-\sqrt{1-x_1^2}, \sqrt{1-x_1^2}[}(x_2) ,$$

mentre per simmetria la d.p. condizionata della v.a. X_1 all'evento $\{X_2 = x_2\}$ risulta

$$f_{X_1|X_2=x_2}(x_1) = \frac{1}{2\sqrt{1 - x_2^2}} \mathbf{1}_{]-\sqrt{1-x_2^2}, \sqrt{1-x_2^2}[}(x_1) ,$$

ovvero le leggi associate alle v.a. condizionate sono Uniformi (si veda la Sezione 6.12). □

5.5. Riferimenti bibliografici

Una trattazione approfondita del concetto di valore atteso condizionato è contenuto in Williams (1992), Rao (2005) e Steyer e Nagel (2017). Altri testi dove è possibile trovare un'esauriente analisi dei problemi tecnici legati al valore atteso condizionato sono Cohen e Elliott (2015), Dudley (2004),

Kallenberg (2002) e Stroock (2013). Per controesempi nell'ambito del valore atteso condizionato si veda Wise e Hall (1993).

5.6. Esercizi svolti

• **Nota.** Al fine di evitare notazioni ridondanti e ripetizioni, nei seguenti esercizi si assume l'esistenza di uno spazio probabilizzato $(\Omega, \mathcal{F}, P)$ e di una σ-algebra $\mathcal{F}_0 \subset \mathcal{F}$, a meno che non venga specificato diversamente.

Sezione 5.1

• **Esercizio 5.1.1.** Sia dato lo spazio probabilizzato $(\Omega, \mathcal{F}, P)$, dove $\Omega =]0,1]$ e $\mathcal{F} = \mathcal{B}(]0,1])$, mentre $P = \lambda$ è la misura di Lebesgue su $]0,1]$. Si consideri inoltre la partizione $\mathcal{E}$ di Ω tale che $\mathcal{E} = \{B_1, \ldots, B_n\}$, dove $B_k =]a_{k-1}, a_k]$ e $a_0 = 0 < a_1 < \ldots < a_n = 1$, e sia $\mathcal{F}_0 = \mathcal{P}(\mathcal{E}) \subset \mathcal{F}$ la σ-algebra generata da $\mathcal{E}$. Se $E \in \mathcal{F}$, si determini la probabilità condizionata $P(E \mid B_k) = \mathrm{E}[\mathbf{1}_E \mid B_k]$, dove $k = 1, \ldots, n$, e $P(E \mid \mathcal{F}_0) = \mathrm{E}[\mathbf{1}_E \mid \mathcal{F}_0]$.

Soluzione. Si noti che risulta

$$P(B_k) = \lambda(]a_{k-1}, a_k]) = a_k - a_{k-1}\,.$$

Dunque, dalla definizione di valore atteso condizionato ad un evento si ha

$$P(E \mid B_k) = \frac{1}{a_k - a_{k-1}} \int_{a_{k-1}}^{a_k} \mathbf{1}_E(x)\, dx = \frac{\lambda(E \cap B_k)}{a_k - a_{k-1}}\,.$$

Inoltre, tenendo presente la definizione di valore atteso condizionato alla σ-algebra $\mathcal{F}_0$, $P(E \mid \mathcal{F}_0)$ è data da

$$P(E \mid \mathcal{F}_0) = \sum_{k=1}^{n} P(E \mid B_k)\, \mathbf{1}_{B_k} = \sum_{k=1}^{n} \frac{\lambda(E \cap B_k)}{a_k - a_{k-1}}\, \mathbf{1}_{B_k}\,.$$

Quindi, $P(E \mid \mathcal{F}_0)$ è una v.a. discreta che assume valori su $S = (P(E \mid B_k))_{k=1}^n$ con legge essenziale data dall'insieme delle probabilità $(P(B_k))_{k=1}^n$. Inoltre, essendo $\mathrm{E}[\mathbf{1}_E] = P(B_k)$ si ha

$$\begin{aligned}\mathrm{E}[P(E \mid \mathcal{F}_0)] &= \sum_{k=1}^{n} \frac{\lambda(E \cap B_k)}{a_k - a_{k-1}}\, \mathrm{E}[\mathbf{1}_{B_k}] = \sum_{k=1}^{n} \lambda(E \cap B_k) \\ &= \lambda\left(E \cap \bigcup_{k=1}^{n} B_k\right) = \lambda(E \cap \Omega) = \lambda(E) = P(E)\,,\end{aligned}$$

dal momento che gli eventi di $\mathcal{E}$ sono incompatibili. □

• **Esercizio 5.1.2.** Si assuma che la v.a. X e la σ-algebra $\mathcal{F}_0$ siano indipendenti, ovvero che

$$P(\{X \le x\} \cap E_0) = P(X \le x) P(E_0)$$

per ogni $x \in \mathbb{R}$ e $E_0 \in \mathcal{F}_0$. Si verifichi che risulta $\{\mathrm{E}[X \mid \mathcal{F}_0] = \mathrm{E}[X]\}$ *q.c.*

Soluzione. Dalla definizione di valore atteso condizionato, per ogni $E_0 \in \mathcal{F}_0$ si ha

$$\int_{E_0} \mathrm{E}[X \mid \mathcal{F}_0]\, dP = \int_{E_0} X\, dP = \int_{\Omega} X\, \mathbf{1}_{E_0}\, dP = \mathrm{E}[X\, \mathbf{1}_{E_0}]\,.$$

Dalle assunzioni fatte la v.a. X e la v.a. $\mathbf{1}_{E_0}$ sono indipendenti e quindi per le proprietà del valore atteso si ha

$$\mathrm{E}[X\,\mathbf{1}_{E_0}] = \mathrm{E}[X]\mathrm{E}[\mathbf{1}_{E_0}]\,,$$

da cui

$$\int_{E_0} \mathrm{E}[X \mid \mathcal{F}_0]\,dP = \mathrm{E}[X]\int_\Omega \mathbf{1}_{E_0}\,dP = \int_\Omega \mathrm{E}[X]\,\mathbf{1}_{E_0}\,dP = \int_{E_0} \mathrm{E}[X]\,dP\,.$$

Dunque, dalla precedente relazione segue l'identità richiesta. □

Sezione 5.2

• **Esercizio 5.2.1.** Sia dato lo spazio probabilizzato $(\Omega, \mathcal{F}, P)$, dove $\Omega =]0,1]$ e $\mathcal{F} = \mathcal{B}(]0,1])$, mentre $P = \lambda$ è la misura di Lebesgue su $]0,1]$ e si consideri inoltre la σ-algebra $\mathcal{F}_0 = \{\emptyset, E, E^c, \Omega\}$ dove $E =]0, \frac{1}{3}]$. Se $X = (X_1, X_2)^{\mathrm{T}}$ è un v.v.a. tale che $X_1(\omega) = \mathbf{1}_E(\omega)$ e $X_2(\omega) = \omega$ per $\omega \in \Omega$, si determini il valore atteso $\mathrm{E}[(X_2 - X_1)^2]$ e i valori attesi $\mathrm{E}[\mathrm{Var}[X_2 \mid \mathcal{F}_0]]$ e $\mathrm{E}[(\mathrm{E}[X_2 \mid \mathcal{F}_0] - X_1)^2]$. Si verifichi inoltre l'identità fra il primo valore atteso e la somma dei rimanenti valori attesi.

Soluzione. Tenendo presente l'Esercizio 3.1.1, si osservi innanzitutto che $X_1^{-1}(B) \in \mathcal{F}_0$ per ogni $B \in \mathcal{B}(\mathbb{R})$. Si ha

$$\mathrm{E}[(X_2 - X_1)^2] = \int_0^1 (\omega - \mathbf{1}_E(\omega))^2\,d\omega = \int_0^1 \omega^2\,d\omega - 2\int_0^{\frac{1}{3}} \omega\,d\omega + \int_0^{\frac{1}{3}} d\omega = \frac{5}{9}\,.$$

Inoltre, essendo $P(E) = \frac{1}{3}$ e $P(E^c) = \frac{2}{3}$, segue

$$\mathrm{E}[X_2 \mid E] = \frac{\mathrm{E}[X_2\,\mathbf{1}_E]}{P(E)} = 3\int_0^{\frac{1}{3}} \omega\,d\omega = \frac{1}{6}$$

e

$$\mathrm{E}[X_2 \mid E^c] = \frac{\mathrm{E}[X_2\,\mathbf{1}_{E^c}]}{P(E^c)} = \frac{3}{2}\int_{\frac{1}{3}}^1 \omega\,d\omega = \frac{2}{3}\,,$$

mentre

$$\mathrm{E}[X_2^2 \mid E] = \frac{\mathrm{E}[X_2^2\,\mathbf{1}_E]}{P(E)} = 3\int_0^{\frac{1}{3}} \omega^2\,d\omega = \frac{1}{27}$$

e

$$\mathrm{E}[X_2^2 \mid E^c] = \frac{\mathrm{E}[X_2^2\,\mathbf{1}_{E^c}]}{P(E^c)} = \frac{3}{2}\int_{\frac{1}{3}}^1 \omega^2\,d\omega = \frac{13}{27}\,.$$

Dal momento che $\mathcal{F}_0$ è una σ-algebra generata da una partizione, allora risulta

$$\mathrm{E}[X_2 \mid \mathcal{F}_0] = \mathrm{E}[X_2 \mid E]\,\mathbf{1}_E + \mathrm{E}[X_2 \mid E^c]\,\mathbf{1}_{E^c} = \frac{1}{6}\,\mathbf{1}_E + \frac{2}{3}\,\mathbf{1}_{E^c}$$

e

$$\mathrm{E}[X_2^2 \mid \mathcal{F}_0] = \mathrm{E}[X_2^2 \mid E]\,\mathbf{1}_E + \mathrm{E}[X_2^2 \mid E^c]\,\mathbf{1}_{E^c} = \frac{1}{27}\,\mathbf{1}_E + \frac{13}{27}\,\mathbf{1}_{E^c}\,,$$

da cui, tenendo presente che $\mathbf{1}_E\mathbf{1}_{E^c} = 0$ per ogni $\omega \in \Omega$,

$$\begin{aligned}\mathrm{Var}[X_2 \mid \mathcal{F}_0] &= \mathrm{E}[X_2^2 \mid \mathcal{F}_0] - \mathrm{E}[X_2 \mid \mathcal{F}_0]^2 \\ &= (\mathrm{E}[X_2^2 \mid E] - \mathrm{E}[X_2 \mid E]^2)\,\mathbf{1}_E + (\mathrm{E}[X_2^2 \mid E^c] - \mathrm{E}[X_2 \mid E^c]^2)\,\mathbf{1}_{E^c} \\ &= \frac{1}{108}\,\mathbf{1}_E + \frac{1}{27}\,\mathbf{1}_{E^c}\,.\end{aligned}$$

Quindi, si ha

$$\mathrm{E}[\mathrm{Var}[X_2 \mid \mathcal{F}_0]] = \mathrm{E}[\tfrac{1}{108}\,\mathbf{1}_E + \tfrac{1}{27}\,\mathbf{1}_{E^c}] = \frac{1}{108}\,P(E) + \frac{1}{27}\,P(E^c) = \frac{1}{36}\,.$$

Inoltre, risulta

$$\begin{aligned}\mathrm{E}[(\mathrm{E}[X_2 \mid \mathcal{F}_0] - X_1)^2] &= \mathrm{E}[(\tfrac{1}{6}\,\mathbf{1}_E + \tfrac{2}{3}\,\mathbf{1}_{E^c} - \mathbf{1}_E)^2] = \mathrm{E}[\tfrac{25}{36}\,\mathbf{1}_E + \tfrac{4}{9}\,\mathbf{1}_{E^c}] \\ &= \frac{25}{36}\,P(E) + \frac{4}{9}\,P(E^c) = \frac{19}{36}\end{aligned}$$

e quindi l'identità richiesta è verificata. □

Sezione 5.3

• **Esercizio 5.3.1.** (Paradosso di Borel) Si consideri il v.v.a. assolutamente continuo $X = (X_1, X_2)^{\mathrm{T}}$ che ammette d.p.c. data da

$$f_X(x_1, x_2) = \mathbf{1}_{[0,1]}(x_1 + x_2)\mathbf{1}_{[0,1]}(x_2)$$

e si verifichi che risulta $\mathrm{E}[X_2 \mid X_1 = 0] = \frac{1}{2}$. Si consideri inoltre il v.v.a. $Y = (Y_1, Y_2)^{\mathrm{T}}$ dove $Y_1 = \frac{X_1+X_2}{X_2}$ e $Y_2 = X_2$ e si verifichi che si ha $\mathrm{E}[Y_2 \mid Y_1 = 1] = \frac{2}{3}$.

Soluzione. La d.p. marginale della v.a. X_1 risulta

$$\begin{aligned}f_{X_1}(x_1) &= \int_{-\infty}^{\infty} \mathbf{1}_{[0,1]}(x_1 + x_2)\mathbf{1}_{[0,1]}(x_2)\,dx_2 \\ &= \int_{-x_1}^{1} \mathbf{1}_{[-1,0]}(x_1)\,dx_2 + \int_{0}^{1-x_1} \mathbf{1}_{[0,1]}(x_1)\,dx_2 = (1 - |x_1|)\,\mathbf{1}_{[-1,1]}(x_1)\,,\end{aligned}$$

da cui

$$\begin{aligned}\mathrm{E}[X_2 \mid X_1 = x_1] &= \int_{-\infty}^{\infty} x_2\,\frac{\mathbf{1}_{[0,1]}(x_1 + x_2)\mathbf{1}_{[0,1]}(x_2)}{(1 - |x_1|)\,\mathbf{1}_{[-1,1]}(x_1)}\,dx_2 \\ &= \frac{\mathbf{1}_{[-1,0]}(x_1)}{1 + x_1}\int_{x_1}^{1} x_2\,dx_2 + \frac{\mathbf{1}_{[0,1]}(x_1)}{1 - x_1}\int_{0}^{1-x_1} x_2\,dx_2 = \frac{1 - x_1}{2}\,\mathbf{1}_{[-1,1]}(x_1)\,.\end{aligned}$$

Quindi, si ha $\mathrm{E}[X_2 \mid X_1 = 0] = \frac{1}{2}$. Per quanto riguarda la trasformata $Y = g(X)$, si ha

$$g^{-1}(y) = (y_2(y_1 - 1), y_2)^{\mathrm{T}}\,,$$

dove $y_1 \in [1, \infty[$ e $y_2 \in [0, 1]$. Dal momento che $|J(g^{-1}(y))| = y_2$, la d.p.c. del v.v.a. Y è data da

$$f_Y(y_1, y_2) = y_2\,\mathbf{1}_{[0,1]}(y_1 y_2)\mathbf{1}_{[0,1]}(y_2)\,,$$

da cui

$$f_{Y_1}(y_1) = \int_{-\infty}^{\infty} y_2\,\mathbf{1}_{[0,1]}(y_1 y_2)\mathbf{1}_{[0,1]}(y_2)\,dy_2 = \int_{0}^{\frac{1}{y_1}} y_2\,\mathbf{1}_{[1,\infty[}(y_1)\,dy_2 = \frac{1}{2y_1^2}\,\mathbf{1}_{[1,\infty[}(y_1)\,.$$

Dunque, risulta

$$\begin{aligned}\mathrm{E}[Y_2 \mid Y_1 = y_1] &= \int_{-\infty}^{\infty} y_2 \, \frac{y_2 \, \mathbf{1}_{[0,1]}(y_1 y_2) \mathbf{1}_{[0,1]}(y_2)}{\frac{1}{2y_1^2} \, \mathbf{1}_{[1,\infty[}(y_1)} \, dy_2 \\ &= 2y_1^2 \, \mathbf{1}_{[1,\infty[}(y_1) \int_0^1 y_2^2 \, dx_2 = \frac{2}{3} \, y_1^2 \, \mathbf{1}_{[1,\infty[}(y_1) \, .\end{aligned}$$

Dalla definizione di Y_1, si osservi che in effetti l'evento $\{X_1 = 0\}$ è "equivalente" all'evento $\{Y_1 = 1\}$. In modo controintuitivo si ha quindi che

$$\mathrm{E}[Y_2 \mid Y_1 = 1] = \frac{2}{3} \neq \mathrm{E}[X_2 \mid X_1 = 0] = \frac{1}{2} \, ,$$

anche se $X_2 = Y_2$. L'apparente paradosso deriva dal fatto che i valori attesi condizionati sono stati ottenuti rispetto a eventi di probabilità nulla. □

• **Esercizio 5.3.2.** Si consideri il v.v.a. $X = (X_1, \ldots, X_n)^{\mathrm{T}}$ le cui componenti sono indipendenti e con la medesima legge. Si assuma inoltre che $Y = \sum_{k=1}^n X_k$. Si verifichi che $\{\mathrm{E}[X_k \mid Y] = n^{-1} Y\}$ $q.c.$ per ogni $k = 1, \ldots, n$.

Soluzione. Dalla definizione di valore atteso condizionato, per ogni $B \in \mathbb{R}$ e $k = 1, \ldots, n$ si ha

$$\int_{Y^{-1}(B)} \mathrm{E}[X_k \mid Y] \, dP = \mathrm{E}[X_k \, \mathbf{1}_{Y^{-1}(B)}] \, .$$

Dunque, dal momento che X_1 e X_k sono equivalenti in legge, risulta $\mathrm{E}[X_k \mathbf{1}_{Y^{-1}(B)}] = \mathrm{E}[X_1 \mathbf{1}_{Y^{-1}(B)}]$, ovvero tenendo presente la precedente relazione si ottiene

$$P(\mathrm{E}[X_k \mid Y] = \mathrm{E}[X_1 \mid Y]) = 1$$

per ogni $k = 2, \ldots, n$. Dunque, si ha $q.c.$

$$\mathrm{E}[X_k \mid Y] = \frac{1}{n} \sum_{j=1}^n \mathrm{E}[X_j \mid Y] = \frac{1}{n} \, \mathrm{E}[\sum_{j=1}^n X_j \mid Y] = \frac{1}{n} \, \mathrm{E}[Y \mid Y] = \frac{1}{n} \, Y \, .$$

Questo risultato può essere esemplificato mediante la seguente applicazione statistica. Se la v.a. X_k si distribuisce con legge di Bernoulli di parametro pari a p, allora $\mathrm{E}[X_k] = p$ e $\mathrm{Var}[X_k] = pq$. Inoltre, è immediato verificare che la v.a. Y si distribuisce con legge Binomiale con parametri n e p (si veda la Sezione 6.1) e quindi $\mathrm{E}[Y] = np$ e $\mathrm{Var}[Y] = npq$. Dunque, dalla precedente relazione si ha

$$\mathrm{E}[\mathrm{E}[X_k \mid Y]] = \frac{1}{n} \, \mathrm{E}[Y] = p$$

e

$$\mathrm{Var}[\mathrm{E}[X_k \mid Y]] = \frac{1}{n^2} \, \mathrm{Var}[Y] = \frac{pq}{n} \, .$$

Quindi, nella terminologia della statistica inferenziale, sia la v.a. X_k che la v.a. $\mathrm{E}[X_k \mid Y]$ sono stimatori corretti del parametro p. Inoltre, posto $x = (x_1, \ldots, x_n)^{\mathrm{T}}$, per $y = 0, 1, \ldots, n$ risulta

$$f_{X|Y=y}(x) = \frac{P(\{X = x\} \cap \{Y = y\})}{P(Y = y)} = \frac{\prod_{k=1}^n p^{x_k} (1-p)^{1-x_k}}{\binom{n}{y} p^y (1-p)^{n-y}} \, \mathbf{1}_{S_y}(x) = \binom{n}{y}^{-1} \mathbf{1}_{S_y}(x) \, ,$$

dove $S_y = \{(x_1, \ldots, x_n) : x_k = 0, 1, k = 1, \ldots, n, \sum_{k=1}^n x_k = y\}$, ovvero la v.a. Y è uno stimatore sufficiente del parametro p, dal momento che legge condizionata del v.v.a. X all'evento $\{Y = y\}$ non

dipende da p. Tuttavia, $\mathrm{Var}[X_k] > \mathrm{Var}[\mathrm{E}[X_k \mid Y]]$ e quindi lo stimatore $\mathrm{E}[X_k \mid Y]$ è più efficiente dello stimatore X_k. Questo risultato segue dal fatto che lo stimatore $\mathrm{E}[X_k \mid Y]$ viene ottenuto dal condizionamento di uno stimatore corretto ad uno stimatore sufficiente. In effetti, in questo esempio è stata considerata un'applicazione del Teorema di Rao-Blackwell. □

• **Esercizio 5.3.3.** Si consideri il v.v.a. assolutamente continuo $X = (X_1, X_2)^{\mathrm{T}}$ che ammette d.p.c. data da

$$f_X(x_1, x_2) = \frac{1}{\pi(1 + x_1^2 + x_2^2)^2}$$

e si determini $\mathrm{E}[X_2 \mid X_1 = x_1]$ e $\mathrm{Var}[X_2 \mid X_1 = x_1]$.

Soluzione. Si ha

$$f_{X_1}(x_1) = \int_{-\infty}^{\infty} \frac{1}{\pi(1 + x_1^2 + x_2^2)^2}\, dx_2 = \frac{1}{2\sqrt{(1 + x_1^2)^3}}\,,$$

da cui

$$\mathrm{E}[X_2 \mid X_1 = x_1] = \int_{-\infty}^{\infty} x_2\, \frac{2\sqrt{(1 + x_1^2)^3}}{\pi(1 + x_1^2 + x_2^2)^2}\, dx_2 = 0\,,$$

dal momento che la funzione integranda è dispari. Quindi, si ottiene

$$\mathrm{Var}[X_2 \mid X_1 = x_1] = \int_{-\infty}^{\infty} x_2^2\, \frac{2\sqrt{(1 + x_1^2)^3}}{\pi(1 + x_1^2 + x_2^2)^2}\, dx_2 = 1 + x_1^2\,.$$

Si osservi che

$$\mathrm{E}[X_2] = \int_{-\infty}^{\infty} x_2\, \frac{1}{2\sqrt{(1 + x_2^2)^3}}\, dx_2 = 0$$

e in effetti risulta $\mathrm{E}[\mathrm{E}[X_2 \mid X_1]] = 0$. Tuttavia, si ha

$$\mathrm{E}[X_2^2] = \int_{-\infty}^{\infty} x_2^2\, \frac{1}{2\sqrt{(1 + x_2^2)^3}}\, dx_2 = \infty$$

e quindi $\mathrm{Var}[X_2]$ non è finita. In effetti, essendo $X_1 \stackrel{\mathcal{L}}{=} X_2$ per simmetria, si ha

$$\mathrm{E}[\mathrm{Var}[X_2 \mid X_1]] = \mathrm{E}[1 + X_1^2] = 1 + \mathrm{E}[X_1^2] = \infty\,.$$

Dunque, può essere $\mathrm{Var}[X_2 \mid X_1 = x_1] < \infty$ per ogni x_1, anche se $\mathrm{E}[\mathrm{Var}[X_2 \mid X_1]] = \infty$. □

Sezione 5.4

• **Esercizio 5.4.1.** Si consideri il v.v.a. assolutamente continuo $X = (X_1, X_2)^{\mathrm{T}}$ che ammette d.p.c.

$$f_X(x_1, x_2) = x_1(x_2 - x_1)\, e^{-x_2}\, \mathbf{1}_{[0,\infty[}(x_1) \mathbf{1}_{[x_1,\infty[}(x_2)\,.$$

Si determini la legge condizionata della v.a. X_2 all'evento $\{X_1 = x_1\}$ e i valori attesi $\mathrm{E}[X_2 \mid X_1 = x_1]$ e $\mathrm{Var}[X_2 \mid X_1 = x_1]$.

Soluzione. Si ha

$$f_{X_1}(x_1) = \int_{x_1}^{\infty} x_1(x_2 - x_1)\, e^{-x_2}\, \mathbf{1}_{[0,\infty[}(x_1)\, dx_2 = x_1\, e^{-x_1}\, \mathbf{1}_{[0,\infty[}(x_1)\,,$$

da cui risulta

$$f_{X_2|X_1=x_1}(x_2) = \frac{f_X(x_1,x_2)}{f_{X_1}(x_1)} = (x_2 - x_1)\, e^{-(x_2-x_1)}\, \mathbf{1}_{[x_1,\infty[}(x_2)\,.$$

Inoltre, si ha

$$\mathrm{E}[X_2 \mid X_1 = x_1] = \int_{x_1}^{\infty} x_2(x_2 - x_1)\, e^{-(x_2-x_1)}\, dx_2 = x_1 + 2\,,$$

ovvero il valore atteso condizionato è una funzione lineare. Infine, si osservi che risulta

$$\mathrm{E}[X_2^2 \mid X_1 = x_1] = \int_{x_1}^{\infty} x_2^2(x_2 - x_1)\, e^{-(x_2-x_1)}\, dx_2 = x_1^2 + 4x_1 + 6\,,$$

da cui $\mathrm{Var}[X_2 \mid X_1 = x_1] = 2$. □

• **Esercizio 5.4.2.** Si consideri il v.v.a. assolutamente continuo $X = (X_1, X_2)^{\mathrm{T}}$ che ammette d.p.c.

$$f_X(x_1,x_2) = \frac{1}{3\pi}\, \mathbf{1}_{[1,4]}(x_1^2 + x_2^2)$$

e si determini la legge condizionata della v.a. X_2 all'evento $\{X_1 = x_1\}$.

Soluzione. Si ha

$$\begin{aligned} f_{X_1}(x_1) &= \frac{1}{3\pi} \int_{-\sqrt{4-x_1^2}}^{\sqrt{4-x_1^2}} \mathbf{1}_{[-2,2]}(x_1)\, dx_2 - \frac{1}{3\pi} \int_{-\sqrt{1-x_1^2}}^{\sqrt{1-x_1^2}} \mathbf{1}_{]-1,1[}(x_1)\, dx_2 \\ &= \frac{2}{3\pi}\, \sqrt{4 - x_1^2}\, \mathbf{1}_{[-2,2]}(x_1) - \frac{2}{3\pi}\, \sqrt{1 - x_1^2}\, \mathbf{1}_{]-1,1[}(x_1)\,, \end{aligned}$$

da cui risulta

$$f_{X_2|X_1=x_1}(x_2) = \frac{\mathbf{1}_{[1,4]}(x_1^2 + x_2^2)}{2\sqrt{4 - x_1^2}\, \mathbf{1}_{[-2,2]}(x_1) - 2\sqrt{1 - x_1^2}\, \mathbf{1}_{]-1,1[}(x_1)}\,,$$

dove $x_1 \in [-2, 2]$. Si noti che la legge condizionata è Uniforme se $|x_1| \in [1, 2]$, mentre è data dalla mistura di due leggi Uniformi se $x_1 \in [-1, 1]$. □

• **Esercizio 5.4.3.** (Leggi coniugate) Per quanto riguarda le notazioni adottate in questo esercizio, si veda la descrizione della legge Binomiale nella Sezione 6.1 e della legge Beta nella Sezione 6.9. Si consideri il v.v.a. $X = (X_1, X_2)^{\mathrm{T}}$ tale che la v.a. marginale X_1 si distribuisce con legge Beta ridotta con parametri pari a $\alpha, \beta \in]0, \infty[$, ovvero X_1 ammette d.p. data

$$f_{X_1}(x_1) = \frac{1}{B(\alpha,\beta)}\, x_1^{\alpha-1}(1 - x_1)^{\beta-1}\, \mathbf{1}_{]0,1[}(x_1)\,,$$

mentre la legge condizionata della v.a. X_2 all'evento $\{X_1 = x_1\}$ è Binomiale con parametri pari a n e x_1, ovvero la f.p. condizionata è data da

$$p_{X_2|X_1=x_1}(x_2) = \binom{n}{x_2}\, x_1^{x_2}(1 - x_1)^{n-x_2}\, \mathbf{1}_{\{0,1,\ldots,n\}}(x_2)\,.$$

Si determini la legge condizionata della v.a. X_1 all'evento $\{X_2 = x_2\}$.

Soluzione. Si assuma che B_1 e B_2 siano due eventi tali che $B_1, B_2 \in \mathcal{B}(\mathbb{R})$. Si osservi che

$$P(X_1 \in B_1) = \frac{1}{B(\alpha,\beta)} \int_{]0,1[\cap B_1} x_1^{\alpha-1}(1-x_1)^{\beta-1}\, dx_1 \,.$$

Inoltre, risulta

$$\begin{aligned} P(X_1 \in B_1, X_2 \in B_2) &= \mathrm{E}[\mathbf{1}_{B_1}(X_1)\mathbf{1}_{B_2}(X_2)] = \mathrm{E}[\mathrm{E}[\mathbf{1}_{B_1}(X_1)\mathbf{1}_{B_2}(X_2) \mid X_1]] \\ &= \mathrm{E}[\mathbf{1}_{B_1}(X_1)\,\mathrm{E}[\mathbf{1}_{B_2}(X_2) \mid X_1]] \,. \end{aligned}$$

Dunque, posto $S_2 = \{0, 1, \ldots, n\}$, per il Teorema di Fubini si ottiene

$$\begin{aligned} P(X_1 \in B_1, X_2 \in B_2) &= \int_{B_1} \left(\sum_{x_2 \in S_2 \cap B_2} p_{X_2|X_1=x_1}(x_2) \right) f_{X_1}(x_1)\, dx_1 \\ &= \sum_{x_2 \in S_2 \cap B_2} \int_{B_1} p_{X_2|X_1=x_1}(x_2) f_{X_1}(x_1)\, dx_1 \,, \\ &= \sum_{x_2 \in S_2 \cap B_2} \binom{n}{x_2} \frac{1}{B(\alpha,\beta)} \int_{]0,1[\cap B_1} x_1^{x_2+\alpha-1}(1-x_1)^{n-x_2+\beta-1}\, dx_1 \,, \end{aligned}$$

da cui

$$P(X_2 \in B_2) = P(X_1 \in \mathbb{R}, X_2 \in B_2) = \sum_{x_2 \in S_2 \cap B_2} \binom{n}{x_2} \frac{B(x_2+\alpha, n-x_2+\beta)}{B(\alpha,\beta)} \,.$$

Dunque, si ha

$$P(X_1 \in B_1, X_2 = x_2) = \binom{n}{x_2} \frac{1}{B(\alpha,\beta)} \int_{]0,1[\cap B_1} x_1^{x_2+\alpha-1}(1-x_1)^{n-x_2+\beta-1}\, dx_1\, \mathbf{1}_{S_2}(x_2) \,,$$

mentre la f.p. della v.a. X_2 è data da

$$p_{X_2}(x_2) = \binom{n}{x_2} \frac{B(x_2+\alpha, n-x_2+\beta)}{B(\alpha,\beta)}\, \mathbf{1}_{S_2}(x_2) \,.$$

Infine, per $x_2 \in S_2$ risulta

$$P(X_1 \in B_1 \mid X_2 = x_2) = \frac{1}{B(x_2+\alpha, n-x_2+\beta)} \int_{]0,1[\cap B_1} x_1^{x_2+\alpha-1}(1-x_1)^{n-x_2+\beta-1}\, dx_1 \,,$$

ovvero la legge condizionata della v.a. X_1 all'evento $\{X_2 = x_2\}$ è la legge Beta ridotta con parametri pari a $(x_2+\alpha)$ e $(n-x_2+\beta-1)$ e quindi si ottiene la seguente d.p. condizionata

$$f_{X_1|X_2=x_2}(x_1) = \frac{1}{B(x_2+\alpha, n-x_2+\beta)}\, x_1^{x_2+\alpha-1}(1-x_1)^{n-x_2+\beta-1}\, \mathbf{1}_{]0,1[}(x_1) \,.$$

Questo è un esempio tipico nella Statistica Bayesiana. La legge relativa ad X_1 è detta *a priori*, mentre la legge condizionata di X_1 all'evento $\{X_2 = x_2\}$ è detta *a posteriori*. Si osservi che in questo esempio la legge a priori e quella a posteriori appartengono della stessa famiglia. In questo caso, si dice che la legge di X_1 e la legge condizionata di X_2 all'evento $\{X_1 = x_1\}$ sono coniugate. □

Pagina intenzionalmente vuota

Capitolo 6

Principali leggi

6.1. Legge Binomiale

Si consideri un esperimento aleatorio con esito dicotomico, ovvero, utilizzando la terminologia dei giochi d'azzardo, suscettibile di assumere due soli risultati ω_1 e ω_2 del tipo "successo" e "insuccesso" di un ipotetico giocatore. Se si ripete n volte l'esperimento aleatorio, il corrispondente spazio fondamentale prodotto è dato da $\Omega = \Omega_1 \times \cdots \times \Omega_n$, dove $\Omega_k = \{\omega_1, \omega_2\}$. Quindi, Ω è costituito da 2^n eventi elementari che coincidono con tutte le possibili sequenze di lunghezza n (distinte anche per l'ordine) di ω_1 e ω_2.

Si supponga inoltre che gli esperimenti aleatori siano condotti in modo indipendente e che l'assegnazione di probabilità su Ω_k sia data da $P_k(\{\omega_1\}) = p$ e $P_k(\{\omega_2\}) = 1 - p = q$, dove $p \in]0, 1[$. In altre parole, l'assegnazione di probabilità è la stessa per ogni esperimento aleatorio. Dunque, la misura di probabilità prodotto P si ottiene probabilizzando i 2^n eventi elementari $\{(\omega_{j_1}, \ldots, \omega_{j_n})\}$ di Ω come

$$P(\{(\omega_{j_1}, \ldots, \omega_{j_n})\}) = \prod_{k=1}^{n} P_k(\{\omega_{j_k}\}) \,,$$

dove $j_k = 1, 2$ e $k = 1, \ldots, n$.

Si consideri la v.a. $X : \Omega \to S$, che ad ogni realizzazione dell'esperimento combinato associa il corrispondente numero di "successi", ovvero il numero di risultati del tipo ω_1 contenuti in ogni $(\omega_{j_1}, \ldots, \omega_{j_n}) \in \Omega$. La v.a. X è evidentemente di tipo discreto e si ha $S = \{0, 1, \ldots, n\}$. Si noti inoltre che la probabilità di ciascun evento elementare di Ω in cui compaiono x "successi" e $(n - x)$ "insuccessi" è pari a $p^x q^{n-x}$, dove $x = 0, 1, \ldots, n$. Inoltre, vi sono $\binom{n}{x}$ eventi elementari di questo tipo, che sono distinti solo per l'ordine in cui i "successi" e gli "insuccessi" si alternano. Dunque, la f.p. della v.a. X è data da

$$p_X(x) = \binom{n}{x} p^x q^{n-x} \, \mathbf{1}_{\{0,1,\ldots,n\}}(x) \,.$$

La legge relativa alla v.a. X è detta *legge Binomiale* di parametri n e p e si denota usualmente con $\mathcal{B}(n, p)$. Inoltre, nel caso particolare in cui $n = 1$, la legge viene detta *legge di Bernoulli* di parametro p, anche se entrambe le leggi sono state introdotte da Jakob Bernoulli.

Si osservi che p_X è in effetti una f.p., in quanto dall'espressione del binomio di Newton si verifica immediatamente che

$$\sum_{x=0}^{n} p_X(x) = \sum_{x=0}^{n} \binom{n}{x} p^x q^{n-x} = (p + q)^n = 1 \,.$$

La v.a. X può essere interpretata anche come una somma di n v.a. indipendenti X_k, ognuna con legge di Bernoulli di parametro p, ovvero si ha $X = \sum_{k=1}^{n} X_k$. Questa affermazione sarà dimostrata con un metodo semplice ed elegante nell'Esempio 7.4.3. Dunque, dal momento che è immediato verificare che $\mathrm{E}[X_k] = p$ e $\mathrm{Var}[X_k] = pq$ e, tenendo presente le proprietà relative alla media e varianza di somme di v.a. indipendenti, si ottiene

$$\mathrm{E}[X] = \sum_{k=1}^{n} \mathrm{E}[X_k] = np$$

e

$$\mathrm{Var}[X] = \sum_{k=1}^{n} \mathrm{Var}[X_k] = npq\ .$$

Questi risultati saranno anche verificati con un differente metodo nell'Esempio 7.2.4.

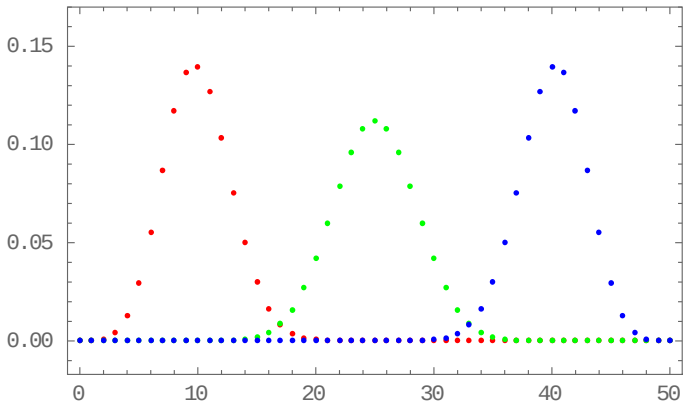


Figura 6.1.1. Funzione di probabilità per la legge Binomiale e $(n, p) = (50, \frac{1}{5}), (50, \frac{1}{2}), (50, \frac{4}{5})$ (rispettivamente in rosso, verde e blu).

6.2. Legge di Poisson

Si consideri una successione $(p_n)_{n\geq 1}$ di elementi di $]0, 1[$, in modo tale che la successione $(np_n)_{n\geq 1}$ converga ad una costante $\lambda \in]0, \infty[$ e si ponga inoltre $q_n = 1 - p_n$. Dunque, la successione $(p_n)_{n\geq 1}$ deve necessariamente convergere a 0. Per ogni n si consideri inoltre la successione $(g_n(x))_{x\geq 0}$, dove

$$g_n(x) = \begin{cases} \binom{n}{x} p_n^x q_n^{n-x} & x \in \{0, 1, \ldots, n\} \\ 0 & x \in \mathbb{N} \setminus \{0, 1, \ldots, n\} \end{cases}$$

In altre parole, $(g_n(x))_{x\geq 0}$ è il prolungamento su $\mathbb{N}$ della legge Binomiale $\mathcal{B}(n, p_n)$. Se $x \in \{0, 1, \ldots, n\}$ si ha

$$g_n(x) = \binom{n}{x} p_n^x q_n^{n-x} = \frac{1}{x!} (np_n)^x (1 - p_n)^{n-x} \prod_{k=0}^{x-1} \left(1 - \frac{k}{n}\right) .$$

Tenendo presente che se $(a_n)_{n\geq 1}$ è una successione di numeri reali tale che $\lim_n a_n = a$ si ha

$$\lim_n \left(1 + \frac{a_n}{n}\right)^n = e^a\ ,$$

allora risulta

$$\lim_n (1 - p_n)^n = \lim_n \left(1 - \frac{np_n}{n}\right)^n = e^{-\lambda}\ .$$

Inoltre, si ha

$$\lim_n (1 - p_n)^{-x} = 1$$

e

$$\lim_n \prod_{k=0}^{x-1} \left(1 - \frac{k}{n}\right) = 1 \,,$$

da cui si ottiene infine

$$\lim_n g_n(x) = e^{-\lambda} \frac{\lambda^x}{x!} \,.$$

Dunque, si può considerare la v.a. X con f.p. limite data da

$$p_X(x) = e^{-\lambda} \frac{\lambda^x}{x!} \mathbf{1}_{\mathbb{N}}(x) \,.$$

La v.a. X è discreta e prende valori su $S = \mathbb{N}$. La legge relativa alla v.a. X è detta *legge di Poisson* di parametro λ e si denota usualmente con $\mathcal{P}(\lambda)$. La legge prende nome dal matematico francese Siméon Denis Poisson (1781-1840), che ne ha considerato le prime applicazioni, anche se la legge era già nota a Abraham de Moivre. La legge di Poisson sarà ottenuta in modo più elegante nell'Esempio 8.1.7.

Si noti che p_X è in effetti una f.p., in quanto dall'espressione della serie esponenziale si ha

$$\sum_{x=0}^{\infty} p_X(x) = e^{-\lambda} \sum_{x=0}^{\infty} \frac{\lambda^x}{x!} = e^{-\lambda} e^{\lambda} = 1 \,.$$

La media della v.a. X è data da

$$\mathrm{E}[X] = e^{-\lambda} \sum_{x=0}^{\infty} x \frac{\lambda^x}{x!} = e^{-\lambda} \sum_{x=1}^{\infty} \frac{\lambda^x}{(x-1)!} = \lambda e^{-\lambda} \sum_{x=0}^{\infty} \frac{\lambda^x}{x!} = \lambda \,.$$

Inoltre, tenendo presente che $\mathrm{E}[X(X-1)] = \mathrm{E}[X^2] - \mathrm{E}[X]$ ed essendo

$$\mathrm{E}[X(X-1)] = e^{-\lambda} \sum_{x=0}^{\infty} x(x-1) \frac{\lambda^x}{x!} = e^{-\lambda} \sum_{x=2}^{\infty} \frac{\lambda^x}{(x-2)!} = \lambda^2 e^{-\lambda} \sum_{x=0}^{\infty} \frac{\lambda^x}{x!} = \lambda^2 \,,$$

si ha $\mathrm{E}[X^2] = \lambda^2 + \lambda$. Dunque, risulta $\mathrm{Var}[X] = \lambda$. Questi risultati saranno anche verificati con un differente metodo nell'Esempio 7.2.5.

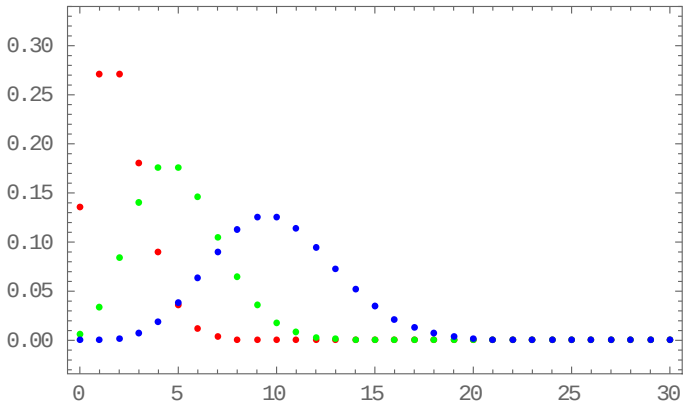


Figura 6.2.1. Funzione di probabilità per la legge di Poisson e $\lambda = 2, 5, 10$ (rispettivamente in rosso, verde e blu).

6.3. Legge Binomiale Negativa

Si consideri di nuovo un esperimento aleatorio con esito dicotomico, suscettibile di assumere due soli risultati ω_1 e ω_2, che al solito rappresentano "successo" e "insuccesso" di un ipotetico giocatore, e si ripeta l'esperimento aleatorio in modo indipendente sino a quando non si siano verificati k "successi". Anche se l'esperimento aleatorio combinato ha termine in pratica dopo che si sono ottenuti k "successi", da un punto di vista formale è conveniente considerare una infinità numerabile di esperimenti aleatori. In altre parole, si assume lo spazio prodotto $\Omega = \prod_{n=1}^{\infty} \Omega_n$, dove ogni singolo spazio fondamentale è dato da $\Omega_n = \{\omega_1, \omega_2\}$. Dunque, gli elementi dello spazio fondamentale prodotto Ω sono tutte le successioni del tipo $(\omega_{j_n})_{n \geq 1}$, dove $j_n = 1, 2$. Tenendo presente la discussione effettuata nella Sezione 2.6, se si effettua l'assegnazione di probabilità su Ω_n in modo tale che $P_n(\{\omega_1\}) = p$ e $P_n(\{\omega_2\}) = 1 - p = q$, dove $p \in]0, 1[$, la probabilità prodotto $P = \bigotimes_{n=1}^{\infty} P_n$ esiste ed è unica.

Si consideri la v.a. $X : \Omega \to S$, che ad ogni elemento di Ω associa il numero di esperimenti aleatori che si devono effettuare oltre il k-esimo esperimento per ottenere k "successi" (in effetti, affinchè si verifichino k "successi", si devono effettuare almeno k esperimenti). Si assuma inoltre che $k = 1, 2, \ldots$. La v.a. X è evidentemente di tipo discreto e risulta $S = \mathbb{N}$. La probabilità di ciascun evento elementare di Ω a cui corrispondono $(k-1)$ "successi" nei primi $(k + x - 1)$ esperimenti aleatori e per cui si ha di nuovo "successo" al $(k + x)$-esimo esperimento aleatorio risulta pari a $p^k q^x$. Inoltre, vi sono $\binom{k+x-1}{k-1}$ eventi elementari di questo tipo e che sono distinti solo per l'ordine in cui i "successi" e gli "insuccessi" si succedono. Dunque, la f.p. della v.a. X è data da

$$p_X(x) = \binom{k + x - 1}{k - 1} p^k q^x \, \mathbf{1}_{\mathbb{N}}(x) \, .$$

La legge relativa alla v.a. X è detta *legge Binomiale Negativa* di parametri k e p e si denota usualmente con $\mathcal{BN}(k, p)$. Inoltre, nel caso particolare in cui $k = 1$, la legge viene comunemente detta *legge Geometrica* di parametro p, anche se qualche autore definisce come legge Geometrica quella relativa alla v.a. $(X + 1)$.

Si osservi che p_X è effettivamente una f.p. Infatti, considerando la serie binomiale negativa, per $a \in]0, 1[$ e $b \in \mathbb{R}^+$ è noto che

$$\sum_{n=0}^{\infty} \binom{b + n - 1}{n} a^n = (1 - a)^{-b} \, ,$$

e quindi si ha

$$\sum_{x=0}^{\infty} p_X(x) = p^k \sum_{x=0}^{\infty} \binom{k + x - 1}{k - 1} q^x = p^k \sum_{x=0}^{\infty} \binom{k + x - 1}{x} q^x = p^k (1 - q)^{-k} = 1 \, .$$

La v.a. X può essere interpretata come una somma di k v.a. indipendenti X_j, ognuna con legge Geometrica di parametro p, ovvero risulta $X = \sum_{j=1}^{k} X_j$. Questa affermazione sarà dimostrata molto semplicemente nell'Esempio 7.4.4. Dunque, risulta (si tenga presente l'Esempio 4.1.1)

$$\mathrm{E}[X_j] = p \sum_{x=0}^{\infty} x q^x = \frac{q}{(1 - q)^2} = \frac{q}{p} \, .$$

Inoltre, per $a \in]0, 1[$ si ha

$$\sum_{n=0}^{\infty} n(n-1) a^n = \frac{2a^2}{(1 - a)^3}$$

da cui

$$\mathrm{E}[X_j(X_j-1)] = p\sum_{x=0}^{\infty} x(x-1)q^x = \frac{2pq^2}{(1-q)^3} = \frac{2q^2}{p^2}\ .$$

e quindi

$$\mathrm{Var}[X_j] = \frac{q}{p^2}\ .$$

Dunque, per le proprietà relative alla media e varianza di somme di v.a. indipendenti si ha

$$\mathrm{E}[X] = \sum_{j=1}^{k} \mathrm{E}[X_j] = \frac{kq}{p}$$

e

$$\mathrm{Var}[X] = \sum_{j=1}^{k} \mathrm{Var}[X_j] = \frac{kq}{p^2}\ .$$

Questi risultati saranno anche verificati con un differente metodo nell'Esempio 7.2.6.

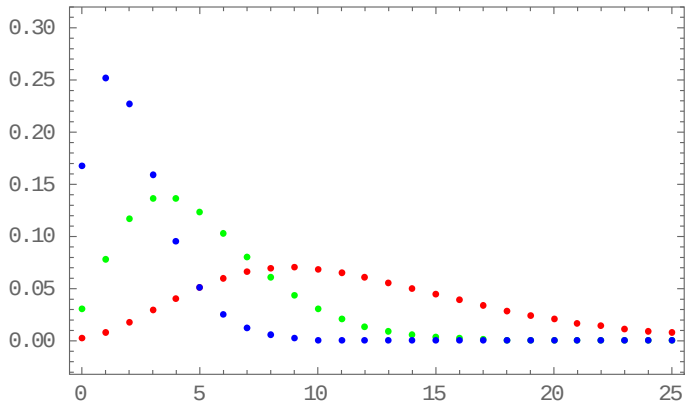


Figura 6.3.1. Funzione di probabilità per la legge Binomiale Negativa e $(k,p) = (5,\frac{3}{10}), (5,\frac{1}{2}), (5,\frac{7}{10})$ (rispettivamente in rosso, verde e blu).

Risulta inoltre interessante ottenere la f.p. limite per $k \to \infty$, qualora $(p_k)_{k\geq 1}$ sia una successione di elementi di $]0,1[$, tale che l'ulteriore successione $(kp_k^{-1}q_k)_{k\geq 1}$ converga ad una costante $\lambda \in]0,\infty[$ e dove $q_k = 1-p_k$. Si noti che $(p_k)_{k\geq 1}$ deve necessariamente convergere a 1. Per $x \in \mathbb{N}$ si ponga

$$g_k(x) = \binom{k+x-1}{k-1} p_k^k q_k^x = \frac{1}{x!}\left(\frac{kq_k}{p_k}\right)^x \left(1+\frac{q_k}{p_k}\right)^{-x-k} \prod_{j=0}^{x-1}\left(1+\frac{j}{k}\right)$$

e per ogni k si consideri la successione $(g_k(x))_{x\geq 0}$. Dal momento che

$$\lim_k \left(1+\frac{q_k}{p_k}\right)^{-k} = \lim_k \left(1+\frac{kp_k^{-1}q_k}{k}\right)^{-k} = e^{-\lambda}\ ,$$

mentre risulta

$$\lim_k \prod_{j=0}^{x-1}\left(1+\frac{j}{k}\right) = 1\ ,$$

si ha

$$\lim_k g_k(x) = e^{-\lambda}\,\frac{\lambda^x}{x!}\,,$$

ovvero la f.p. limite è quella di una v.a. con legge di Poisson $\mathcal{P}(\lambda)$. Questo risultato sarà ottenuto in modo più elegante nell'Esempio 8.1.8.

Si osservi infine che il parametro k può in generale assumere un qualsiasi valore positivo non necessariamente intero, anche se in questo caso la genesi della v.a. non si ottiene dallo schema degli esperimenti ripetuti. In effetti, la serie binomiale negativa può essere definita per ogni k positivo e le proprietà ottenute in precedenza possono essere estese anche a questo caso.

6.4. Legge Ipergeometrica

Si consideri un'urna composta da N palline di cui D sono rosse e $(N-D)$ sono nere. Si suppone ovviamente che D sia un intero positivo tale che $D \le N$. Si consideri inoltre l'esperimento aleatorio che consiste nell'estrarre in blocco n palline, dove n è un intero positivo tale che $n \le N$. Lo spazio fondamentale Ω relativo all'esperimento risulta costituito da $\binom{N}{n}$ risultati che corrispondono a tutte le possibili scelte di N palline a gruppi di n. Se si suppone che l'estrazione sia regolare, allora si può considerare un'assegnazione equiprobabile, ovvero la probabilità di ciascun evento elementare di Ω risulta pari a $\binom{N}{n}^{-1}$.

Si consideri la v.a. $X : \Omega \to S$, che ad ogni risultato associa il numero di numero di palline rosse estratte. La v.a. X è di tipo discreto e si ha inoltre $S = \{\max(0, n-N+D), \dots, \min(n, D)\}$. Gli estremi dell'insieme S derivano dal fatto che, se $n \ge N-D$ almeno $(n-N+D)$ palline estratte devono essere rosse, mentre se $n \ge D$ si possono estrarre al più D palline rosse. Inoltre il numero di eventi elementari per cui si hanno x palline rosse estratte risulta $\binom{D}{x}\binom{N-D}{n-x}$, che corrisponde al numero di scelte di D palline rosse a gruppi di x combinate con le scelte di $(N-D)$ palline nere a gruppi di $(n-x)$. Dunque, la f.p. della v.a. X è data da

$$p_X(x) = \frac{\binom{D}{x}\binom{N-D}{n-x}}{\binom{N}{n}}\,\mathbf{1}_S(x)\,.$$

La legge relativa alla v.a. X è detta *legge Ipergeometrica* di parametri n, D e N, e si denota usualmente con $\mathcal{I}(n, D, N)$.

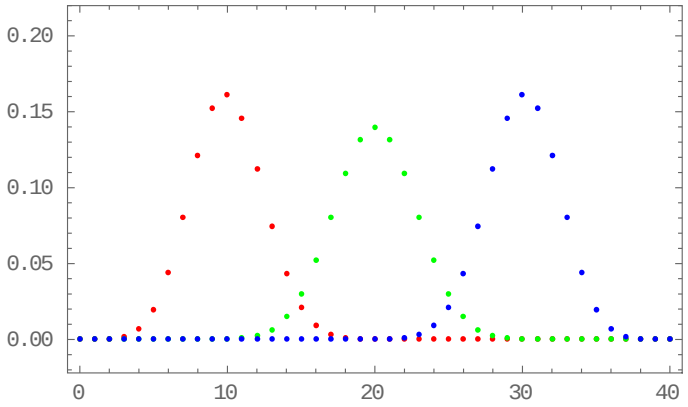


Figura 6.4.1. Funzione di probabilità per la legge Ipergeometrica e $(n, D, N) = (40, 50, 200), (40, 100, 200), (40, 150, 200)$ (rispettivamente in rosso, verde e blu).

Si osservi che p_X è effettivamente una f.p. Infatti, se $a, b, c \in \mathbb{N}$ allora l'identità di Vandermonde afferma che

$$\binom{a+b}{c} = \sum_{k=0}^{c} \binom{a}{k}\binom{b}{c-k},$$

da cui, si ha

$$\sum_{x\in S} p_X(x) = \binom{N}{n}^{-1} \sum_{x\in S} \binom{D}{x}\binom{N-D}{n-x} = \binom{N}{n}^{-1}\binom{N}{n} = 1\,.$$

L'identità di Vandermonde si ottiene per un caso particolare della funzione ipergeometrica, che conferisce dunque il nome alla legge.

Sulla base dell'espressione della f.p. relativa alla legge Ipergeometrica $\mathcal{I}(n-1, D-1, N-1)$, la media della v.a. X è data da

$$\mathrm{E}[X] = \sum_{x=\max(0,n-N+D)}^{\min(n,D)} x\,\frac{\binom{D}{x}\binom{N-D}{n-x}}{\binom{N}{n}} = \frac{nD}{N} \sum_{x=\max(0,n-N+D-1)}^{\min(n-1,D-1)} \frac{\binom{D-1}{x}\binom{N-D}{n-1-x}}{\binom{N-1}{n-1}} = \frac{nD}{N}$$

e, tenendo presente l'espressione della f.p. relativa alla legge Ipergeometrica $\mathcal{I}(n-2, D-2, N-2)$, si ha inoltre

$$\begin{aligned}\mathrm{E}[X(X-1)] &= \sum_{x=\max(0,n-N+D)}^{\min(n,D)} x(x-1)\,\frac{\binom{D}{x}\binom{N-D}{n-x}}{\binom{N}{n}} \\ &= \frac{nD(n-1)(D-1)}{N(N-1)} \sum_{x=\max(0,n-N+D-2)}^{\min(n-2,D-2)} \frac{\binom{D-2}{x}\binom{N-D}{n-2-x}}{\binom{N-2}{n-2}} = \frac{nD(n-1)(D-1)}{N(N-1)}\end{aligned}$$

da cui

$$\mathrm{Var}[X] = \frac{nD}{N}\,\frac{N-D}{N}\,\frac{N-n}{N-1}\,.$$

La legge Ipergeometrica può essere ottenuta anche come legge condizionata. Più esattamente, date N v.a. indipendenti X_k, ognuna con legge di Bernoulli di parametro p, si consideri le v.a. $X = \sum_{k=1}^{D} X_k$ e $Y = \sum_{k=1}^{N} X_k$. La f.p. condizionata della v.a. X all'evento $\{Y = n\}$ è data da (si veda l'Esercizio 6.4.1)

$$p_{X|Y=n}(x) = \frac{\binom{D}{x}p^x q^{D-x}\binom{N-D}{n-x}p^{n-x}q^{N-D-n+x}}{\binom{N}{n}p^n q^{N-n}}\,\mathbf{1}_S(x) = \frac{\binom{D}{x}\binom{N-D}{n-x}}{\binom{N}{n}}\,\mathbf{1}_S(x)\,.$$

Risulta interessante ottenere la f.p. limite per $N \to \infty$, qualora $(D_N)_{N\geq 1}$ sia una successione con $D_N \leq N$ per ogni N e tale che la successione $(N^{-1}D_N)_{N\geq 1}$ converga ad una costante $p \in]0,1[$. Per ogni N, si consideri l'insieme $(g_N(x))_{x=0}^{n}$, dove $g_N(x)$ è il prolungamento su $\{0,1,\ldots,n\}$ della legge Ipergeometrica $\mathcal{I}(n, D_N, N)$. Se $x \in S$, si ottiene

$$\begin{aligned}g_N(x) &= \frac{\binom{D_N}{x}\binom{N-D_N}{n-x}}{\binom{N}{n}} = \binom{n}{x}\,\frac{\prod_{k=0}^{x-1}(D_N-k)\prod_{k=0}^{n-x-1}(N-D_N-k)}{\prod_{k=0}^{n-1}(N-k)} \\ &= \binom{n}{x}\,\frac{\prod_{k=0}^{x-1}(\frac{D_N}{N}-\frac{k}{N})\prod_{k=0}^{n-x-1}(1-\frac{D_N}{N}-\frac{k}{N})}{\prod_{k=0}^{n-1}(1-\frac{k}{N})}\end{aligned}$$

Dal momento che si ha

$$\lim_N \prod_{k=0}^{x-1} \left(\frac{D_N}{N} - \frac{k}{N} \right) = p^x$$

e

$$\lim_N \prod_{k=0}^{n-x-1} \left(1 - \frac{D_N}{N} - \frac{k}{N} \right) = q^{n-x} ,$$

dove $q = 1 - p$, mentre

$$\lim_N \prod_{k=0}^{n-1} \left(1 - \frac{k}{N} \right) = 1,$$

si ottiene

$$\lim_N g_N(x) = \binom{n}{x} p^x q^{n-x} ,$$

ovvero la f.p. limite è quella di una v.a. con legge Binomiale $\mathcal{B}(n, p)$.

Si osservi infine che se le palline sono estratte dall'urna reimmettendo dopo ogni singola estrazione la pallina nell'urna, ovvero in caso di estrazioni indipendenti, allora la legge della v.a. X che rappresenta il numero di palline rosse estratte è Binomiale $\mathcal{B}(n, \frac{D}{N})$.

6.5. Legge Multinomiale

Nella presente Sezione viene introdotta la generalizzazione dell'esperimento aleatorio analizzato per la legge Binomiale. Più esattamente, si considera un esperimento aleatorio suscettibile di k possibili risultati e si ripete n volte l'esperimento aleatorio, in modo tale che lo spazio fondamentale prodotto associato è dato da $\Omega = \Omega_1 \times \cdots \times \Omega_n$ con $\Omega_l = \{\omega_1, \ldots, \omega_k\}$. Quindi, Ω è costituito da k^n eventi elementari che coincidono con tutte le possibili sequenze di lunghezza n (distinte anche per l'ordine) di $\omega_1, \ldots, \omega_k$.

Si supponga che gli esperimenti aleatori siano condotti indipendentemente e che l'assegnazione di probabilità su Ω_l sia la stessa per ogni esperimento aleatorio, ovvero che risulti $P_l(\{\omega_j\}) = p_j \in]0, 1[$ per ogni $l = 1, \ldots, n$ con $j = 1, \ldots, k$. Ovviamente, si deve avere $\sum_{j=1}^k p_j = 1$. In modo simile a quanto visto nella Sezione 6.1, la misura di probabilità prodotto P si ottiene probabilizzando i k^n eventi elementari $\{(\omega_{j_1}, \ldots, \omega_{j_n})\}$ di Ω come

$$P(\{(\omega_{j_1}, \ldots, \omega_{j_n})\}) = \prod_{l=1}^n P_l(\{\omega_{j_l}\}) ,$$

dove $j_l = 1, \ldots, k$ e $l = 1, \ldots, n$.

Si consideri il v.v.a. $X : \Omega \to S$, che ad ogni risultato associa il vettore $x = (x_1, \ldots, x_k)^{\mathrm{T}}$, dove x_j rappresenta il numero di elementi del tipo ω_j in ogni $(\omega_{j_1}, \ldots, \omega_{j_n}) \in \Omega$. Il v.v.a. $X = (X_1, \ldots, X_k)^{\mathrm{T}}$ è evidentemente di tipo discreto e $S \subset \mathbb{R}^k$ è tale che

$$S = \{(x_1, \ldots, x_k) : x_j \in \{0, 1, \ldots, n\}, \sum_{j=1}^k x_j = n\} .$$

Inoltre, la probabilità di ciascun evento elementare di Ω in cui si compare x_1 volte il risultato ω_1, …, x_k volte il risultato ω_k (dove $\sum_{j=1}^k x_j = n$), risulta pari a $\prod_{j=1}^k p_j^{x_j}$. Dal momento che vi sono

$$\binom{n}{x_1 \dots x_k} = \frac{n!}{\prod_{j=1}^{k} x_j!}$$

di tali eventi elementari e che sono distinti per l'ordine in cui gli eventi $\omega_1, \dots, \omega_k$ si succedono, la f.p.c. del v.v.a. X è data da

$$p_X(x) = p_X(x_1, \dots, x_k) = \binom{n}{x_1 \dots x_k} \prod_{j=1}^{k} p_j^{x_j} \, \mathbf{1}_S(x_1, \dots, x_k) \, .$$

Si noti che l'insieme S è contenuto su un iperpiano di $\mathbb{R}^k$ e che in effetti risulta $\{\sum_{j=1}^{k} X_j = n\}$ *q.c.*, ovvero il v.v.a. X è linearmente degenere. La legge relativa al v.v.a. X è detta *legge Multinomiale* di parametri n e $p = (p_1, \dots, p_k)^{\mathrm{T}}$, e si denota usualmente con $\mathcal{M}(n, p)$. Si osservi che p_X è in effetti una f.p.c., in quanto per il teorema multinomiale si ha

$$\sum_{(x_1,\dots,x_k)\in S} p_X(x_1, \dots, x_k) = \sum_{(x_1,\dots,x_k)\in S} \binom{n}{x_1 \dots x_k} \prod_{j=1}^{k} p_j^{x_j} = \left(\sum_{j=1}^{k} p_j \right)^n = 1 \, .$$

Si può dimostrare facilmente che la legge della componente marginale X_j è Binomiale $\mathcal{B}(n, p_j)$ (si veda l'Esempio 7.4.1). Inoltre, in modo analogo alla legge Binomiale, il v.v.a. X può essere espresso come somma di n v.v.a. indipendenti, ognuno dei quali con legge Multinomiale $\mathcal{M}(1, p)$ (si veda l'Esempio 7.4.5). Sulla base di questa osservazione, in modo simile a quanto fatto per la legge Binomiale, si può provare che il vettore delle medie del v.v.a. X è dato da $\mu = np$, mentre la matrice di varianza-covarianza è data da

$$\Sigma = n(\mathrm{diag}(p) - pp^{\mathrm{T}}) \, ,$$

con varianza generalizzata che deve necessariamente risultare $\det(\Sigma) = 0$. In modo alternativo, questi risultati possono essere ottenuti tenendo presente l'Esempio 7.4.2.

Si osservi infine che per $k = 2$ e $p = (p_1, p_2)^{\mathrm{T}}$ la legge $\mathcal{M}(n, p)$ è in effetti "equivalente" alla legge $\mathcal{B}(n, p_1)$, dal momento che si ha $\{X_2 = n - X_1\}$ *q.c.*, mentre risulta $\binom{n}{x_1\, x_2} = \binom{n}{x_1}$ e $p_2 = 1 - p_1$. Più esattamente, si dovrebbe affermare che, se la v.a. Y è distribuita con legge Binomiale $\mathcal{B}(n, p_1)$, allora risulta $(X_1, X_2) \stackrel{\mathcal{L}}{=} (Y, n - Y)$.

6.6. Leggi con parametri di posizione e di scala

Si osservi innanzitutto che due v.a. X e Z, con rispettive f.r. F_X e F_Z, sono dette dello *stesso tipo* se esistono due costanti $a, b \in \mathbb{R}$ per cui

$$X \stackrel{\mathcal{L}}{=} a + bZ \, ,$$

ovvero quando $F_Z(x) = F_X(a + bx)$ per ogni $x \in \mathbb{R}$. Si consideri dunque una v.a. X con f.r. data da

$$F_X(x) = G\Big(\frac{x - a}{b}\Big) \, ,$$

dove G è una f.r., mentre $a \in \mathbb{R}$ e $b \in \,]0, \infty[$. In pratica, per una data f.r. G, la precedente relazione descrive una famiglia di leggi al variare di a e b. Inoltre, a è detto *parametro di posizione* e b è detto *parametro di scala*. Se si consideri la trasformazione $Z = \frac{X-a}{b}$, risulta

$$F_Z(z) = P(Z \le z) = P(X \le a + bz) = F_X(a + bz) = G(z)$$

per ogni $z \in \mathbb{R}$. In altre parole, le due v.a. X e Z sono dello stesso tipo. Inoltre, è evidente che la precedente trasformazione consente di ottenere la v.a. Z con f.r. che non dipende dai parametri a e b. Per questo motivo si dice che la v.a. Z possiede la *legge ridotta* (o *legge standardizzata*) all'interno della famiglia di leggi. Analogamente, la v.a. Z viene comunemente detta *standardizzata*, anche se in modo leggermente improprio. In effetti, questa definizione dovrebbe essere solamente adottata per la trasformazione di standardizzazione, ovvero quando si ha $a = \mu_X$ e $b = \sigma_X$. In pratica, risulta quindi conveniente lavorare con la legge ridotta e successivamente considerare la famiglia di leggi che si ottiene al variare dei parametri di posizione e di scala (questo è il motivo per cui negli Esempi 4.1.3 e 4.1.4 sono state considerate le leggi ridotte). Infine, se la legge è caratterizzata da altri parametri oltre a quello di posizione e di scala, questi vengono detti *parametri di forma*.

Se X è una v.a. assolutamente continua, allora la d.p. di X è esprimibile attraverso la d.p. della v.a. standardizzata, ovvero

$$f_X(x) = \frac{1}{b}\, f_Z\Big(\frac{x-a}{b}\Big)\,,$$

mentre per quanto riguarda la f.r. risulta ovviamente

$$F_X(x) = F_Z\Big(\frac{x-a}{b}\Big)\,.$$

Si noti infine che il momento di ordine r della v.a. X è esprimibile attraverso i momenti della v.a. standardizzata Z, ovvero

$$\mu_{X,r} = \mathrm{E}[X^r] = \mathrm{E}[(a+bZ)^r] = \mathrm{E}[\sum_{k=1}^{r} \binom{r}{k} a^k (bZ)^{r-k}] = \sum_{k=1}^{r} \binom{r}{k} a^k b^{r-k} \mu_{Z,r-k}\,,$$

da cui $\mu_X = a + b\mu_Z$. Analogamente, il momento centrale di ordine r della v.a. X risulta

$$\mathrm{E}[(X-\mu_X)^r] = \mathrm{E}[(a+bZ-a-b\mu_Z)^r] = b^r \mathrm{E}[(Z-\mu_Z)^r]\,,$$

da cui $\sigma_X^2 = b^2\sigma_Z^2$. In particolare, se $a = \mu_X$ e $b = \sigma_X$, allora $\mu_Z = 0$ e $\sigma_Z = 1$.

• **Esempio 6.6.1.** (Legge di Gumbel) La legge introdotta dal matematico tedesco Emil Julius Gumbel (1891-1966) è associata alla f.r.

$$F_X(x) = e^{-e^{-\frac{x-a}{b}}}\,,$$

dove $a \in \mathbb{R}$ e $b \in \,]0,\infty[$. Evidentemente, una v.a. X che possiede la precedente f.r. è assolutamente continua e ammette d.p. data da

$$f_X(x) = \frac{1}{b}\, e^{-\frac{x-a}{b} - e^{-\frac{x-a}{b}}}\,.$$

Se si considera la trasformazione $Z = \frac{X-a}{b}$, si ha

$$F_Z(z) = e^{-e^{-z}}\,,$$

ovvero la f.r. della v.a. standardizzata Z non dipende dai parametri a e b. Inoltre, la v.a. Z ammette d.p. data da

$$f_Z(z) = e^{-z-e^{-z}}\,.$$

Si può verificare infine che $\mu_Z = \gamma$ e $\sigma_Z^2 = \frac{\pi^2}{6}$, dove $\gamma = -\int_0^\infty \log(x) e^{-x} dx \simeq 0.577$ rappresenta la costante di Eulero-Mascheroni. Si ha dunque $\mu_X = a + b\gamma$, mentre $\sigma_X^2 = \frac{b^2\pi^2}{6}$. □

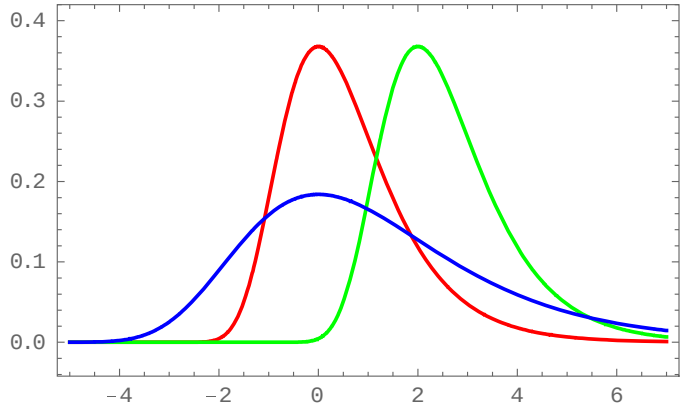


Figura 6.6.1. Densità di probabilità per la legge di Gumbel e $(a, b) = (0, 1), (2, 1), (0, 2)$ (rispettivamente in rosso, verde e blu).

• **Esempio 6.6.2.** (Legge di Weibull) L'ingegnere e statistico svedese Ernst Hjalmar Waloddi Weibull (1887-1979) ha introdotto la legge a cui è associata la f.r.

$$F_X(x) = \left(1 - e^{-\left(\frac{x-a}{b}\right)^k}\right) \mathbf{1}_{]a,\infty[}(x) \,,$$

dove $a \in \mathbb{R}$, $b \in]0, \infty[$ e $k \in]0, \infty[$. Evidentemente, una v.a. X che possiede la precedente f.r. è assolutamente continua e ammette d.p. data da

$$f_X(x) = \frac{k}{b}\left(\frac{x-a}{b}\right)^{k-1} e^{-\left(\frac{x-a}{b}\right)^k} \mathbf{1}_{]a,\infty[}(x) \,.$$

Se si considera la trasformazione $Z = \frac{X-a}{b}$, si ha

$$F_Z(z) = (1 - e^{-z^k})\, \mathbf{1}_{]0,\infty[}(z) \,,$$

ovvero la f.r. della v.a. Z non dipende dai parametri a e b. Tuttavia, la f.r. continua a dipendere dal parametro k, che quindi risulta essere un parametro di forma. Inoltre, la v.a. Z ammette d.p. data da

$$f_Z(z) = k z^{k-1}\, e^{-z^k}\, \mathbf{1}_{]0,\infty[}(z) \,.$$

Si può verificare che $\mu_Z = \Gamma(1 + k^{-1})$ e $\sigma_Z^2 = \Gamma(1 + 2k^{-1}) - \Gamma(1 + k^{-1})^2$, dove Γ rappresenta la funzione gamma di Eulero (si veda la Sezione 6.8 per una precisa definizione di questa funzione). Si ha $\mu_X = a + b\Gamma(1 + k^{-1})$, mentre $\sigma_X^2 = b^2(\Gamma(1 + 2k^{-1}) - \Gamma(1 + k^{-1})^2)$. □

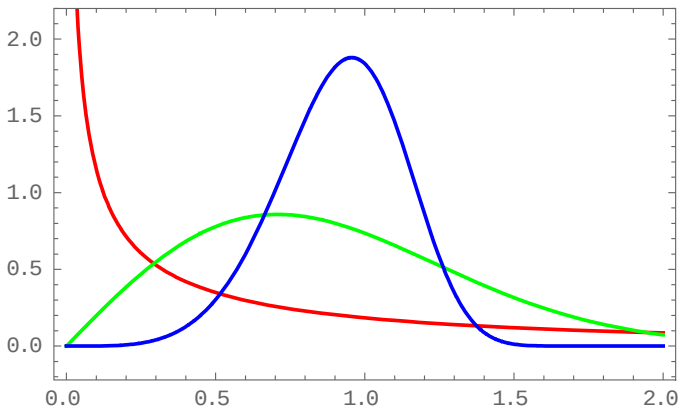


Figura 6.6.2. Densità di probabilità per la legge di Weibull e $(a, b, k) = (0, 1, \frac{1}{2}), (0, 1, 2), (0, 1, 5)$ (rispettivamente in rosso, verde e blu).

6.7. Legge Normale

Si consideri la v.a. assolutamente continua X che ammette d.p. data da

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

dove $\mu \in \mathbb{R}$ e $\sigma \in]0, \infty[$. La legge associata alla v.a. X è detta *legge Normale* di parametri μ e σ, e viene indicata comunemente con $\mathcal{N}(\mu, \sigma^2)$. La legge è anche detta Gaussiana in onore del matematico tedesco Johann Carl Friedrich Gauss (1777-1855) che ne studiò le applicazioni, anche se in effetti la legge fu introdotta da Abraham de Moivre.

Figura 6.7.1. Johann Carl Friedrich Gauss (1777-1855).

Se si considera la trasformazione di standardizzazione $Z = \frac{X-\mu}{\sigma}$, si ha

$$\phi(z) = f_Z(z) = \frac{1}{\sqrt{2\pi}}\, e^{-\frac{1}{2}z^2} ,$$

ovvero la d.p. della v.a. Z non dipende dai parametri μ e σ, che in effetti sono rispettivamente i parametri di posizione e di scala. Si noti che ϕ (e di conseguenza f_X) è in effetti una d.p. essendo

$$\left(\int_{-\infty}^{\infty} e^{-\frac{1}{2}z^2}\, dz\right)^2 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-\frac{1}{2}(z^2+y^2)}\, dzdy = \int_0^{\infty}\int_0^{2\pi} r\, e^{-\frac{1}{2}r^2}\, drd\theta = 2\pi ,$$

sulla base della trasformazione in coordinate polari. La legge associata alla v.a. Z è detta *legge Normale ridotta*. Si noti che la notazione ϕ per la d.p. f_Z è universalmente adottata, così come la notazione

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

per la f.r. F_Z. Di conseguenza, la f.r. di X è data da

$$F_X(x) = \Phi\Big(\frac{x-\mu}{\sigma}\Big) .$$

Dal momento che la v.a. Z è simmetrica rispetto all'origine si ha $\mathrm{E}[Z] = 0$, mentre

$$\mathrm{Var}[Z] = \mathrm{E}[Z^2] = \int_{-\infty}^{\infty} z^2\, e^{-\frac{1}{2}z^2}\, dz = 1 .$$

Si ha dunque $\mathrm{E}[X] = \mu$ e $\mathrm{Var}[X] = \sigma^2$, che rende più chiaro il motivo per cui si adotta la notazione $\mathcal{N}(\mu, \sigma^2)$ per indicare la legge Normale.

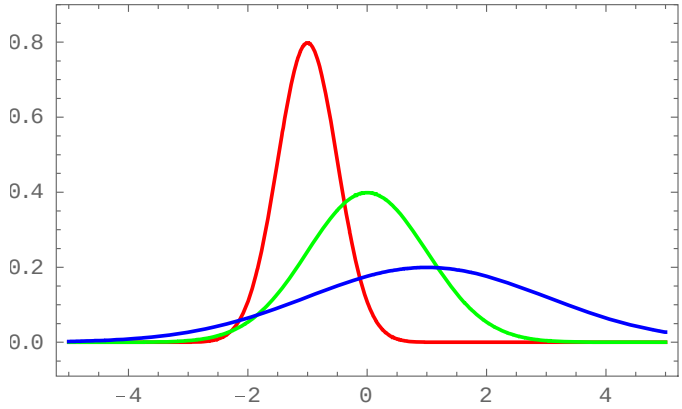


Figura 6.7.2. Densità di probabilità per la legge Normale e $(\mu, \sigma) = (-1, \frac{1}{2}), (0, 1), (1, 2)$ (rispettivamente in rosso, verde e blu).

6.8. Legge Gamma

Si consideri la v.a. assolutamente continua X che ammette d.p. data da

$$f_X(x) = \frac{1}{b\Gamma(k)} \left(\frac{x-a}{b}\right)^{k-1} e^{-\frac{x-a}{b}} \mathbf{1}_{]a,\infty[}(x) \,,$$

dove $a \in \mathbb{R}$ e $b \in]0, \infty[$, mentre $k \in]0, \infty[$. La legge associata alla v.a. X è detta *legge Gamma* di parametri a, b e k, e viene indicata comunemente con $\mathcal{G}(a, b, k)$. Se si considera la trasformazione di standardizzazione $Z = \frac{X-a}{b}$, si ha

$$f_Z(z) = \frac{1}{\Gamma(k)} z^{k-1} e^{-z} \mathbf{1}_{]0,\infty[}(z) \,,$$

ovvero la d.p. della v.a. Z non dipende dai parametri a e b, che rappresentano rispettivamente i parametri di posizione e di scala, mentre k è dunque un parametro di forma. La legge associata alla v.a. Z è detta *legge Gamma ridotta*. Si osservi che f_Z (e di conseguenza f_X) è in effetti una d.p., essendo

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

la nota funzione Gamma di Eulero (da cui ovviamente prende il nome anche la legge). La funzione Gamma è stata introdotta dal grande matematico svizzero Leonhard Euler (1707-1783), che viene comunemente italianizzato in Eulero. La f.r. relativa alla v.a. Z è data da

$$F_Z(z) = \frac{1}{\Gamma(k)} \gamma(k, z) \, \mathbf{1}_{]0,\infty[}(z) \,,$$

dove

$$\gamma(k, z) = \int_0^z u^{k-1} e^{-u} \, du$$

rappresenta la cosiddetta funzione Gamma incompleta. Di conseguenza, la f.r. di X è data da

$$F_X(x) = \frac{1}{\Gamma(k)} \gamma\Big(k, \frac{x-a}{b}\Big) \, \mathbf{1}_{]a,\infty[}(x) \,.$$

Dal momento che risulta $\Gamma(k+1) = k\Gamma(k)$, si ha

$$\mathrm{E}[Z] = \frac{1}{\Gamma(k)} \int_0^\infty z^k \, e^{-z} \, dz = \frac{\Gamma(k+1)}{\Gamma(k)} = k \, ,$$

e

$$\mathrm{E}[Z^2] = \frac{1}{\Gamma(k)} \int_0^\infty z^{k+1} \, e^{-z} \, dz = \frac{\Gamma(k+2)}{\Gamma(k)} = k(k+1) \, ,$$

da cui si ha anche $\mathrm{Var}[X] = k$. Di conseguenza, si ha anche $\mathrm{E}[X] = a + bk$ e $\mathrm{Var}[X] = b^2 k$.

Si noti che la legge Gamma contiene numerosi casi particolari di rilevante importanza sia pratica che teorica. Ad esempio, per $k = 1$ si ottiene la cosiddetta *legge Esponenziale*, e in questo caso la v.a. X ammette d.p. data da

$$f_X(x) = \frac{1}{b} \, e^{-\frac{x-a}{b}} \, \mathbf{1}_{]a,\infty[}(x) \, ,$$

con relativa f.r. data da

$$F_X(x) = \left(1 - e^{-\frac{x-a}{b}}\right) \mathbf{1}_{]a,\infty[}(x) \, .$$

Inoltre, per $a = 0$ e $b = 2$ e se $k = \frac{n}{2}$ dove $n = 1, 2, \ldots$, si ottiene la *legge Chi-quadrato* con n gradi di libertà, che viene indicata di solito con χ^2_n. In questo caso la v.a. X ammette d.p. data da

$$f_X(x) = \frac{1}{2\Gamma\left(\frac{n}{2}\right)} \left(\frac{x}{2}\right)^{\frac{1}{2}n-1} e^{-\frac{1}{2}x} \, \mathbf{1}_{]0,\infty[}(x) \, .$$

Si ha inoltre $\mathrm{E}[X] = n$ e $\mathrm{Var}[X] = 2n$. Tenendo presente l'Esempio 3.7.3 e dal momento che $\Gamma(\frac{1}{2}) = \sqrt{\pi}$, è immediato dimostrare che la legge del quadrato di una v.a. X con legge Normale ridotta $\mathcal{N}(0,1)$ è Chi-quadrato χ^2_1. Inoltre, la v.a. X con legge χ^2_n si può esprimere come una somma dei quadrati di n v.a. indipendenti con legge Normale ridotta $\mathcal{N}(0,1)$ (si veda l'Esempio 7.3.4).

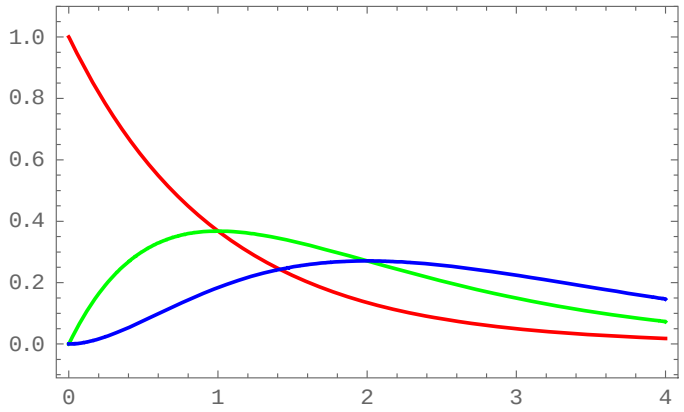


Figura 6.8.1. Densità di probabilità per la legge Gamma e $(a, b, k) = (0, 1, 1), (0, 1, 2), (0, 1, 3)$ (rispettivamente in rosso, verde e blu).

6.9. Legge Beta

Si consideri la v.a. assolutamente continua X che ammette d.p. data da

$$f_X(x) = \frac{\Gamma(\alpha+\beta)}{b\Gamma(\alpha)\Gamma(\beta)} \left(\frac{x-a}{b}\right)^{\alpha-1} \left(1 - \frac{x-a}{b}\right)^{\beta-1} \mathbf{1}_{]a,a+b[}(x) \, ,$$

dove $a \in \mathbb{R}$ e $b \in]0, \infty[$, mentre $\alpha, \beta \in]0, \infty[$. La legge associata alla v.a. X è detta *legge Beta* di parametri a, b, α e β, e viene indicata con $\mathcal{BE}(a, b, \alpha, \beta)$. Se si considera la trasformazione di standardizzazione $Z = \frac{X-a}{b}$, si ha

$$f_Z(z) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\, z^{\alpha-1}(1-z)^{\beta-1}\, \mathbf{1}_{]0,1[}(z)\,,$$

ovvero la d.p. della v.a. Z non dipende dai parametri a e b, che rappresentano rispettivamente i parametri di posizione e di scala, mentre α e β sono dunque parametri di forma. La legge associata alla v.a. Z è detta *legge Beta ridotta*. Si osservi che f_Z (e di conseguenza f_X) è in effetti una d.p., essendo

$$B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)} = \int_0^1 u^{\alpha-1}(1-u)^{\beta-1}\, du$$

la funzione Beta di Eulero (da cui ovviamente prende il nome anche la legge).

La d.p. f_Z (e di conseguenza la d.p. f_X) può assumere svariate morfologie al variare di α e β. In particolare, per $\alpha = \beta$, f_Z è simmetrica, per $\alpha, \beta \in]1, \infty[$, f_Z è campanulare, per $\alpha, \beta \in]0, 1[$, f_Z è a forma di U, mentre f_Z è crescente o decrescente se $\alpha \in]0, 1[$ o $\beta \in]0, 1[$.

La f.r. relativa alla v.a. Z è data da

$$F_Z(z) = I_z(\alpha, \beta)\, \mathbf{1}_{]0,1[}(z) + \mathbf{1}_{[1,\infty[}(z)\,,$$

dove

$$I_z(\alpha, \beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \int_0^z u^{\alpha-1}(1-u)^{\beta-1}\, du$$

rappresenta la cosiddetta funzione Beta incompleta regolarizzata. Dunque, la f.r. di X è data da

$$F_X(x) = I_{\frac{x-a}{b}}(\alpha, \beta)\, \mathbf{1}_{]a,a+b[}(x) + \mathbf{1}_{[a+b,\infty[}(x)\,.$$

Si noti che nel caso particolare in cui $\alpha = \beta = 1$, la legge associata alla v.a. X è detta *legge Uniforme* su $]a, a+b[$, e quindi la v.a. X ammette d.p. data da

$$f_X(x) = \frac{1}{b}\, \mathbf{1}_{]a,a+b[}(x)\,,$$

con corrispondente f.r. data da

$$F_X(x) = \frac{x}{b}\, \mathbf{1}_{]a,a+b[}(x) + \mathbf{1}_{[a+b,\infty[}(x)\,.$$

Dal momento che $\Gamma(k+1) = k\Gamma(k)$, si ha

$$\mathrm{E}[Z] = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \int_0^1 z^{\alpha}(1-z)^{\beta-1}\, dz = \frac{\alpha}{\alpha+\beta}\,,$$

e

$$\mathrm{E}[Z^2] = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \int_0^1 z^{\alpha+1}(1-z)^{\beta-1}\, dz = \frac{\alpha(\alpha+1)}{(\alpha+\beta)(\alpha+\beta+1)}\,,$$

da cui si ha anche

$$\mathrm{Var}[Z] = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}\,.$$

La media e la varianza della v.a. X sono quindi date da

$$\mathrm{E}[X] = a + b\,\frac{\alpha}{\alpha+\beta}\,,$$

mentre

$$\mathrm{Var}[X] = b^2\,\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta-1)}\,.$$

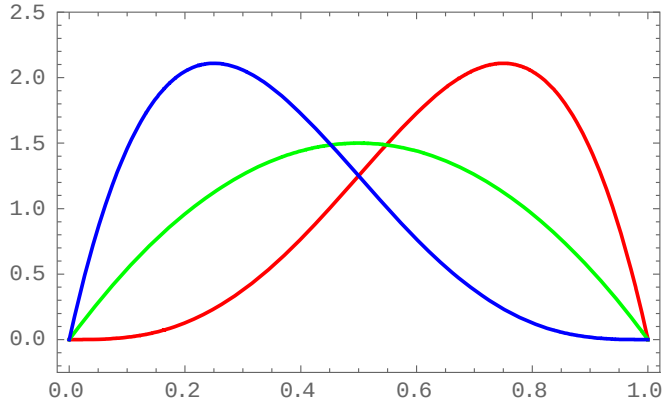


Figura 6.9.1. Densità di probabilità per la legge Beta e $(a,b,\alpha,\beta) = (0,1,4,2), (0,1,2,2), (0,1,2,4)$ (rispettivamente in rosso, verde e blu).

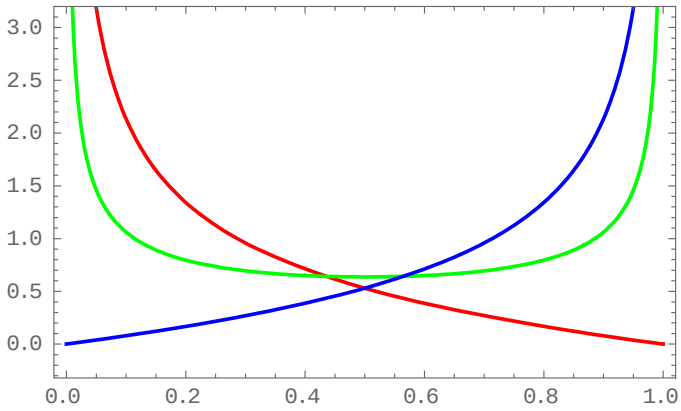


Figura 6.9.2. Densità di probabilità per la legge Beta e $(a,b,\alpha,\beta) = (0,1,\frac{1}{2},2), (0,1,\frac{1}{2},\frac{1}{2}), (0,1,2,\frac{1}{2})$ (rispettivamente in rosso, verde e blu).

La legge Beta nella forma ridotta può essere ottenuta considerando la v.a. data dal rapporto $\frac{Z_1}{Z_1+Z_2}$, dove Z_1 e Z_2 sono due v.a. indipendenti con legge Gamma ridotta, rispettivamente di parametri α e β. Si consideri dunque la trasformata $(Y_1,Y_2)^{\mathrm{T}} = g(Z_1,Z_2)$, dove $g(z_1,z_2) = (\frac{z_1}{z_1+z_2}, z_2)^{\mathrm{T}}$. Dal momento che risulta $g^{-1}(y_1,y_2) = (\frac{y_1y_2}{1-y_1}, y_2)^{\mathrm{T}}$, allora $|J(g^{-1}(y_1,y_2))| = \frac{|y_2|}{(1-y_1)^2}$, e quindi

$$f_{(Y_1,Y_2)}(y) = f_{(Z_1,Z_2)}\left(\frac{y_1y_2}{1-y_1}, y_2\right)\frac{|y_2|}{(1-y_1)^2}\,.$$

La d.p.m. della componente $Y_1 = \frac{Z_1}{Z_1+Z_2}$ è dunque data da

$$f_{Y_1}(y_1) = \int_{-\infty}^{\infty} f_{(Z_1,Z_2)}\left(\frac{y_1 y_2}{1-y_1}, y_2\right) \frac{|y_2|}{(1-y_1)^2}\, dy_2 = \int_{-\infty}^{\infty} f_{Z_1}\left(\frac{y_1 y_2}{1-y_1}\right) f_{Z_2}(y_2)\, \frac{|y_2|}{(1-y_1)^2}\, dy_2 \,,$$

ovvero, sostituendo opportunamente, si ha

$$\begin{aligned} f_{Y_1}(y_1) &= \frac{1}{\Gamma(\alpha)\Gamma(\beta)}\, y_1^{\alpha-1}(1-y_1)^{-\alpha-1}\, \mathbf{1}_{]0,1[}(y_1) \int_0^{\infty} y_2^{\alpha+\beta-1}\, e^{-\frac{y_2}{1-y_1}}\, dy_2 \\ &= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\, y_1^{\alpha-1}\,(1-y_1)^{\beta-1}\, \mathbf{1}_{]0,1[}(y_1)\,, \end{aligned}$$

che risulta essere appunto la d.p. di una v.a. distribuita con legge Beta ridotta di parametri α e β.

6.10. Legge *t* di Student e *F* di Snedecor

Le cosiddette leggi t di Student e F di Snedecor sono frequentemente utilizzate nell'ambito della Statistica inferenziale classica. Si consideri le v.a. indipendenti X_1 e X_2 distribuite rispettivamente con leggi $\chi^2_{n_1}$ e $\chi^2_{n_2}$ e la v.a.

$$X = \frac{n_1^{-1} X_1}{n_2^{-1} X_2} = \frac{n_2}{n_1}\, \frac{X_1}{X_2}\,.$$

Risulta evidente che la v.a. X è assolutamente continua. Al fine di ottenere la legge della v.a. X, si consideri innanzitutto la v.a. $Y = \frac{X_1}{X_2}$. Tenendo presente l'Esempio 3.7.5, la v.a. Y ammette d.p. data da

$$\begin{aligned} f_Y(y) &= \int_{-\infty}^{\infty} f_{X_1}(x) f_{X_2}(xy)|x|\, dx \\ &= \frac{1}{2^{\frac{1}{2}(n_1+n_2)}\Gamma\left(\frac{n_1}{2}\right)\Gamma\left(\frac{n_2}{2}\right)}\, y^{\frac{1}{2}n_1-1}\, \mathbf{1}_{]0,\infty[}(y) \int_0^{\infty} x^{\frac{1}{2}(n_1+n_2)-1}\, e^{-\frac{1}{2}(1+y)x}\, dx \\ &= \frac{\Gamma\left(\frac{n_1+n_2}{2}\right)}{\Gamma\left(\frac{n_1}{2}\right)\Gamma\left(\frac{n_2}{2}\right)}\, y^{\frac{1}{2}n_1-1}\,(1+y)^{-\frac{1}{2}(n_1+n_2)}\, \mathbf{1}_{]0,\infty[}(y)\,. \end{aligned}$$

Dunque, la v.a. $X = \frac{n_2}{n_1} Y$ ammette d.p. data da

$$f_X(x) = \frac{\Gamma\left(\frac{n_1+n_2}{2}\right)}{\Gamma\left(\frac{n_1}{2}\right)\Gamma\left(\frac{n_2}{2}\right)} \left(\frac{n_1}{n_2}\right)^{\frac{1}{2}n_1} x^{\frac{1}{2}n_1-1} \left(1+\frac{n_1}{n_2}\, x\right)^{-\frac{1}{2}(n_1+n_2)} \mathbf{1}_{]0,\infty[}(x)\,.$$

Figura 6.9.1. Ronald Aylmer Fisher (1890-1962).

La legge associata alla v.a. X è detta *legge F di Snedecor* con n_1 e n_2 gradi di libertà, e viene indicata comunemente con F_{n_1,n_2}. La legge dovrebbe essere più correttamente denominata di Fisher-Snedecor, dal momento che fu inizialmente introdotta dallo statistico e genetista inglese Ronald Aylmer Fisher (1890-1962) e poi applicata dal matematico statunitense George Waddel Snedecor (1881-1974).

Si tenga presente che, se $n_2 = 3, 4, \ldots$, si ha

$$\mathrm{E}[X_2^{-1}] = \frac{1}{4\,\Gamma\left(\frac{n_2}{2}\right)} \int_0^\infty \left(\frac{x}{2}\right)^{\frac{1}{2}n_2-2} e^{-\frac{x}{2}}\,dx = \frac{\Gamma\left(\frac{n_2}{2}-1\right)}{2\,\Gamma\left(\frac{n_2}{2}\right)} = \frac{1}{n_2-2}\,,$$

mentre, se $n_2 = 4, 5, \ldots$, risulta

$$\mathrm{E}[X_2^{-2}] = \frac{1}{8\,\Gamma\left(\frac{n_2}{2}\right)} \int_0^\infty \left(\frac{x}{2}\right)^{\frac{1}{2}n_2-3} e^{-\frac{x}{2}}\,dx = \frac{\Gamma\left(\frac{n_2}{2}-2\right)}{4\,\Gamma\left(\frac{n_2}{2}\right)} = \frac{1}{(n_2-2)(n_2-4)}\,.$$

Dunque, sulla base della definizione della v.a. X e dal momento che X_1 e X_2 sono indipendenti, si ottiene

$$\mathrm{E}[X] = \frac{n_2}{n_1}\,\mathrm{E}[X_1]\mathrm{E}[X_2^{-1}] = \frac{n_2}{n_2-2}\,,$$

se $n_2 = 3, 4, \ldots$, e

$$\mathrm{Var}[X] = \frac{n_2^2}{n_1^2}\,\mathrm{E}[X_1^2]\mathrm{E}[X_2^{-2}] - \mathrm{E}[X]^2 = \frac{2n_2^2(n_1+n_2-2)}{n_1(n_2-2)^2(n_2-4)}\,,$$

se $n_2 = 4, 5, \ldots$. Dunque, se $n_2 = 1, 2$ la media non è finita, mentre se $n_2 = 1, 2, 3, 4$ la varianza non è finita.

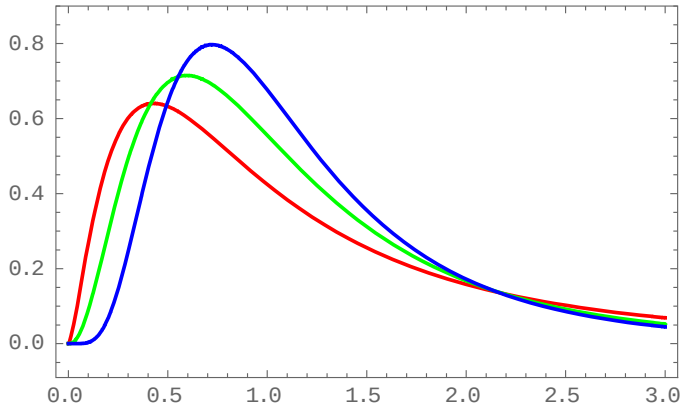


Figura 6.10.1. Densità di probabilità per la legge F di Snedecor e $(n_1, n_2) = (5, 5), (7, 10), (15, 10)$ (rispettivamente in rosso, verde e blu).

In secondo luogo, si consideri le v.a. indipendenti X_1 e X_2 rispettivamente con legge $\mathcal{N}(0, 1)$ e legge χ^2_n e la v.a.

$$X = \frac{X_1}{\sqrt{n^{-1}X_2}}\,.$$

È evidente che la v.a. X è assolutamente continua. Al fine di ottenere la legge della v.a. X, si osservi che la v.a. $Y = \frac{nX_1^2}{X_2}$ si distribuisce con legge di Snedecor $F_{1,n}$, dal momento che la v.a. X_1^2 si distribuisce con legge Chi-quadrato χ^2_1. Dunque, si ha

$$f_Y(y) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\,\Gamma\left(\frac{n}{2}\right)}\, y^{-\frac{1}{2}}\left(1+\frac{y}{n}\right)^{-\frac{1}{2}(n+1)} \mathbf{1}_{]0,\infty[}(y)\,,$$

poichè $\Gamma(\frac{1}{2}) = \sqrt{\pi}$. Inoltre, tenendo presente che $X_1 \stackrel{\mathcal{L}}{=} -X_1$ e che $X = \mathrm{sgn}(X_1)\sqrt{Y}$, sulla base della discussione successiva alla Proposizione 3.7.1 si ha

$$f_X(x) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\,\Gamma\left(\frac{n}{2}\right)}\left(1+\frac{x^2}{n}\right)^{-\frac{1}{2}(n+1)}\,.$$

La legge associata alla v.a. X è detta *legge t di Student* con n gradi di libertà, e viene indicata comunemente con t_n. La distribuzione venne introdotta dallo statistico inglese William Sealy Gosset (1876-1937), che pubblicò il risultato sotto lo pseudonimo di "Student" dal momento che la birreria presso la quale era impiegato vietava ai propri dipendenti di pubblicare articoli, affinché questi non divulgassero segreti di produzione. Si noti inoltre che per $n = 1$ si ottiene la legge di Cauchy (si veda Esempio 4.1.4).

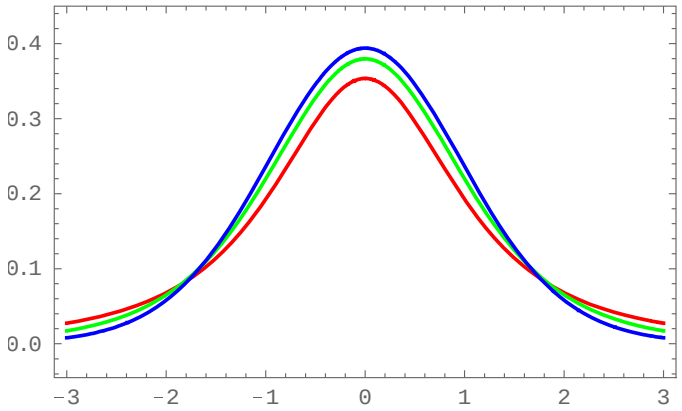


Figura 6.10.2. Densità di probabilità per la legge t di Student e $n = 2, 5, 20$ (rispettivamente in rosso, verde e blu).

Tenendo presente la definizione della v.a. X, dal momento che X_1 e X_2 sono indipendenti e che $\mathrm{E}[X_1] = 0$ e $\mathrm{E}[X_2^{-\frac{1}{2}}]$ è finito se $n = 2, 3, \ldots$, la media della v.a. X è data da

$$\mathrm{E}[X] = \sqrt{n}\,\mathrm{E}[X_1]\mathrm{E}[X_2^{-\frac{1}{2}}] = 0\,,$$

se $n = 2, 3, \ldots$. Inoltre, dal momento che si ha $\mathrm{E}[X_1^2] = 1$ e per quanto visto in precedenza in questa Sezione si ha $\mathrm{E}[X_2^{-1}] = \frac{1}{n-2}$ se $n = 3, 4, \ldots$, allora

$$\mathrm{Var}[X] = \mathrm{E}[X^2] = n\mathrm{E}[X_1^2]\mathrm{E}[X_2^{-1}] = \frac{n}{n-2}\,,$$

se $n = 3, 4, \ldots$. Dunque, se $n = 1$ la media non è definita, mentre se $n = 1, 2$ la varianza non è definita.

6.11. Legge Normale Multivariata

Si consideri il v.v.a. assolutamente continuo $X = (X_1, \ldots, X_n)^{\mathrm{T}}$ che ammette d.p.c. data da

$$f_X(x) = \det(2\pi\Sigma)^{-\frac{1}{2}}\, e^{-\frac{1}{2}(x-\mu)^{\mathrm{T}}\Sigma^{-1}(x-\mu)}\,,$$

dove $x = (x_1, \dots, x_n)^{\mathrm{T}}$, mentre $\mu \in \mathbb{R}^n$ rappresenta il vettore delle medie e Σ è la matrice di varianza-covarianza, che si assume definita positiva. La legge associata al v.v.a. X è detta *legge Normale Multivariata* di parametri μ e Σ, e viene indicata comunemente con $\mathcal{N}_n(\mu, \Sigma)$.

Si osservi che f_X è in effetti una d.p.c. Infatti, considerando il cambio di variabile $y = \Sigma^{-\frac{1}{2}}(x - \mu)$ dove $y = (y_1, \dots, y_n)^{\mathrm{T}}$ e con Jacobiano dato da $\det(\Sigma)^{\frac{1}{2}}$, si ha

$$\begin{aligned}\int_{\mathbb{R}^n} e^{-\frac{1}{2}(x-\mu)^{\mathrm{T}}\Sigma^{-1}(x-\mu)}\,dx &= \det(\Sigma)^{\frac{1}{2}} \int_{\mathbb{R}^n} e^{-\frac{1}{2}y^{\mathrm{T}}y}\,dy = \det(\Sigma)^{\frac{1}{2}} \int_{\mathbb{R}} \dots \int_{\mathbb{R}} \left(\prod_{k=1}^{n} e^{-\frac{1}{2}y_k^2}\right) dy_1 \dots dy_n \\ &= \det(\Sigma)^{\frac{1}{2}} \prod_{k=1}^{n} \int_{\mathbb{R}} e^{-\frac{1}{2}y_k^2}\,dy_k = (2\pi)^{\frac{n}{2}} \det(\Sigma)^{\frac{1}{2}} = \det(2\pi\Sigma)^{\frac{1}{2}} .\end{aligned}$$

Inoltre, la d.p.c. risulta costante per tutti i vettori x che appartengono al luogo geometrico $(x - \mu)^{\mathrm{T}}\Sigma^{-1}(x - \mu) = c$, con $c \in]0, \infty[$, ovvero è costante su un iper-ellissoide. Questo iper-ellissoide è centrato in μ, con assi le cui direzioni sono determinate dalle direzioni degli autovettori di Σ e le cui lunghezze sono pari a $2\sqrt{c\lambda_k}$, dove λ_k rappresenta il k-esimo autovalore di Σ, con $k = 1, \dots, n$.

Se X è un v.v.a. con legge Normale Multivariata, la legge di ogni scelta $(j_1, \dots, j_k)$ di k v.a. $(X_{j_1}, \dots, X_{j_k})^{\mathrm{T}}$ con $k = 1, \dots, n-1$, è ancora Normale Multivariata con vettore delle medie e matrice di varianza-covarianza costituiti rispettivamente dalle medie e dalle varianze e covarianze delle v.a. marginali che lo compongono (si veda l'Esempio 7.3.1). Da questo segue ovviamente che ogni componente marginale X_k è distribuito con legge Normale $\mathcal{N}(\mu_{X_k}, \sigma^2_{X_k})$. Risulta inoltre valido per una legge Normale Multivariata la seguente Proposizione.

Proposizione 6.11.1. *Il v.v.a. $X = (X_1, \dots, X_n)^{\mathrm{T}}$ con legge $\mathcal{N}_n(\mu, \Sigma)$ possiede componenti reciprocamente indipendenti se e solo se Σ è una matrice diagonale.*

Dimostrazione. Se X possiede componenti marginali indipendenti la dimostrazione è immediata in quanto l'indipendenza implica assenza di correlazione e quindi $\sigma_{X_k X_j} = 0$ per ogni $k \neq j = 1, \dots, n$. Inversamente, se $\sigma_{X_k X_j} = 0$ per ogni $k \neq j = 1, \dots, n$, allora risulta $\det(2\pi\Sigma) = \prod_{k=1}^{n} 2\pi\sigma^2_{X_k}$ e $\Sigma^{-1} = \mathrm{diag}(\sigma^{-2}_{X_1}, \dots, \sigma^{-2}_{X_n})$, da cui

$$(x - \mu)^{\mathrm{T}}\Sigma^{-1}(x - \mu) = \sum_{k=1}^{n} \frac{(x_k - \mu_{X_k})^2}{\sigma^2_{X_k}} .$$

Sostituendo queste espressioni nella d.p.c. del v.v.a. X si ottiene infine

$$\begin{aligned} f_X(x) &= \frac{1}{\prod_{k=1}^{n} \sqrt{2\pi}\sigma_{X_k}} \exp\left(- \sum_{k=1}^{n} \frac{(x_k - \mu_{X_k})^2}{2\sigma^2_{X_k}} \right) \\ &= \prod_{k=1}^{n} \frac{1}{\sqrt{2\pi}\sigma_{X_k}} \exp\left(- \frac{(x_k - \mu_{X_k})^2}{2\sigma^2_{X_k}} \right) = \prod_{k=1}^{n} f_{X_k}(x_k) , \end{aligned}$$

che è quanto si voleva dimostrare. □

In pratica la Proposizione 6.11.1 afferma che nel caso di v.v.a. con legge Normale Multivariata, l'assenza di correlazione fra le v.a. marginali è condizione sufficiente e necessaria per l'indipendenza delle stesse.

6.12. Legge di Dirichlet

In questa Sezione viene introdotta la generalizzazione multivariata della legge Beta ridotta. Si consideri il v.v.a. assolutamente continuo $Z = (Z_1, \ldots, Z_n)^{\mathrm{T}}$ che ammette d.p.c. data da

$$f_Z(z) = f_Z(z_1, \ldots, z_n) = \frac{\Gamma(\alpha_0)}{\prod_{k=1}^n \Gamma(\alpha_k)} \prod_{k=1}^n z_k^{\alpha_k - 1} \, \mathbf{1}_S(z_1, \ldots, z_n) \, ,$$

dove $z = (z_1, \ldots, z_n)^{\mathrm{T}}$, mentre $\alpha = (\alpha_1, \ldots, \alpha_n)^{\mathrm{T}}$ è un vettore di parametri tale che $\alpha_k \in]0, \infty[$ per $k = 1, \ldots, n$ e $\alpha_0 = \sum_{k=1}^n \alpha_k$, e

$$S = \{(z_1, \ldots, z_n) : z_k \in]0, 1[, \sum_{k=1}^n z_k = 1\} \, .$$

Si osservi che l'insieme S è contenuto su un iperpiano di $\mathbb{R}^n$ e che in effetti risulta $\{\sum_{k=1}^n Z_k = 1\}$ $q.c.$, ovvero il v.v.a. Z è linearmente degenere. La legge associata al v.v.a. Z è detta *legge di Dirichlet* di parametro α, e viene indicata comunemente con $\mathcal{D}(\alpha)$. La legge prende nome dal matematico tedesco Johann Peter Dirichlet (1805-1859). Inoltre, f_Z è in effetti una d.p.c., essendo

$$\frac{\prod_{k=1}^n \Gamma(\alpha_k)}{\Gamma(\alpha_0)} = \int_S \prod_{k=1}^n z_k^{\alpha_k - 1} \, dz_1 \ldots dz_n$$

la funzione Beta multivariata.

Al fine di determinare le proprietà del v.v.a. Z è conveniente estendere un risultato visto in precedenza per la legge Beta ridotta. Sia $X = (X_1, \ldots, X_n)^{\mathrm{T}}$ un v.v.a. con componenti marginali indipendenti, in modo tale che la v.a. X_k è distribuita con legge Gamma $\mathcal{G}(0, 1, \alpha_k)$. Inoltre, si consideri la trasformata $Y = (Y_1, \ldots, Y_n)^{\mathrm{T}} = g(X)$, dove

$$g(x) = \left(\frac{x_1}{\sum_{k=1}^n x_k}, \ldots, \frac{x_{n-1}}{\sum_{k=1}^n x_k}, \sum_{k=1}^n x_k \right)^{\mathrm{T}} \, ,$$

per cui risulta $g^{-1}(y) = (y_1 y_n, \ldots, y_{n-1} y_n, (1 - \sum_{k=1}^{n-1} y_k) y_n)^{\mathrm{T}}$ e dunque $|J(g^{-1}(y))| = y_n^{n-1}$. Dal momento che

$$f_X(x_1, \ldots, x_n) = \prod_{k=1}^n \frac{1}{\Gamma(\alpha_k)} \, x_k^{\alpha_k - 1} \, e^{-x_k} \, \mathbf{1}_{]0,\infty[}(x_k) \, ,$$

allora si ha

$$f_Y(y_1, \ldots, y_n) = \frac{1}{\prod_{k=1}^n \Gamma(\alpha_k)} \left(\prod_{k=1}^{n-1} y_k^{\alpha_k - 1} \right) \left(1 - \sum_{k=1}^{n-1} y_k \right)^{\alpha_n - 1} y_n^{\alpha_0 - 1} e^{-y_n} \, \mathbf{1}_T(y_1, \ldots, y_{n-1}) \mathbf{1}_{]0,\infty[}(y_n) \, ,$$

dove

$$T = \{(y_1, \ldots, y_{n-1}) : y_k \in]0, 1[, \sum_{k=1}^{n-1} y_k \leq 1\} \, .$$

Dunque, posto $V = (Y_1, \ldots, Y_{n-1})^{\mathrm{T}}$ si ha

$$f_V(y_1,\ldots,y_{n-1}) = \frac{1}{\prod_{k=1}^n \Gamma(\alpha_k)} \left(\prod_{k=1}^{n-1} y_k^{\alpha_k-1}\right) \left(1-\sum_{k=1}^{n-1} y_k\right)^{\alpha_n-1} \mathbf{1}_T(y_1,\ldots,y_{n-1}) \int_0^\infty y_n^{\alpha_0-1} e^{-y_n}\, dy_n$$
$$= \frac{\Gamma(\alpha_0)}{\prod_{k=1}^n \Gamma(\alpha_k)} \left(\prod_{k=1}^{n-1} y_k^{\alpha_k-1}\right) \left(1-\sum_{k=1}^{n-1} y_k\right)^{\alpha_n-1} \mathbf{1}_T(y_1,\ldots,y_{n-1})\,.$$

Il v.v.a. linearmente degenere $(Z_1,\ldots,Z_n)^{\mathrm{T}} = (Y_1,\ldots,Y_{n-1},1-\sum_{k=1}^{n-1}Y_k)^{\mathrm{T}}$ si distribuisce quindi con legge di Dirichlet $\mathcal{D}(\alpha)$.

Dal precedente risultato e tenendo presente quanto detto nella Sezione 6.9, la componente marginale Z_k del v.v.a. Z si distribuisce con legge $\mathcal{BE}(0,1,\alpha_k,\alpha_0-\alpha_k)$. Inoltre, ogni scelta $(j_1,\ldots,j_k)$ di k v.a. $(Z_{j_1},\ldots,Z_{j_k})^{\mathrm{T}}$ con $k=2,\ldots,n-1$, si distribuisce ancora con legge di Dirichlet. Infine, dal medesimo risultato si può anche dimostrare che il vettore delle medie del v.v.a. Z è dato da $\mu = \frac{1}{\alpha_0}\alpha$, mentre la matrice di varianza-covarianza risulta

$$\Sigma = \frac{1}{\alpha_0^2(\alpha_0+1)} \left(\alpha_0 \operatorname{diag}(\alpha) - \alpha\alpha^{\mathrm{T}}\right),$$

con varianza generalizzata che deve necessariamente risultare $\det(\Sigma) = 0$.

Si osservi infine che per $n=2$ e $\alpha = (\alpha_1,\alpha_2)^{\mathrm{T}}$ la legge $\mathcal{D}(\alpha)$ è in effetti "equivalente" alla legge $\mathcal{BE}(0,1,\alpha_1,\alpha_2)$, dal momento che si ha $\{Z_2 = 1-Z_1\}$ $q.c.$ Più correttamente, se la v.a. Y si distribuisce con legge Beta $\mathcal{BE}(0,1,\alpha_1,\alpha_2)$, allora si ha $(Z_1,Z_2) \overset{\mathcal{L}}{=} (Y,1-Y)$.

6.13. Riferimenti bibliografici

Una trattazione enciclopedica delle leggi di probabilità e delle loro applicazioni nelle varie discipline è contenuta nella serie di volumi di Johnson *et al.* (2005), Johnson *et al.* (1994, 1995, 1997) e Kotz *et al.* (2000). Inoltre, una esposizione più sintetica di questi argomenti è data in Forbes *et al.* (2011). Per quanto riguarda le leggi relative a variabili aleatorie discrete, si dovrebbero consultare anche i testi di Charalambides (2005) e Johnson e Kotz (1977).

6.14. Esercizi svolti

• **Nota.** Al fine di evitare notazioni ridondanti e ripetizioni, nei seguenti esercizi si assume l'esistenza di uno spazio probabilizzato $(\Omega,\mathcal{F},P)$, a meno che non venga specificato diversamente.

Sezione 6.1

• **Esercizio 6.1.1.** Due monete bilanciate vengono lanciate n volte in modo indipendente. Si determini la probabilità di ottenere lo stesso numero di volte la faccia contrassegnata dalla testa per entrambe le monete.

Soluzione. Si considerino le v.a. indipendenti X_1 e X_2 che descrivono il numero di volte che si ottiene la faccia contrassegnata dalla testa rispettivamente per la prima e la seconda moneta. Le v.a. X_1 e X_2 sono entrambe distribuite con legge Binomiale $\mathcal{B}(n,\frac{1}{2})$. Tenendo presente le proprietà dei coefficienti Binomiali, e in particolare la relazione che assicura che la f.p. di una v.a. con legge Ipergeometrica è in effetti tale, la probabilità richiesta è data da

$$P(X_1 = X_2) = \sum_{x=0}^{n} P(X_1 = x)P(X_2 = x) = 2^{-2n} \sum_{x=0}^{n} \binom{n}{x}^2$$
$$= 2^{-2n} \sum_{x=0}^{n} \binom{n}{x}\binom{n}{n-x} = \binom{2n}{n} 2^{-2n} .$$

Sulla base della formula di Stirling, si ha inoltre che

$$\binom{2n}{n} 2^{-2n} \sim \frac{1}{\sqrt{\pi n}} ,$$

ovvero la probabilità richiesta tende a zero per $n \to \infty$. □

• **Esercizio 6.1.2.** (Problema delle scatole di fiammiferi di Banach) Un matematico possiede una scatola con n fiammiferi in ognuna delle sue due tasche. Il matematico prende un fiammifero da una delle due scatole, scelta con la medesima probabilità, finchè non trova una scatola vuota. Si determini la media del numero di fiammiferi rimasti nell'altra scatola.

Soluzione. L'esperimento aleatorio è equivalente a quello del lancio di una moneta equilibrata fino a quando si ottiene $(n+1)$ volte la faccia contrassegnata dalla testa o dalla croce. In effetti, una scatola viene trovata vuota dopo che sono stati estratti n fiammiferi e viene effettuato un ulteriore tentativo. Se la v.a. X rappresenta il numero di lanci che vengono effettuati, allora X è discreta con legge essenziale definita su $S = \{n+1, \ldots, 2n+1\}$. Dunque, l'evento $\{X = x\}$ si verifica se si è ottenuto n volte la faccia contrassegnata dalla testa (o dalla croce) nei primi $(x-1)$ lanci e il medesimo risultato al lancio x. Tenendo presente la legge Binomiale e che la probabilità di ottenere $(n+1)$ volte la faccia contrassegnata dalla testa è uguale a quella relativa alla croce, la f.p. della v.a. X è data da

$$p_X(x) = 2\binom{x-1}{n}\left(\frac{1}{2}\right)^{x-1}\frac{1}{2}\,\mathbf{1}_{\{n+1,\ldots,2n+1\}}(x) = \binom{x-1}{n} 2^{-x+1}\,\mathbf{1}_{\{n+1,\ldots,2n+1\}}(x) .$$

Si osservi che p_X è in effetti una f.p., dal momento che sulla base dell'Esercizio 3.3.4 si ha

$$\sum_{x=n+1}^{2n+1} \binom{x-1}{n} 2^{-x+1} = \sum_{x=0}^{n} \binom{n+x}{n} 2^{-n-x} = 1 .$$

Inoltre, si ha

$$\mathrm{E}[X] = \sum_{x=n+1}^{2n+1} x \binom{x-1}{n} 2^{-x+1} = (n+1) \sum_{x=n+1}^{2n+1} \binom{x}{n+1} 2^{-x+1}$$
$$= 2(n+1) \sum_{x=n+2}^{2n+2} \binom{x-1}{n+1} 2^{-x+1} ,$$

ovvero, aggiungendo e sottraendo in modo opportuno e tenendo di nuovo presente l'Esercizio 3.3.4, si ottiene

$$\mathrm{E}[X] = 2(n+1)\left(\sum_{x=n+2}^{2n+3} \binom{x-1}{n+1} 2^{-x+1} - \binom{2n+2}{n+1} 2^{-2n-2}\right)$$
$$= 2(n+1)\left(1 - \frac{(2n+2)(2n+1)}{(n+1)^2}\binom{2n}{n} 2^{-2n-2}\right) = 2(n+1) - (2n+1)\binom{2n}{n} 2^{-2n} .$$

Se la v.a. Y rappresenta il numero di fiammiferi rimasti nella scatola non vuota, allora risulta $Y = 2n + 1 - X$ e dunque la media richiesta è data da

$$\mathrm{E}[Y] = 2n + 1 - \mathrm{E}[X] = (2n+1)\binom{2n}{n} 2^{-2n} - 1 \,.$$

Tenendo presente l'Esercizio 6.1.1, si ha infine

$$\mathrm{E}[Y] \sim \frac{2n+1}{\sqrt{\pi n}} - 1 \,.$$

Questo classico problema è attribuito al matematico polacco Wladyslaw Hugo Dionizy Steinhaus (1887-1972) ed è riferito al grande matematico polacco Stefan Banach (1892-1945), che era un accanito fumatore. □

Sezione 6.2

• **Esercizio 6.2.1.** Si considerino le v.a. indipendenti X_1 e X_2 distribuite rispettivamente con legge di Poisson $\mathcal{P}(\lambda_1)$ e $\mathcal{P}(\lambda_2)$. Si determini la legge condizionata della v.a. X_1 all'evento $\{X_1 + X_2 = n\}$ con $n \in \mathbb{N}$.

Soluzione. Tenendo presente la Formula di Bayes, la f.p. condizionata della v.a. X_1 all'evento $\{X_1 + X_2 = n\}$ è data da

$$\begin{aligned} p_{X_1|X_1+X_2=n}(x_1) &= P(X_1 = x_1 \mid X_1 + X_2 = n) = \frac{P(X_1 + X_2 = n \mid X_1 = x_1) P(X_1 = x_1)}{P(X_1 + X_2 = n)} \\ &= \frac{P(X_2 = n - x_1) P(X_1 = x_1)}{P(X_1 + X_2 = n)} \, \mathbf{1}_{\{0,1,\ldots,n\}}(x_1) \,. \end{aligned}$$

Inoltre, sulla base dell'Esercizio 3.7.5, la f.p. della v.a. $Y = X_1 + X_2$ è data da

$$\begin{aligned} p_Y(y) &= \sum_{x=0}^{y} e^{-\lambda_1} \frac{\lambda_1^x}{x!} \, e^{-\lambda_2} \frac{\lambda_2^{y-x}}{(y-x)!} \, \mathbf{1}_{\mathbb{N}}(y) = e^{-(\lambda_1+\lambda_2)} \frac{1}{y!} \, \mathbf{1}_{\mathbb{N}}(y) \sum_{x=0}^{y} \binom{y}{x} \lambda_1^x \lambda_2^{y-x} \\ &= e^{-(\lambda_1+\lambda_2)} \frac{(\lambda_1+\lambda_2)^y}{y!} \, \mathbf{1}_{\mathbb{N}}(y) \,, \end{aligned}$$

ovvero la v.a. $(X_1 + X_2)$ è distribuita con legge di Poisson $\mathcal{P}(\lambda_1 + \lambda_2)$. Sulla base di questi risultati si ha

$$\begin{aligned} p_{X_1|X_1+X_2=n}(x_1) &= \frac{e^{-\lambda_2} \frac{\lambda_2^{n-x_1}}{(n-x_1)!} \, e^{-\lambda_1} \frac{\lambda_1^{x_1}}{x_1!}}{e^{-(\lambda_1+\lambda_2)} \frac{(\lambda_1+\lambda_2)^n}{n!}} \, \mathbf{1}_{\{0,1,\ldots,n\}}(x_1) \\ &= \binom{n}{x_1} \left(\frac{\lambda_1}{\lambda_1+\lambda_2}\right)^{x_1} \left(1 - \frac{\lambda_1}{\lambda_1+\lambda_2}\right)^{n-x_1} \mathbf{1}_{\{0,1,\ldots,n\}}(x_1) \,, \end{aligned}$$

ovvero la v.a. X_1 condizionata all'evento $\{X_1 + X_2 = n\}$ si distribuisce con legge Binomiale $\mathcal{B}(n, p)$ dove $p = \frac{\lambda_1}{\lambda_1+\lambda_2}$. □

• **Esercizio 6.2.2.** (Problema della segretaria distratta, terza parte) Si ottenga il limite della successione $(g_n(x))_{x\geq 0}$, dove

$$g_n(x) = \begin{cases} \sum_{k=x}^{n} \binom{k}{x} \frac{(-1)^{k-x}}{k!} & x \in \{0, 1, \ldots, n\} \\ 0 & x \in \mathbb{N} \setminus \{0, 1, \ldots, n\} \,, \end{cases}$$

ovvero del prolungamento su $\mathbb{N}$ della legge considerata nell'Esercizio 2.2.1.

Soluzione. Se $x \in \{0, 1, \dots, n\}$ si ha

$$g_n(x) = \frac{1}{x!} \sum_{k=x}^{n} \frac{(-1)^{k-x}}{(k-x)!} = \frac{1}{x!} \sum_{k=0}^{n-x} \frac{(-1)^k}{k!}$$

e quindi risulta

$$\lim_n g_n(x) = \frac{1}{x!} \sum_{k=0}^{\infty} \frac{(-1)^k}{k!} = \frac{e}{x!} \, .$$

Dunque, si può considerare la v.a. X con f.p. limite data da

$$p_X(x) = \frac{e}{x!} \, \mathbf{1}_{\mathbb{N}}(x) \, ,$$

ovvero la v.a. X è distribuita con legge di Poisson $\mathcal{P}(1)$. Quindi, per n abbastanza elevato, la probabilità $P(F_m)$ considerata nell'Esercizio 2.2.1 può essere approssimata con $p_X(m)$. □

Sezione 6.3

• **Esercizio 6.3.1.** (Legge Ipergeometrica Negativa) Si considerino le v.a. indipendenti X_1 e X_2 rispettivamente con legge Binomiale Negativa $\mathcal{B}(k_1, p)$ e $\mathcal{B}(k_2, p)$. Si determini la legge condizionata della v.a. X_1 all'evento $\{X_1 + X_2 = n\}$ con $n \in \mathbb{N}$.

Soluzione. In modo simile all'Esercizio 6.2.1, la f.p. condizionata della v.a. X_1 all'evento $\{X_1 + X_2 = n\}$ è data da

$$p_{X_1|X_1+X_2=n}(x_1) = \frac{P(X_2 = n - x_1)P(X_1 = x_1)}{P(X_1 + X_2 = n)} \, \mathbf{1}_{\{0,1,\dots,n\}}(x_1) \, .$$

Dal momento che la v.a. $(X_1 + X_2)$ è distribuita con la legge Binomiale Negativa $\mathcal{B}(k_1 + k_2, p)$ (si veda l'Esempio 7.5.5), risulta

$$p_{X_1|X_1+X_2=n}(x_1) = \frac{\binom{k_1+x_1-1}{k_1-1}\binom{k_2+n-x_1-1}{k_2-1}}{\binom{k_1+k_2+n-1}{k_1+k_2-1}} \, \mathbf{1}_{\{0,1,\dots,n\}}(x_1) \, .$$

Al fine di descrivere la legge condizionata della v.a. X_1 all'evento $\{X_1 + X_2 = n\}$, si consideri un'urna composta da N palline di cui D sono rosse e $(N - D)$ sono nere. Si suppone ovviamente che D sia un intero positivo tale che $D \leq N$. L'esperimento aleatorio consiste nell'estrarre una pallina per volta in modo casuale e senza reinserimento. Le palline vengono estratte fino a quando non si ottengono m palline rosse, dove m è un intero positivo tale che $m \leq D$. Si consideri dunque la v.a. X che rappresenta il numero di palline nere estratte prima di ottenere m palline rosse. Si dimostra facilmente che la f.p. della v.a. X è data da

$$p_X(x) = \frac{\binom{m+x-1}{m-1}\binom{N-m-x}{D-m}}{\binom{N}{D}} \, \mathbf{1}_{\{0,1,\dots,N-D\}}(x) \, .$$

La legge relativa alla v.a. X è detta Ipergeometrica Negativa di parametri m, D e N. La legge condizionata della v.a. X_1 all'evento $\{X_1 + X_2 = n\}$ è quindi Ipergeometrica Negativa di parametri $m = k_1$, $D = k_1 + k_2 - 1$ e $N = k_1 + k_2 + n - 1$. Infine, si noti che per $k_1 = k_2 = 1$ si ha

$$p_{X_1|X_1+X_2=n}(x_1) = \frac{1}{n+1} \, \mathbf{1}_{\{0,1,\dots,n\}}(x_1) \, ,$$

ovvero la legge condizionata è in questo caso la legge Uniforme discreta su $\{0, 1, \dots, n\}$ (si veda l'Esercizio 7.2.2). □

• **Esercizio 6.3.2.** (Problema del collezionista di figurine) Si supponga di voler collezionare un insieme di figurine di n tipologie distinte. Le figurine sono confezionate in buste che contengono una singola figurina. Le buste vengono acquistate una per volta in modo indipendente e la probabilità che una busta contenga una figurina di una determinata tipologia è identica per tutte le tipologie ad ogni acquisto. Se la v.a. X rappresenta il numero di buste acquistate per completare la collezione, si determini $\mathrm{E}[X]$ e $\mathrm{Var}[X]$.

Soluzione. Si consideri il vettore di v.v.a. $(X_1, \dots, X_n)$, dove la componente marginale X_k rappresenta il numero di buste acquistate per collezionare la k-esima tipologia di figurina dopo che sono state collezionate $(k-1)$ tipologie. Dunque, sulla base delle assunzioni le componenti del v.v.a. sono indipendenti e si ha $X = \sum_{k=1}^{n} X_k$. Inoltre, risulta $X_k = Y_k + 1$ dove la v.a. Y_k è distribuita con legge Geometrica di parametro $\frac{n-k+1}{n}$. Quindi, si ha

$$\mathrm{E}[X] = \sum_{k=1}^{n} \mathrm{E}[Y_k + 1] = \sum_{k=1}^{n} \frac{n}{n-k+1} = n \sum_{k=1}^{n} \frac{1}{k} = nH_n \,,$$

dove H_n rappresenta l'n-esimo numero armonico. Per le proprietà dei numeri armonici si ottiene inoltre che

$$\mathrm{E}[X] = n\log(n) + \gamma n + \frac{1}{2} + O(n^{-1}) \,,$$

dove $\gamma \simeq 0.5772$ rappresenta la costante di Eulero-Mascheroni. Risulta anche

$$\begin{aligned} \mathrm{Var}[X] &= \sum_{k=1}^{n} \mathrm{Var}[Y_k] = \sum_{k=1}^{n} \frac{n(k-1)}{(n-k+1)^2} = n \sum_{k=1}^{n} \frac{n-k}{k^2} \\ &= n^2 \sum_{k=1}^{n} \frac{1}{k^2} - n \sum_{k=1}^{n} \frac{1}{k} = n^2 H_{n,2} - nH_n \,, \end{aligned}$$

dove $H_{n,2}$ rappresenta l'n-esimo numero armonico del secondo ordine. Si noti che

$$\mathrm{Var}[X] < n^2 H_{n,2} < n^2 \zeta(2) = \frac{\pi^2 n^2}{6}$$

e dunque dalla disuguaglianza di Chebyshev si ottiene anche

$$P(|X - nH_n| \geq cn) \leq \frac{\pi^2}{6c^2} \,. \qquad \square$$

• **Esercizio 6.3.3.** (Proprietà dell'assenza di memoria) Si consideri la v.a. discreta X con f.p. data da

$$p_X(x) = pq^{x-1} \mathbf{1}_{\{1,2,\dots\}}(x) \,,$$

dove $p \in]0, 1[$ e $q = 1 - p$. Si osservi che risulta $X = Z + 1$ dove Z è distribuita con legge Geometrica di parametro p e, in effetti, alcuni autori definiscono come legge Geometrica quella relativa alla v.a. X. Si verifichi che la legge della v.a. X possiede la proprietà dell'assenza di memoria, ovvero che

$$P(X > x + y \mid X > x) = P(X > y) \,,$$

dove $x, y \in \{1, 2, \dots\}$. Si verifichi inoltre che la legge considerata è l'unica a soddisfare questa proprietà fra quelle definite sui numeri naturali.

Soluzione. Dal momento che

$$P(X > x + y \mid X > x) = \frac{P(\{X > x + y\} \cap \{X > x\})}{P(X > x)} = \frac{P(X > x + y)}{P(X > x)} ,$$

posto $g(x) = P(X > x)$, la proprietà dell'assenza di memoria può essere riformulata attraverso la seguente relazione funzionale

$$g(x + y) = g(x)g(y) .$$

Tenendo presente che per $a \in]0, 1[$ si ha

$$\sum_{k=1}^{n} a^k = \frac{a(1 - a^n)}{1 - a} ,$$

per $x \in \{1, 2, \dots\}$ risulta

$$g(x) = 1 - P(X \leq x) = 1 - \sum_{u=1}^{x} pq^{u-1} = q^x .$$

Quindi, è immediato verificare che la legge considerata possiede la proprietà dell'assenza di memoria. Al fine di verificare l'unicità, si consideri una qualsiasi v.a. Y con legge essenziale definita su $\mathbb{Z}^+$ e, senza perdita di generalità, si ponga $P(Y = 1) = p$ e $q = 1 - p$. Posto $h(y) = P(Y > y)$ e tenendo presente la relazione funzionale che caratterizza l'assenza di memoria, per $y \in \{1, 2, \dots\}$ si ha

$$h(y) = h(y - 1)h(1) = h(y - 2)h(1)^2 = \dots = h(1)^y = q^y .$$

Dunque, $h = g$ e la legge della v.a. Y coincide con quella della v.a. X. Infine, si può verificare che per la v.a. Z con legge Geometrica vale una proprietà di assenza di memoria del tipo

$$P(Z > x + y \mid Z \geq x) = P(Z > y) . \qquad \square$$

Sezione 6.4

• **Esercizio 6.4.1.** Si considerino le v.a. indipendenti X_1 e X_2 rispettivamente con legge Binomiale $\mathcal{B}(n_1, p)$ e $\mathcal{B}(n_2, p)$. Si determini la legge condizionata della v.a. X_1 all'evento $\{X_1 + X_2 = k\}$ con $k \in \{0, \dots, n_1 + n_2\}$.

Soluzione. In modo simile all'Esercizio 6.2.1, la f.p. condizionata della v.a. X_1 all'evento $\{X_1 + X_2 = k\}$ è data da

$$p_{X_1|X_1+X_2=k}(x_1) = \frac{P(X_2 = k - x_1)P(X_1 = x_1)}{P(X_1 + X_2 = k)} \mathbf{1}_S(x_1) ,$$

dove $S = \{\max(0, k - n_2), \dots, \min(k, n_1)\}$. Inoltre, sulla base dell'Esercizio 3.7.5, la f.p. della v.a. $Y = X_1 + X_2$ è data da

$$\begin{aligned} p_Y(y) &= \sum_{x=0}^{y} \binom{n_1}{x} p^x (1 - p)^{n_1 - x} \binom{n_2}{y - x} p^{y-x} (1 - p)^{n_2 - y + x} \mathbf{1}_{\{0,1,\dots,y\}}(y) \\ &= p^y (1 - p)^{n_1 + n_2 - y} \mathbf{1}_{\{0,1,\dots,y\}}(y) \sum_{x=0}^{y} \binom{n_1}{x} \binom{n_2}{y - x} \\ &= \binom{n_1 + n_2}{y} p^y (1 - p)^{n_1 + n_2 - y} \mathbf{1}_{\{0,1,\dots,y\}}(y) , \end{aligned}$$

ovvero la v.a. $(X_1 + X_2)$ è distribuita con legge Binomiale $\mathcal{B}(n_1 + n_2, p)$. Dunque, sostituendo opportunamente risulta

$$p_{X_1|X_1+X_2=k}(x_1) = \frac{\binom{n_1}{x_1}\binom{n_2}{k-x_1}}{\binom{n_1+n_2}{k}} \mathbf{1}_S(x_1) .$$

La legge condizionata della v.a. X_1 all'evento $\{X_1 + X_2 = k\}$ è dunque la legge Ipergeometrica $\mathcal{I}(k, n_1, n_1 + n_2)$. □

Sezione 6.5

• **Esercizio 6.5.1.** Si consideri il v.v.a. $X = (X_1, \ldots, X_k)^{\mathrm{T}}$ con componenti marginali indipendenti e tali che la v.a. X_j è distribuita con legge di Poisson $\mathcal{P}(\lambda_j)$. Si determini la legge condizionata del v.v.a. X all'evento $\{\sum_{j=1}^k X_j = n\}$ con $n \in \mathbb{N}$.

Soluzione. Tenendo presenti le considerazioni fatte nell'Esercizio 6.2.1, la f.p.c. condizionata del v.v.a. X all'evento $\{\sum_{j=1}^k X_j = n\}$ è data da

$$p_{X|\sum_{j=1}^k X_j=n}(x) = \frac{\prod_{j=1}^k P(X_j = x_j)}{P(\sum_{j=1}^k X_j = n)} \mathbf{1}_S(x) ,$$

dove $x = (x_1, \ldots, x_k)^{\mathrm{T}}$ e $S = \{(x_1, \ldots, x_k) : x_j \in \{0, 1, \ldots, n\}, \sum_{j=1}^k x_j = n\}$. Dal momento che la v.a. $\sum_{j=1}^k X_j$ si distribuisce con legge di Poisson $\mathcal{P}(\sum_{j=1}^k \lambda_j)$ (si veda l'Esempio 7.5.4), risulta

$$p_{X|\sum_{j=1}^k X_j}(x) = \frac{\prod_{j=1}^k e^{-\lambda_j} \frac{\lambda_j^{x_j}}{x_j!}}{e^{-\sum_{j=1}^k \lambda_j} \frac{(\sum_{j=1}^k \lambda_j)^n}{n!}} \mathbf{1}_S(x) = \binom{n}{x_1 \ldots x_k} \prod_{j=1}^k \left(\frac{\lambda_j}{\sum_{j=1}^k \lambda_j}\right)^{x_j} \mathbf{1}_S(x) .$$

La legge condizionata del v.v.a. X all'evento $\{\sum_{j=1}^k X_j = n\}$ è quindi Multinomiale $\mathcal{M}(n, p)$, dove $p = (p_1, \ldots, p_k)^{\mathrm{T}}$ con $p_j = \frac{\lambda_j}{\sum_{j=1}^k \lambda_j}$. □

• **Esercizio 6.5.2.** Si consideri il lancio di 12 dadi da gioco equilibrati, supponendo che i lanci siano fatti in modo indipendente. Si determini la probabilità che ogni faccia si presenti due volte.

Soluzione. Si consideri il v.v.a. $X = (X_1, \ldots, X_6)^{\mathrm{T}}$, dove la v.a. X_k rappresenta il numero di volte che si è presentata la faccia contrassegnata con k punti. La legge del v.v.a. X è Multinomiale di parametri 12 e $(\frac{1}{6}, \ldots, \frac{1}{6})^{\mathrm{T}}$. Quindi, se p_X rappresenta la f.p.c. del v.v.a. X, la probabilità richiesta è data da

$$p_X(2, \ldots, 2) = \binom{12}{2 \ldots 2} \left(\frac{1}{6}\right)^{12} = \frac{1925}{559872} \simeq 0.003 .$$

Sezione 6.6

• **Esercizio 6.6.1.** (Legge di Burr del terzo tipo) Si consideri la v.a. X con f.r.

$$F_X(x) = \left(1 + \left(\frac{x}{b}\right)^{-p}\right)^{-q} \mathbf{1}_{]0,\infty[}(x) ,$$

dove $b, p, q \in]0, \infty[$. Si verifichi che F_X è una f.r.

Soluzione. Si osservi che F_X è effettivamente la f.r. di una v.a. assolutamente continua dal momento che

$$F_X(x) = \int_0^x \frac{pq}{b} \left(\frac{u}{b}\right)^{-p-1} \left(1 + \left(\frac{u}{b}\right)^{-p}\right)^{-q-1} du \,,$$

ovvero X ammette d.p. data da

$$f_X(x) = \frac{pq}{b} \left(\frac{x}{b}\right)^{-p-1} \left(1 + \left(\frac{x}{b}\right)^{-p}\right)^{-q-1} \mathbf{1}_{]0,\infty[}(x) \,.$$

Risulta immediato verificare che il parametro di posizione è dato da $a = 0$. Dunque, se si considera la trasformazione $Z = \frac{X}{b}$, si ha

$$F_Z(z) = (1 + z^{-p})^{-q} \, \mathbf{1}_{]0,\infty[}(z) \,,$$

ovvero la f.r. della v.a. standardizzata Z non dipende dal parametro b, che quindi risulta essere un parametro di scala. Dunque, p e q sono parametri di forma. La legge prende il nome dallo statistico Irving Wingate Burr (1908-1989). □

Sezione 6.7

• **Esercizio 6.7.1.** Si consideri la v.a. X con legge Normale $\mathcal{N}(0, \sigma^2)$ e si determini $\mathrm{E}[e^{aX}]$ dove $a \in \mathbb{R}$.

Soluzione. Mediante il completamento di quadrato, si ha

$$\mathrm{E}[e^{aX}] = \int_{-\infty}^{\infty} e^{ax} \, \frac{1}{\sqrt{2\pi\sigma}} \, e^{-\frac{x^2}{2\sigma^2}} \, dx = e^{\frac{1}{2}a^2\sigma^2} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma}} \, e^{-\frac{1}{2\sigma^2}(x - a\sigma^2)^2} \, dx \,.$$

Si osservi che la funzione integranda nell'ultima espressione è la d.p. di una v.a. con legge Normale $\mathcal{N}(a\sigma^2, \sigma^2)$ e dunque risulta

$$\mathrm{E}[e^{aX}] = e^{\frac{1}{2}a^2\sigma^2} \,. \qquad \square$$

• **Esercizio 6.7.2.** (Legge Normale asimmetrica) Si consideri la v.a. X che ammette d.p. data da

$$f_X(x) = 2\phi(x)\Phi(\alpha x) \,,$$

dove $\alpha \in \mathbb{R}$. Si verifichi che f_X è in effetti una d.p.

Soluzione. Si noti che f_X è una funzione a valori positivi. Inoltre, risulta

$$\begin{aligned} \int_{-\infty}^{\infty} f_X(x) \, dx &= 2 \int_{-\infty}^{0} \phi(x)\Phi(\alpha x) \, dx + 2 \int_0^{\infty} \phi(x)\Phi(\alpha x) \, dx \\ &= 2 \int_0^{\infty} \phi(-x)\Phi(-\alpha x) \, dx + 2 \int_0^{\infty} \phi(x)\Phi(\alpha x) \, dx \\ &= 2 \int_0^{\infty} \phi(x)\Phi(-\alpha x) \, dx + 2 \int_0^{\infty} \phi(x)\Phi(\alpha x) \, dx \,, \end{aligned}$$

essendo ϕ una funzione pari. Quindi, tenendo presente che $\Phi(x) + \Phi(-x) = 1$, si ha

$$\int_{-\infty}^{\infty} f_X(x) \, dx = 2 \int_0^{\infty} \phi(x)(\Phi(\alpha x) + \Phi(-\alpha x)) \, dx = 1 \,,$$

ovvero f_X è una d.p. La legge associata alla v.a. X è detta Normale asimmetrica e α è un parametro di forma legato all'asimmetria della distribuzione. Evidentemente, la legge Normale ridotta è un caso particolare di questa famiglia e viene ottenuta per $\alpha = 0$. □

• **Esercizio 6.7.3.** (Legge Lognormale) Si consideri la v.a. X tale che $X = e^Z$, dove Z è una v.a. con legge Normale ridotta $\mathcal{N}(0,1)$ e se ne determini la d.p. f_X. Inoltre, considerata la v.a. assolutamente continua Y con d.p. data da

$$f_Y(y) = f_X(y)(1 + \sin(2\pi\log(y)))\,,$$

si verifichi che $\mathrm{E}[X^r] = \mathrm{E}[Y^r]$ per ogni $r \in \mathbb{N}$.

Soluzione. Dal momento che la v.a. Z ammette densità data da ϕ e che la funzione $g(z) = e^z$ è biunivoca, allora si ha

$$f_X(x) = \phi(\log(x))\,\frac{1}{x}\,\mathbf{1}_{]0,\infty[}(x) = \frac{1}{x\sqrt{2\pi}}\,e^{-\frac{1}{2}\log^2(x)}\,\mathbf{1}_{]0,\infty[}(x)\ .$$

La legge della v.a. X è detta Lognormale ridotta. Inoltre, per ogni $r = 0, 1, \ldots$ risulta

$$\begin{aligned}\int_0^\infty y^r \sin(2\pi\log(y)) f_X(y)\,dy &= \int_0^\infty y^{r-1}\sin(2\pi\log(y))\,\phi(\log(y))\,dy \\ &= \int_{-\infty}^\infty e^{rz}\sin(2\pi z)\,\phi(z)\,dz = e^{\frac{r^2}{2}}\int_{-\infty}^\infty \sin(2\pi z)\,\phi(z-r)\,dz\,,\end{aligned}$$

dove si è considerato un completamento del quadrato in modo simile all'Esempio 6.7.1. Dunque, si ha

$$\begin{aligned}\int_0^\infty y^r \sin(2\pi\log(y)) f_X(y)\,dy &= e^{\frac{1}{2}r^2}\int_{-\infty}^\infty \sin(2\pi(z-r))\,\phi(z-r)\,dz \\ &= e^{\frac{1}{2}r^2}\int_{-\infty}^\infty \sin(2\pi z)\,\phi(z)\,dz = e^{\frac{1}{2}r^2}\,\mathrm{E}[\sin(2\pi Z)] = 0\,,\end{aligned}$$

dove si è tenuto presente che $\sin(2\pi z) = \sin(2\pi(z-r))$. Quindi, segue che $\mathrm{E}[X^r] = \mathrm{E}[Y^r]$ per ogni $r \in \mathbb{N}$, ovvero si è ottenuto un esempio di v.a. con leggi differenti e per cui tutti i momenti sono coincidenti. □

Sezione 6.8

• **Esercizio 6.8.1.** (Legge di Laplace) Si considerino le v.a. indipendenti X_1 e X_2 con legge Esponenziale ridotta $\mathcal{G}(0,1,1)$ e si determini la d.p. della v.a. $Y = X_1 - X_2$.

Soluzione. Sulla base dell'espressione della d.p. per la v.a. differenza di v.a. assolutamente continue, la d.p. di Y è data da

$$f_Y(y) = \int_{-\infty}^\infty f_{X_1}(x) f_{X_2}(x+y)\,dx = \int_{-\infty}^\infty e^{-x}\,\mathbf{1}_{]0,\infty[}(x)\,e^{-x-y}\,\mathbf{1}_{]0,\infty[}(x+y)\,dx$$

ovvero, dal momento che $\mathbf{1}_{]0,\infty[}(x)\mathbf{1}_{]0,\infty[}(x+y) = \mathbf{1}_{]\min(0,-y),\infty[}(x)$, si ha

$$f_Y(y) = e^{-y}\int_{\min(0,-y)}^\infty e^{-2x}\,dx = \frac{1}{2}\,e^{-y-2\min(0,-y)} = \frac{1}{2}\,e^{-|y|}\ .$$

La v.a. Y si distribuisce con la cosiddetta legge di Laplace. □

• **Esercizio 6.8.2.** Si consideri la v.a. X con legge Esponenziale ridotta. Si verifichi che la legge della v.a. X possiede la proprietà dell'assenza di memoria, ovvero che

$$P(X > x + y \mid X > x) = P(X > y) \,,$$

dove $x, y \in \mathbb{R}^+$.

Soluzione. Tenendo presente l'Esercizio 6.3.3 e assumendo $g(x) = P(X > x)$, la proprietà dell'assenza di memoria può essere di nuovo espressa attraverso la relazione funzionale $g(x + y) = g(x)g(y)$. Dunque, per $x \in \mathbb{R}^+$ si ha

$$g(x) = 1 - P(X \leq x) = e^{-x}$$

ed è immediato verificare che la legge considerata possiede la proprietà dell'assenza di memoria. Si può verificare che la legge Esponenziale è l'unica a soddisfare questa proprietà fra quelle relative a v.a. assolutamente continue definite su $\mathbb{R}^+$. La dimostrazione è basata sulla soluzione dell'equazione funzionale di Cauchy, con l'assunzione che g sia una funzione monotona decrescente. □

• **Esercizio 6.8.3.** Si consideri il v.v.a. $X = (X_1, \ldots, X_k)^{\mathrm{T}}$ dove le v.a. marginali sono indipendenti e la legge della v.a. X_j è Gamma ridotta $\mathcal{G}(0, 1, \alpha + \frac{j-1}{k})$. Si verifichi che la legge della v.a. Y, dove

$$Y^k = k^k \prod_{j=1}^{k} X_j \,,$$

è Gamma ridotta $\mathcal{G}(0, 1, k\alpha)$.

Soluzione. Si tenga presente che la legge Gamma è univocamente determinata dalla successione dei momenti. Inoltre, il momento di ordine r della v.a. X_j è dato da

$$\mathrm{E}[X_j^r] = \frac{1}{\Gamma\left(\alpha + \frac{j-1}{k}\right)} \int_0^\infty x^{\alpha + r + \frac{j-1}{k} - 1} \, e^{-x} \, dx = \frac{\Gamma\left(\alpha + r + \frac{j-1}{k}\right)}{\Gamma\left(\alpha + \frac{j-1}{k}\right)} \,.$$

Dunque, tenendo presente la formula di moltiplicazione per la funzione Gamma, ovvero

$$\Gamma(kz) = \frac{k^{kz - \frac{1}{2}}}{(2\pi)^{\frac{1}{2}(k-1)}} \prod_{j=1}^{k} \Gamma\left(z + \frac{j-1}{k}\right) \,,$$

risulta

$$\mathrm{E}[Y^{kr}] = \mathrm{E}[k^{kr} \prod_{j=1}^{k} X_j^r] = k^{kr} \prod_{j=1}^{k} \mathrm{E}[X_j^r] = k^{kr} \prod_{j=1}^{k} \frac{\Gamma\left(\alpha + r + \frac{j-1}{k}\right)}{\Gamma\left(\alpha + \frac{j-1}{k}\right)} = \frac{\Gamma(k\alpha + kr)}{\Gamma(k\alpha)} \,.$$

Dunque, dal momento che $k \in \mathbb{Z}^+$ anche $s = kr \in \mathbb{Z}^+$ e si ha

$$\mathrm{E}[Y^s] = \frac{\Gamma(k\alpha + s)}{\Gamma(k\alpha)} \,,$$

che in effetti rappresenta il momento di ordine s di una v.a. distribuita con legge Gamma ridotta $\mathcal{G}(0, 1, k\alpha)$. □

Sezione 6.9

• **Esercizio 6.9.1.** Si consideri il v.v.a. $X = (X_1, X_2)^{\mathrm{T}}$ dove X_1 e X_2 sono v.a. indipendenti rispettivamente con legge Beta ridotta $\mathcal{BE}(0, 1, \alpha, \beta)$ e $\mathcal{BE}(0, 1, \alpha + \beta, \gamma)$. Si verifichi che la legge della v.a. $Y = X_1 X_2$ è Beta ridotta $\mathcal{BE}(0, 1, \alpha, \beta + \gamma)$.

Soluzione. Si tenga presente che la legge Beta è univocamente determinata dalla successione dei momenti. Inoltre, il momento di ordine r della v.a. X_1 è dato da

$$\mathrm{E}[X_1^r] = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \int_0^1 x^{\alpha+r-1}(1-x)^{\beta-1}\,dx = \frac{\Gamma(\alpha+\beta)\Gamma(\alpha+r)}{\Gamma(\alpha)\Gamma(\alpha+\beta+r)}\,.$$

In modo simile, si ha

$$\mathrm{E}[X_2^r] = \frac{\Gamma(\alpha+\beta+\gamma)\Gamma(\alpha+\beta+r)}{\Gamma(\alpha+\beta)\Gamma(\alpha+\beta+\gamma+r)}\,.$$

Inoltre, si ha

$$\mathrm{E}[Y^r] = \mathrm{E}[X_1^r X_2^r] = \mathrm{E}[X_1^r]\mathrm{E}[X_2^r] = \frac{\Gamma(\alpha+\beta+\gamma)\Gamma(\alpha+r)}{\Gamma(\alpha)\Gamma(\alpha+\beta+\gamma+r)}\,,$$

che è in effetti il momento di ordine r di una v.a. con legge Beta ridotta $\mathcal{BE}(0,1,\alpha,\beta+\gamma)$. □

Sezione 6.10

• **Esercizio 6.10.1.** Si consideri la v.a. X distribuita con legge $F_{n,n}$ di Snedecor. Si verifichi che la v.a.

$$Y = \frac{\sqrt{n}}{2}\left(\sqrt{X} - \frac{1}{\sqrt{X}}\right)$$

è distribuita con legge t_n di Student.

Soluzione. La v.a. X ammette d.p. data da

$$f_X(x) = \frac{\Gamma(n)}{\Gamma\left(\frac{n}{2}\right)^2}\, x^{\frac{1}{2}n-1}(1+x)^{-n}\,\mathbf{1}_{]0,\infty[}(x)\,.$$

Si consideri inizialmente la trasformata $U = \sqrt{X}$, per cui è immediato ottenere che

$$f_U(u) = \frac{2\Gamma(n)}{\Gamma\left(\frac{n}{2}\right)^2}\, u^{n-1}(1+u^2)^{-n}\,\mathbf{1}_{]0,\infty[}(u)\,.$$

Successivamente, si consideri la trasformata $V = \frac{1}{2}(U - \frac{1}{U})$. Si noti che la funzione $g(u) = \frac{1}{2}(u - \frac{1}{u})$ è biunivoca per $u \in]0,\infty[$, e si ha $g^{-1}(v) = v + \sqrt{1+v^2}$. Dal momento che risulta

$$\left|\frac{d}{dv}\, g^{-1}(v)\right| = 1 + \frac{v}{\sqrt{1+v^2}}\,,$$

allora si ha

$$f_V(v) = \frac{\Gamma(n)}{2^{n-1}\Gamma\left(\frac{n}{2}\right)^2}\,(1+v^2)^{-\frac{1}{2}(n+1)}\,.$$

Inoltre, sulla base della formula di duplicazione per la funzione Gamma, si ha la relazione

$$\Gamma(n) = \frac{2^{n-1}}{\sqrt{\pi}}\,\Gamma\left(\frac{n}{2}\right)\Gamma\left(\frac{n+1}{2}\right)$$

e, considerando la trasformata $Y = \sqrt{n}V$, si ha infine

$$f_Y(y) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\,\Gamma\left(\frac{n}{2}\right)} \left(1 + \frac{y^2}{n}\right)^{-\frac{1}{2}(n+1)} ,$$

ovvero la v.a. Y è distribuita con legge t_n di Student. □

Sezione 6.11

• **Esercizio 6.11.1.** Si consideri il v.v.a. $X = (X_1, X_2)^{\mathrm{T}}$ distribuito con legge Normale Multivariata $\mathcal{N}_2(\mu, \Sigma)$, dove

$$\mu = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \Sigma = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix},$$

e $\rho \in [-1, 1]$. Si determini la d.p. della v.a. $Y = \arctan(\frac{X_2}{X_1})$.

Soluzione. Dal momento che $\det(\Sigma) = 1 - \rho^2$ e

$$\Sigma^{-1} = \frac{1}{1-\rho^2} \begin{pmatrix} 1 & -\rho \\ -\rho & 1 \end{pmatrix},$$

la d.p. del v.v.a. X è data da

$$f_X(x_1, x_2) = \frac{1}{2\pi\sqrt{1-\rho^2}}\, e^{-\frac{1}{2(1-\rho^2)}(x_1^2 - 2\rho x_1 x_2 + x_2^2)} .$$

Sulla base dei risultati sul rapporto di v.a. assolutamente continue, la d.p. della v.a. $Z = \frac{X_2}{X_1}$ è data da

$$\begin{aligned} f_Z(z) &= \int_{-\infty}^{\infty} f_X(x, xz)\,|x|\,dx = \frac{1}{2\pi\sqrt{1-\rho^2}} \int_{-\infty}^{\infty} |x|\, e^{-\frac{1}{2(1-\rho^2)}(1 - 2\rho z + z^2)x^2}\,dx \\ &= \frac{\sqrt{1-\rho^2}}{2\pi(1 - 2\rho z + z^2)} \int_{-\infty}^{\infty} |x|\, e^{-\frac{1}{2}x^2}\,dx = \frac{\sqrt{1-\rho^2}}{\pi(1 - 2\rho z + z^2)} . \end{aligned}$$

Dal momento che $Y = \arctan(Z)$ e che la funzione $g(z) = \arctan(z)$ è biunivoca, allora si ha

$$f_Y(y) = f_Z(\tan(y))\, \frac{1}{\cos^2(y)}\, \mathbf{1}_{[-\frac{\pi}{2}, \frac{\pi}{2}]}(y) = \frac{\sqrt{1-\rho^2}}{\pi(1 - \rho\sin(2y))}\, \mathbf{1}_{[-\frac{\pi}{2}, \frac{\pi}{2}]}(y) . \qquad \square$$

• **Esercizio 6.11.2.** Si consideri il v.v.a. $X = (X_1, X_2)^{\mathrm{T}}$ con legge Normale Multivariata $\mathcal{N}_2(\mu, \Sigma)$ introdotto nell'Esercizio 6.11.1. Si determini la legge condizionata della v.a. X_2 all'evento $\{X_1 = x_1\}$ e il valore atteso condizionato $\mathrm{E}[X_2 \mid X_1 = x_1]$.

Soluzione. Tenendo presente l'espressione della d.p. del v.v.a. X considerato nell'Esercizio 6.11.1, si ha

$$\begin{aligned} f_{X_1}(x_1) &= \int_{-\infty}^{\infty} \frac{1}{2\pi\sqrt{1-\rho^2}}\, e^{-\frac{1}{2(1-\rho^2)}(x_1^2 - 2\rho x_1 x_2 + x_2^2)}\,dx_2 \\ &= \frac{1}{\sqrt{2\pi}}\, e^{-\frac{1}{2}x_1^2} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi(1-\rho^2)}}\, e^{-\frac{1}{2(1-\rho^2)}(x_2 - \rho x_1)^2}\,dx_2 . \end{aligned}$$

Si osservi che la funzione integranda nell'ultima espressione è la d.p. di una v.a. distribuita con legge Normale $\mathcal{N}(\rho x_1, 1-\rho^2)$ e dunque risulta

$$f_{X_1}(x_1) = \frac{1}{\sqrt{2\pi}}\, e^{-\frac{1}{2}x_1^2}\,,$$

ovvero la v.a. X_1 è distribuita con legge Normale $\mathcal{N}(0,1)$. Per simmetria anche la v.a. X_2 è equivalente in legge alla v.a. X_1. Dunque, si ha

$$f_{X_2|X_1=x_1}(x_2) = \frac{f_X(x_1,x_2)}{f_{X_1}(x_1)} = \frac{1}{\sqrt{2\pi(1-\rho^2)}}\, e^{-\frac{1}{2(1-\rho^2)}(x_2-\rho x_1)^2}\,,$$

ovvero la legge condizionata richiesta è Normale $\mathcal{N}(\rho x_1, 1-\rho^2)$, un risultato desumibile dalla prima espressione. Ovviamente, si ha $\mathrm{E}[X_2 \mid X_1 = x_1] = \rho x_1$ e dunque il valore atteso condizionato è una funzione lineare. Infine, si noti che $\mathrm{Var}[X_2 \mid X_1 = x_1] = 1-\rho^2$ è una costante. □

Sezione 6.12

• **Esercizio 6.12.1.** (Statistica ordinata) Dato il v.v.a. assolutamente continuo $Z = (Z_1, \ldots, Z_{n+1})^{\mathrm{T}}$ distribuito con legge di Dirichlet $\mathcal{D}(\alpha)$ dove $\alpha = (1, \ldots, 1)^{\mathrm{T}}$, si determini la d.p.c. della trasformata $Y = (Y_1, \ldots, Y_{n+1})^{\mathrm{T}} = g(Z)$, dove

$$g(z) = (z_1, z_1 + z_2, \ldots, z_1 + \ldots + z_{n+1})^{\mathrm{T}}\,.$$

Si determini inoltre le leggi delle componenti marginali del v.v.a. Z.

Soluzione. Dal momento che risulta $g^{-1}(y) = (y_1, y_2 - y_1, \ldots, y_{n+1} - y_n)^{\mathrm{T}}$, si ha $|J(g^{-1}(y))| = 1$. Dunque, risulta

$$f_Y(y_1, \ldots, y_{n+1}) = n!\, \mathbf{1}_S(y_1, \ldots, y_{n+1})\,,$$

dove $S = \{(y_1, \ldots, y_{n+1}) : 0 \le y_1 \le y_2 \le \ldots \le y_{n+1} = 1\}$. Inoltre, per le proprietà della legge di Dirichlet si ha

$$Y_k = \frac{\sum_{l=1}^{k} X_l}{\sum_{l=1}^{n+1} X_l}\,,$$

dove $(X_1, \ldots, X_{n+1})^{\mathrm{T}}$ è un v.v.a. con componenti indipendenti distribuiti con la medesima legge Esponenziale ridotta $\mathcal{G}(0,1,1)$. Evidentemente, la v.a. Y_{n+1} è degenere, ovvero $P(Y_{n+1} = 1) = 1$. Per $k = 1, \ldots, n$, tenendo presente le proprietà della legge Beta, la v.a. Y_k si distribuisce dunque con legge Beta ridotta di parametri k e $(n+1-k)$, ovvero ammette d.p. data da

$$f_{Y_k}(y) = \frac{\Gamma(n+1)}{\Gamma(k)\Gamma(n+1-k)}\, y^{k-1}(1-y)^{n-k}\, \mathbf{1}_{]0,1[}(y) = n \binom{n-1}{k-1} y^{k-1}(1-y)^{n-k}\, \mathbf{1}_{]0,1[}(y)\,.$$

Si osservi che il v.v.a. Y risulta fondamentale nella teoria relativa alla statistica ordinata nell'inferenza statistica. □

Capitolo 7

Funzioni caratteristiche e generatrici

7.1. Funzioni caratteristiche

La funzione caratteristica di una v.a. è una particolare funzione a valori complessi, le cui proprietà analitiche sono molto utili per ottenere leggi di probabilità o momenti di v.a. trasformate. Nel prossimo Capitolo, la funzione caratteristica sarà anche centrale nello studio del comportamento asintotico di successioni di v.a.

Prima di introdurre la definizione di funzione caratteristica è conveniente estendere il concetto di v.a. all'insieme dei numeri complessi. Per quanto riguarda la notazione adottata in questo Capitolo, $\mathrm{i} = \sqrt{-1}$ rappresenta come al solito l'unità immaginaria. Inoltre, dato il numero complesso $c = a + \mathrm{i}b$, allora la parte reale e immaginaria di c sono rispettivamente denotate con $\Re(c) = a$ e $\Im(c) = b$, mentre il modulo di c è dato da $|c| = \sqrt{a^2 + b^2}$.

Definizione 7.1.1. Se $(\Omega, \mathcal{F}, P)$ è uno spazio probabilizzato, un'applicazione $X : \Omega \to \mathbb{C}$ è detta *variabile aleatoria complessa* se $\Re(X)$ e $\Im(X)$ sono variabili aleatorie. □

Se X è una v.a. complessa, allora il relativo valore atteso è definito come

$$\mathrm{E}[X] = \mathrm{E}[\Re(X)] + \mathrm{i}\,\mathrm{E}[\Im(X)] \,.$$

Affinchè il valore atteso di una v.a. complessa esista finito è sufficiente che il valore atteso

$$\mathrm{E}[|X|] = \mathrm{E}[\sqrt{\Re(X)^2 + \Im(X)^2}]$$

esista finito. Infatti, le proprietà del valore atteso di una v.a. complessa sono simili a quelle di una v.a. reale e in particolare si ha $|\mathrm{E}[X]| \leq \mathrm{E}[|X|]$ (si veda Billingsley, 1995, p.218). Inoltre, vale anche $\mathrm{E}[cX] = c\mathrm{E}[X]$ per $c \in \mathbb{C}$.

La definizione formale di funzione caratteristica è quindi la seguente.

Definizione 7.1.2. Data la v.a. X definita sullo spazio probabilizzato $(\Omega, \mathcal{F}, P)$, si dice *funzione caratteristica* (f.c.) della v.a. X la funzione $\varphi_X : \mathbb{R} \to \mathbb{C}$

$$\varphi_X(t) = \mathrm{E}[\cos(tX)] + \mathrm{i}\,\mathrm{E}[\sin(tX)] = \mathrm{E}[e^{\mathrm{i}tX}] \,,$$

dove $t \in \mathbb{R}$. □

Nel linguaggio della Teoria della Misura, la f.c. è in effetti la trasformata di Fourier (a meno di una costante moltiplicativa). Tenendo presente quanto detto nella Sezione 4.1, se la v.a. X è discreta a valori su insieme numerabile S con f.p. p_X, la f.c. si riduce a

$$\varphi_X(t) = \sum_{x \in S} \cos(tx)\, p_X(x) + \mathrm{i} \sum_{x \in S} \sin(tx)\, p_X(x) = \sum_{x \in S} e^{\mathrm{i}tX}\, p_X(x) \,,$$

mentre, se la v.a. assolutamente continua X ammette d.p. f_X, la f.c. si riduce a

$$\varphi_X(t) = \int_{-\infty}^{\infty} \cos(tx)\, f_X(x)\, dx + \mathrm{i} \int_{-\infty}^{\infty} \sin(tx)\, f_X(x)\, dx = \int_{-\infty}^{\infty} e^{\mathrm{i}tx}\, f_X(x)\, dx\,.$$

• **Esempio 7.1.1.** Si consideri la v.a. X distribuita con legge di Bernoulli $\mathcal{B}(1, p)$. In questo caso, la f.c. della v.a. X è data da

$$\varphi_X(t) = \sum_{x=0}^{1} \cos(tx)\, p^x q^{1-x} + \mathrm{i} \sum_{x=0}^{1} \sin(tx)\, p^x q^{1-x} = q + p\cos(t) + \mathrm{i}p\sin(t) = q + pe^{\mathrm{i}t}\,.$$

Se si considera più in generale la v.a. X distribuita con legge Binomiale $\mathcal{B}(n, p)$, tenendo presente l'espressione del binomio di Newton in $\mathbb{C}$, si ha

$$\varphi_X(t) = \sum_{x=0}^{n} e^{\mathrm{i}tx} \binom{n}{x} p^x q^{n-x} = \sum_{x=0}^{n} \binom{n}{x} (pe^{\mathrm{i}t})^x q^{n-x} = (q + pe^{\mathrm{i}t})^n\,. \qquad \square$$

• **Esempio 7.1.2.** Si consideri la v.a. X distribuita con legge di Poisson $\mathcal{P}(\lambda)$. Tenendo presente l'espressione della serie esponenziale in $\mathbb{C}$, si ha

$$\varphi_X(t) = \sum_{x=0}^{\infty} e^{\mathrm{i}tx}\, e^{-\lambda}\, \frac{\lambda^x}{x!} = e^{-\lambda} \sum_{x=0}^{\infty} \frac{(\lambda e^{\mathrm{i}t})^x}{x!} = e^{\lambda(e^{\mathrm{i}t}-1)}\,. \qquad \square$$

• **Esempio 7.1.3.** Si consideri la v.a. X distribuita con legge Binomiale Negativa $\mathcal{BN}(k, p)$. Tenendo presente l'espressione della serie binomiale negativa in $\mathbb{C}$, si ha

$$\varphi_X(t) = \sum_{x=0}^{\infty} e^{\mathrm{i}tx} \binom{k+x-1}{k-1} p^k q^x = p^k \sum_{x=0}^{\infty} \binom{k+x-1}{x} (qe^{\mathrm{i}t})^x = \left(\frac{p}{1-qe^{\mathrm{i}t}}\right)^k\,. \qquad \square$$

• **Esempio 7.1.4.** Si consideri la v.a. Z distribuita con legge Normale ridotta $\mathcal{N}(0, 1)$. Tenendo presente che Z è simmetrica rispetto all'origine, si ha

$$\mathrm{E}[\sin(tZ)] = \int_{-\infty}^{\infty} \sin(tz)\, \frac{1}{\sqrt{2\pi}}\, e^{-\frac{1}{2}z^2}\, dz = 0\,,$$

e quindi

$$\varphi_Z(t) = \mathrm{E}[\cos(tZ)] = \int_{-\infty}^{\infty} \cos(tz)\, \frac{1}{\sqrt{2\pi}}\, e^{-\frac{1}{2}z^2}\, dz = e^{-\frac{1}{2}t^2}\,.$$

Per ulteriori dettagli su questo risultato si veda l'Esercizio 7.1.1. Si osservi che in questo caso $\varphi_Z(t)$ è una funzione puramente reale. In generale, risulta immediato verificare che la f.c. è una funzione puramente reale se una v.a. è simmetrica rispetto all'origine. $\square$

• **Esempio 7.1.5.** Si consideri la v.a. Z distribuita con legge Esponenziale ridotta $\mathcal{G}(0, 1, 1)$. Dal momento che

$$\mathrm{E}[\cos(tZ)] = \int_{0}^{\infty} \cos(tz)\, e^{-z}\, dz = \frac{1}{1+t^2}$$

e

$$\mathrm{E}[\sin(tZ)] = \int_0^\infty \sin(tz)\, e^{-z}\, dz = \frac{t}{1+t^2}\,,$$

allora si ha

$$\varphi_Z(t) = \frac{1}{1+t^2} + \mathrm{i}\,\frac{t}{1+t^2} = (1-\mathrm{i}t)^{-1}\,.$$

Più in generale, se si considera la v.a. Z distribuita con legge Gamma ridotta $\mathcal{G}(0,1,k)$, tenendo presente le espressioni delle serie esponenziale e binomiale negativa in $\mathbb{C}$, si ha

$$\begin{aligned}\varphi_Z(t) &= \int_0^\infty e^{\mathrm{i}tz}\,\frac{1}{\Gamma(k)}\,z^{k-1}\,e^{-z}\,dz = \frac{1}{\Gamma(k)}\int_0^\infty \sum_{n=0}^\infty \frac{(\mathrm{i}tz)^n}{n!}\,z^{k-1}\,e^{-z}\,dz \\ &= \frac{1}{\Gamma(k)}\sum_{n=0}^\infty \frac{(\mathrm{i}t)^n}{n!}\int_0^\infty z^{n+k-1}\,e^{-z}\,dz = \sum_{n=0}^\infty \frac{\Gamma(n+k)}{n!\Gamma(k)}\,(\mathrm{i}t)^n \\ &= \sum_{n=0}^\infty \binom{n+k-1}{n}\,(\mathrm{i}t)^n = (1-\mathrm{i}t)^{-k}\,.\end{aligned}$$

Ovviamente, l'espressione coincide con quella ottenuta per la legge Esponenziale ridotta se $k=1$. □

Di seguito vengono considerate una serie di proprietà che riguardano l'esistenza della f.c., la continuità uniforme della f.c. e l'espressione della f.c. di particolari v.a. trasformate.

Proposizione 7.1.3. *Se X è una v.a. definita sullo spazio probabilizzato $(\Omega,\mathcal{F},P)$ con f.c. φ_X, allora $|\varphi_X(t)| \le 1$ per ogni $t\in\mathbb{R}$ e in particolare $\varphi_X(0)=1$.*

Dimostrazione. Tenendo presente che per ogni $x\in\mathbb{R}$ si ha

$$|e^{\mathrm{i}tx}| = |\cos(tx) + \mathrm{i}\sin(tx)| = \sqrt{\cos(tx)^2+\sin(tx)^2} = 1\,,$$

allora

$$|\varphi_X(t)| = |\mathrm{E}[e^{\mathrm{i}tX}]| \le \mathrm{E}[|e^{\mathrm{i}tX}|] = 1\,.$$

Inoltre, si ha

$$|\varphi_X(0)| = \int dP = 1\,.$$ □

Proposizione 7.1.4. *Se X è una v.a. definita sullo spazio probabilizzato $(\Omega,\mathcal{F},P)$ con f.c. φ_X, allora φ_X è uniformemente continua su $\mathbb{R}$.*

Dimostrazione. Per $h\in\mathbb{R}$ si ha

$$\begin{aligned}|\varphi_X(t+h)-\varphi_X(t)| &= |\mathrm{E}[e^{\mathrm{i}tX}(e^{\mathrm{i}hX}-1)]| \le \mathrm{E}[|e^{\mathrm{i}tX}(e^{\mathrm{i}hX}-1)|] \\ &= \mathrm{E}[|e^{\mathrm{i}tX}|\cdot|e^{\mathrm{i}hX}-1|] = \mathrm{E}[|e^{\mathrm{i}hX}-1|]\,,\end{aligned}$$

essendo $|e^{\mathrm{i}tX}|=1$. Dal momento che si hanno le disuguaglianze $|e^{\mathrm{i}hx}-1|\le 2$ e $|e^{\mathrm{i}hx}-1|\le|hx|$ per ogni $x\in\mathbb{R}$, per $a\in\,]0,\infty[$ risulta

$$\begin{aligned}\mathrm{E}[|e^{\mathrm{i}hX}-1|] &= \mathrm{E}[|e^{\mathrm{i}hX}-1|\,\mathbf{1}_{[0,a]}(|X|)] + \mathrm{E}[|e^{\mathrm{i}hX}-1|\,\mathbf{1}_{]a,\infty[}(|X|)] \\ &\le \mathrm{E}[|hX|\,\mathbf{1}_{[0,a]}(|X|)] + 2\,\mathrm{E}[\mathbf{1}_{]a,\infty[}(|X|)] = |h|\,\mathrm{E}[|X|\,\mathbf{1}_{[0,a]}(|X|)] + 2P(|X|>a) \\ &\le a|h| + 2P(|X|>a)\,.\end{aligned}$$

Dunque, fissato $\varepsilon>0$, se h e a sono scelti in modo tale che $a|h|<\frac{\varepsilon}{2}$ e $P(|X|>a)<\frac{\varepsilon}{4}$, si ha

$$|\varphi_X(t+h) - \varphi_X(t)| \leq a|h| + 2P(|X| > a) < \varepsilon \,,$$

da cui segue la continuità uniforme, dal momento che ε non dipende da t. □

Proposizione 7.1.5. *Se X è una v.a. definita sullo spazio probabilizzato $(\Omega, \mathcal{F}, P)$ con f.c. φ_X, allora la f.c. della v.a. trasformata $Y = a + bX$, dove $a, b \in \mathbb{R}$, è data da*

$$\varphi_Y(t) = e^{\mathrm{i}at}\, \varphi_X(bt) \,.$$

Dimostrazione. Si ha

$$\varphi_Y(t) = \mathrm{E}[e^{\mathrm{i}t(a+bX)}] = e^{\mathrm{i}at}\, \mathrm{E}[e^{\mathrm{i}btX}] = e^{\mathrm{i}at}\, \varphi_X(bt) \,,$$

che è quanto si voleva dimostrare. □

• **Esempio 7.1.6.** Si consideri la v.a. X distribuita con legge Normale $\mathcal{N}(\mu, \sigma^2)$. Tenendo presente l'Esempio 7.1.4, dal momento che $X = \mu + \sigma Z$, si ha

$$\varphi_X(t) = e^{\mathrm{i}\mu t}\, \varphi_Z(\sigma t) = e^{\mathrm{i}\mu t - \frac{1}{2}\sigma^2 t^2} \,.$$

In questo caso, al contrario della legge ridotta $\mathcal{N}(0,1)$, la f.c. φ_X non è puramente reale, dal momento che la v.a. X non è simmetrica rispetto all'origine. □

• **Esempio 7.1.7.** Si consideri la v.a. X distribuita con legge Gamma $\mathcal{G}(a, b, k)$. Tenendo presente l'Esempio 7.1.5, dal momento che $X = a + bZ$, si ha

$$\varphi_X(t) = e^{\mathrm{i}at}\, \varphi_Z(bt) = e^{\mathrm{i}at}\, (1 - \mathrm{i}bt)^{-k} \,.$$

In particolare, se la v.a. X si distribuisce con legge Chi-quadrato χ^2_n, dalla precedente espressione con $a = 0$, $b = 2$ e $k = \frac{n}{2}$ risulta

$$\varphi_X(t) = (1 - 2\mathrm{i}t)^{-\frac{1}{2}n} \,. \qquad \square$$

Le seguenti Proposizioni forniscono rispettivamente un'importante relazione fra la f.c. e i momenti di una v.a., e lo sviluppo in serie della f.c. basato sui momenti in un intorno di zero.

Proposizione 7.1.6. *Se la v.a. X definita sullo spazio probabilizzato $(\Omega, \mathcal{F}, P)$ possiede momento di ordine r finito, allora la relativa f.c. φ_X è derivabile r volte e*

$$\varphi_X^{(s)}(0) = \mathrm{i}^s \mathrm{E}[X^s]$$

per $s = 1, \ldots, r$.

Dimostrazione. Dalla definizione di derivata di funzione complessa si ha

$$\frac{d}{dt}\, e^{\mathrm{i}tx} = \frac{d}{dt} \cos(tx) + \mathrm{i}\, \frac{d}{dt} \sin(tx) = \mathrm{i}x\, e^{\mathrm{i}tx} \,.$$

Dunque, per ogni $t \in \mathbb{R}$, dal Teorema A.7 risulta che

$$\varphi_X^{(1)}(t) = \frac{d}{dt}\, \mathrm{E}[e^{\mathrm{i}tX}] = \mathrm{E}[\frac{d}{dt}\, e^{\mathrm{i}tX}] = \mathrm{E}[\mathrm{i}X\, e^{\mathrm{i}tX}] \,.$$

La funzione $\varphi_X^{(1)}$ è definita, dal momento che, tenendo presente le ipotesi fatte, si ha

$$|\mathrm{E}[\mathrm{i}X\, e^{\mathrm{i}tX}]| \leq \mathrm{E}[|\mathrm{i}X\, e^{\mathrm{i}tX}|] = \mathrm{E}[|\mathrm{i}| \cdot |X| \cdot |e^{\mathrm{i}tX}|] = \mathrm{E}[|X|] < \infty \,.$$

In particolare, risulta

$$\varphi_X^{(1)}(0) = \mathrm{i}\,\mathrm{E}[X]\,.$$

In maniera analoga, si ha

$$\varphi_X^{(2)}(t) = \frac{d}{dt}\,\mathrm{E}[\mathrm{i}X\,e^{\mathrm{i}tX}] = \mathrm{E}[\mathrm{i}^2X^2\,e^{\mathrm{i}tX}]\,,$$

che è definita, essendo per le ipotesi fatte

$$|\mathrm{E}[\mathrm{i}^2X^2\,e^{\mathrm{i}tX}]| \leq \mathrm{E}[|\mathrm{i}^2X^2\,e^{\mathrm{i}tX}|] = \mathrm{E}[X^2\,|e^{\mathrm{i}tX}|] = \mathrm{E}[X^2] < \infty\,.$$

Dunque, si ha

$$\varphi_X^{(2)}(0) = \mathrm{i}^2\,\mathrm{E}[X^2]\,.$$

Iterando il procedimento r volte si ottiene la tesi. □

• **Esempio 7.1.8.** Si consideri la v.a. Z distribuita con legge Normale ridotta $\mathcal{N}(0,1)$. La v.a. Z possiede momenti di ogni ordine, essendo $\mathrm{E}[Z^r] = 0$ se r è dispari, mentre risulta

$$\begin{aligned}\mathrm{E}[Z^r] &= \int_{-\infty}^{\infty} z^r\,\frac{1}{\sqrt{2\pi}}\,e^{-\frac{1}{2}z^2}\,dz = \frac{2}{\sqrt{2\pi}}\int_0^{\infty} z^r\,e^{-\frac{1}{2}z^2}\,dz \\ &= \frac{2^{\frac{1}{2}r}}{\Gamma(\frac{1}{2})}\int_0^{\infty} y^{\frac{1}{2}(r+1)-1}\,e^{-y}\,dy = \frac{2^{\frac{1}{2}r}\Gamma(\frac{r+1}{2})}{\Gamma(\frac{1}{2})}\end{aligned}$$

se r è pari. Tenendo presente la Formula di Duplicazione per la funzione Gamma di Eulero, la precedente espressione può essere anche riscritta come

$$\mathrm{E}[Z^r] = \frac{\Gamma(r)}{2^{\frac{1}{2}r-1}\Gamma(\frac{r}{2})} = \frac{r!}{2^{\frac{1}{2}r}(\frac{r}{2})!}\,.$$

In base alla Proposizione 7.1.5, la f.c. ammette dunque derivata di ogni ordine. In particolare, tenendo presente l'Esempio 7.1.4, si ha

$$\varphi_Z^{(1)}(t) = \frac{d}{dt}\,e^{-\frac{1}{2}t^2} = -\,t\varphi_Z(t)\,,$$

da cui $\varphi_Z^{(1)}(0) = 0$, mentre

$$\varphi_Z^{(2)}(t) = \frac{d^2}{dt^2}\,e^{-\frac{1}{2}t^2} = (t^2-1)\varphi_Z(t)\,,$$

da cui da cui $\varphi_Z^{(2)}(0) = -\,1$. Confermando i precedenti risultati, si ha dunque $\mathrm{E}[Z] = \mathrm{i}^{-1}\varphi_Z^{(1)}(0) = 0$, mentre $\mathrm{Var}[Z] = \mathrm{E}[Z^2] = \mathrm{i}^{-2}\varphi_Z^{(2)}(0) = 1$. □

• **Esempio 7.1.9.** Si consideri la v.a. Z con legge di Cauchy ridotta analizzata nell'Esempio 4.1.4. Dal medesimo Esempio e tenendo presente la Proposizione 4.3.3, la v.a. Z non possiede momento di ogni ordine. In effetti, dal momento che la v.a. Z è simmetrica rispetto all'origine, la f.c. è data da

$$\varphi_Z(t) = \mathrm{E}[\cos(tZ)] = \int_{-\infty}^{\infty} \cos(tz)\,\frac{1}{\pi(1+z^2)}\,dz = e^{-|t|}\,,$$

che appunto non risulta derivabile in $t = 0$. Per maggiori dettagli sulla precedente espressione, si veda l'Esercizio 7.3.2. □

Proposizione 7.1.7. *Se la v.a. X definita sullo spazio probabilizzato $(\Omega, \mathcal{F}, P)$ con f.c. φ_X possiede momento di ordine r finito, allora per $t \to 0$ si ha*

$$\varphi_X(t) = \sum_{s=0}^{r} \frac{(\mathrm{i}t)^s}{s!} \mathrm{E}[X^s] + o(|t|^r) \, .$$

Dimostrazione. Si veda Billingsley (1995, p.341).

• **Esempio 7.1.10.** Si consideri la v.a. Z con legge Normale ridotta $\mathcal{N}(0,1)$. Per $t \to 0$ si ha dunque

$$\varphi_Z(t) = 1 + \mathrm{i}t \, \mathrm{E}[Z] + \frac{(\mathrm{i}t)^2}{2} \mathrm{E}[Z^2] + o(t^2) = 1 - \frac{t^2}{2} + o(t^2) \, . \qquad \square$$

In alcune applicazioni della Teoria della Probabilità (ad esempio in molti problemi della statistica matematica) risulta piuttosto semplice ottenere la f.c. di alcune v.a. Il problema che sorge di conseguenza consiste nel risalire alla f.r. di queste v.a., se è nota la corrispondente f.c. I prossimi Teoremi consentono appunto di ottenere la f.r. di una v.a. (eventualmente la f.p. e la d.p. nel caso di v.a. discrete e di v.a. assolutamente continue) a partire dalla f.c.

Teorema 7.1.8. (Teorema di inversione di Lévy) *Si consideri la v.a. X definita sullo spazio probabilizzato $(\Omega, \mathcal{F}, P)$ con f.r. F_X e f.c. φ_X. Se $x, y \in \mathbb{R}$ sono punti di continuità di F_X tali che $x < y$, allora si ha*

$$F_X(y) - F_X(x) = \lim_{T \to \infty} \frac{1}{2\pi} \int_{-T}^{T} \frac{e^{\mathrm{i}tx} - e^{\mathrm{i}ty}}{\mathrm{i}t} \varphi_X(t) \, dt \, .$$

Dimostrazione. Si veda Billingsley (1995, p.346).

Teorema 7.1.9. *Si consideri la v.a. X definita sullo spazio probabilizzato $(\Omega, \mathcal{F}, P)$ con f.c. φ_X. Se $P(X = x) > 0$, allora si ha*

$$P(X = x) = \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} e^{-\mathrm{i}tx} \varphi_X(t) \, dt \, .$$

Dimostrazione. Si veda Billingsley (1995, p.354).

Teorema 7.1.10. *Si consideri la v.a. X definita sullo spazio probabilizzato $(\Omega, \mathcal{F}, P)$ con f.c. φ_X. Se $\int_{-\infty}^{\infty} |\varphi_X(t)| dt < \infty$, la v.a. X è assolutamente continua e ammette d.p.*

$$f_X(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-\mathrm{i}tx} \varphi_X(t) \, dt \, .$$

Dimostrazione. Si veda Billingsley (1995, p.347).

Si noti che dal Teorema 7.1.8 segue anche la proprietà di unicità della f.c., ovvero, date le v.a. X e Y, allora $X \stackrel{\mathcal{L}}{=} Y$ se e solo se $\varphi_X = \varphi_Y$. La f.c. deve appunto la sua denominazione al fatto che "caratterizza" in modo univoco la corrispondente legge.

• **Esempio 7.1.11.** Si consideri la v.a. Z con legge di Cauchy ridotta (si veda l'Esempio 4.1.4). Tenendo presente l'Esempio 7.1.9, si ha

$$\int_{-\infty}^{\infty} |\varphi_X(t)| \, dt = \int_{-\infty}^{\infty} e^{-|t|} \, dt = 2 < \infty$$

e dal momento che la f.c. φ_X è simmetrica rispetto all'origine, si ottiene

$$f_X(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-\mathrm{i}tx}\, e^{-|t|}\, dt = \frac{1}{2\pi}\int_{-\infty}^{\infty} \cos(tx)\, e^{-|t|}\, dt = \frac{1}{\pi}\int_{0}^{\infty} \cos(tx)\, e^{-t}\, dt = \frac{1}{\pi(1+x^2)}\,,$$

come ovviamente doveva risultare. □

7.2. Funzioni generatrici

Nel caso in cui si consideri una v.a. discreta X a valori su $\mathbb{N}$ (o su un suo sottoinsieme), è conveniente considerare la cosiddetta funzione generatrice di probabilità piuttosto che la f.c. Formalmente, si ha la seguente definizione.

Definizione 7.2.1. Data la v.a. discreta X definita sullo spazio probabilizzato $(\Omega, \mathcal{F}, P)$ a valori su $S \subseteq \mathbb{N}$ e con f.p. p_X, si dice *funzione generatrice di probabilità* (f.g.) (o semplicemente *funzione generatrice*) della v.a. X la funzione $G_X : [0,1] \to \mathbb{R}$ tale che

$$G_X(t) = \mathrm{E}[t^X] = \sum_{x=0}^{\infty} t^x p_X(x)\,,$$

con la convenzione che $0^0 = 1$. □

Nel linguaggio della Teoria della Misura la f.g. corrisponde alla funzione generatrice di Laplace. Risulta immediato verificare che per ciascun punto di $[0,1]$ la f.g. G_X coincide per definizione con la serie di potenze avente $(p_X(x))_{x\geq 0}$ come successione dei coefficienti. Inoltre, la serie di potenze ha raggio di convergenza almeno pari ad 1 dal momento che converge nel punto $t = 1$, essendo $G_X(1) = 1$. Inoltre, dal momento che $G_X(0) = p_X(0)$, la f.g. è definita per ogni valore dell'intervallo $[0,1]$ ed è crescente in questo intervallo essendo $\mathrm{E}[t^X] \leq \mathrm{E}[s^X]$ per ogni $t < s$.

• **Esempio 7.2.1.** Se si considera la v.a. X distribuita con legge Binomiale $\mathcal{B}(n,p)$, tenendo presente l'espressione del binomio di Newton, si ha

$$G_X(t) = \sum_{x=0}^{n} t^x \binom{n}{x} p^x q^{n-x} = \sum_{x=0}^{n} \binom{n}{x} (pt)^x q^{n-x} = (q+pt)^n\,.$$

Evidentemente, si ha $G_X(0) = p_X(0) = q^n$ e $G_X(1) = (q+p)^n = 1$, mentre G_X è una funzione crescente. □

• **Esempio 7.2.2.** Si consideri la v.a. X distribuita con legge di Poisson $\mathcal{P}(\lambda)$. Tenendo presente l'espressione della serie esponenziale, si ha

$$G_X(t) = \sum_{x=0}^{\infty} t^x\, e^{-\lambda}\, \frac{\lambda^x}{x!} = e^{-\lambda} \sum_{x=0}^{\infty} \frac{(\lambda t)^x}{x!} = e^{\lambda(t-1)}\,.$$ □

• **Esempio 7.2.3.** Si consideri la v.a. X con legge Binomiale Negativa $\mathcal{BN}(k,p)$. Tenendo presente l'espressione della serie binomiale negativa, si ha

$$G_X(t) = \sum_{x=0}^{\infty} t^x \binom{k+x-1}{k-1} p^k\, q^x = p^k \sum_{x=0}^{\infty} \binom{k+x-1}{x} (qt)^x = \left(\frac{p}{1-qt}\right)^k\,.$$ □

La seguente Proposizione mette in relazione la f.p. con la f.g. e giustifica in effetti la denominazione di G_X.

Proposizione 7.2.2. *Data la v.a. discreta X definita sullo spazio probabilizzato $(\Omega, \mathcal{F}, P)$ a valori su $S \subseteq \mathbb{N}$ con f.p. p_X e f.g. G_X, si ha*

$$p_X(x) = \frac{G_X^{(x)}(0)}{x!}$$

dove $x \in \mathbb{N}$.

Dimostrazione. Dal momento che una serie di potenze può essere derivata termine a termine per ogni valore nel raggio di convergenza, per $t \in [0, 1[$ si ha

$$G_X^{(1)}(t) = \sum_{n=1}^{\infty} n t^{n-1} p_X(n)$$

e

$$G_X^{(2)}(t) = \sum_{n=2}^{\infty} n(n-1) t^{n-2} p_X(n) \,,$$

mentre in generale si ha

$$G_X^{(x)}(t) = \sum_{n=x}^{\infty} n(n-1)\dots(n-x+1) t^{n-x} p_X(n) \,.$$

Dunque, si ottiene $G_X^{(1)}(0) = p_X(1)$ e $G_X^{(2)}(0) = 2p_X(2)$, mentre in generale risulta

$$G_X^{(x)}(0) = x! p_X(x) \,.$$

Tenendo presente che $G_X^{(0)}(0) = p_X(0)$ si ha dunque la tesi. □

Dalla precedente Proposizione segue anche la proprietà di unicità della f.g., ovvero, date le v.a. X e Y, allora si ha $X \stackrel{\mathcal{L}}{=} Y$ se e solo se $G_X = G_Y$. La seguente Proposizione fornisce la relazione fra la f.g. e i momenti di una v.a.

Proposizione 7.2.3. *Data la v.a. discreta X definita sullo spazio probabilizzato $(\Omega, \mathcal{F}, P)$ a valori su $S \subseteq \mathbb{N}$ con f.p. p_X e f.g. G_X, se la v.a. X possiede momento di ordine r finito, si ha*

$$G_X^{(s)}(1) = \mathrm{E}[X(X-1)\dots(X-s+1)]$$

per $s = 1, \dots, r$.

Dimostrazione. Tenendo presente la Proposizione 7.2.2, $G_X^{(s)}$ è positiva e crescente in $[0, 1[$ e quindi ammette un limite sinistro per $t \to 1$ (che può non essere finito). Si ha

$$\lim_{t \to 1^-} G_X^{(1)}(t) = \sum_{x=0}^{\infty} x p_X(x) = \mathrm{E}[X]$$

e

$$\lim_{t \to 1^-} G_X^{(2)}(t) = \sum_{x=0}^{\infty} x(x-1) p_X(x) = \mathrm{E}[X(X-1)] \,,$$

mentre in generale si ha

$$\lim_{t\to 1^-} G_X^{(s)}(t) = \sum_{x=0}^{\infty} x(x-1)\dots(x-s+1)p_X(x) = \mathrm{E}[X(X-1)\dots(X-s+1)]\,.$$

Dal momento che risulta $\mathrm{E}[X(X-1)\dots(X-s+1)] \le \mathrm{E}[X^s] < \infty$, in quanto per ipotesi $\mathrm{E}[X^s]$ è finito per $s = 1, \dots, r$, si ha quindi la tesi. □

In particolare dalla precedente Proposizione si ha che

$$\mathrm{E}[X] = G_X^{(1)}(1)$$

se $\mathrm{E}[X]$ è finito, mentre

$$\mathrm{Var}[X] = \mathrm{E}[X(X-1)] + \mathrm{E}[X] - \mathrm{E}[X]^2 = G_X^{(2)}(1) + G_X^{(1)}(1) - G_X^{(1)}(1)^2\,,$$

se $\mathrm{E}[X^2]$ è finito.

• **Esempio 7.2.4.** Se si considera la v.a. X distribuita con legge Binomiale $\mathcal{B}(n,p)$, si ha

$$G_X^{(1)}(t) = \frac{d}{dt}\,(q+pt)^n = np(q+pt)^{n-1}$$

e

$$G_X^{(2)}(t) = \frac{d^2}{dt^2}\,(q+pt)^n = n(n-1)p^2(q+pt)^{n-2}\,.$$

Dunque, risulta $\mathrm{E}[X] = np$ e $\mathrm{E}[X(X-1)] = n(n-1)p^2$, da cui è immediato verificare che $\mathrm{Var}[X] = npq$, un risultato noto dalla Sezione 6.1. □

• **Esempio 7.2.5.** Se si considera la v.a. X distribuita con legge di Poisson $\mathcal{P}(\lambda)$, si ha

$$G_X^{(1)}(t) = \frac{d}{dt}\,e^{\lambda(t-1)} = \lambda\,e^{\lambda(t-1)}$$

e

$$G_X^{(2)}(t) = \frac{d^2}{dt^2}\,e^{\lambda(t-1)} = \lambda^2\,e^{\lambda(t-1)}\,.$$

Dunque, risulta $\mathrm{E}[X] = \lambda$ e $\mathrm{E}[X(X-1)] = \lambda^2$, da cui $\mathrm{Var}[X] = \lambda$, un risultato verificato nella Sezione 6.2. □

• **Esempio 7.2.6.** Se si considera la v.a. X distribuita con legge Binomiale Negativa $\mathcal{BN}(k,p)$, si ha

$$G_X^{(1)}(t) = \frac{d}{dt}\left(\frac{p}{1-qt}\right)^k = \frac{kq}{p}\left(\frac{p}{1-qt}\right)^{k+1}$$

e

$$G_X^{(2)}(t) = \frac{d^2}{dt^2}\left(\frac{p}{1-qt}\right)^k = \frac{k(k+1)q^2}{p^2}\left(\frac{p}{1-qt}\right)^{k+2}\,.$$

Dunque, risulta $\mathrm{E}[X] = \frac{kq}{p}$ e $\mathrm{E}[X(X-1)] = \frac{k(k+1)q^2}{p^2}$, da cui è immediato verificare che $\mathrm{Var}[X] = \frac{kq}{p^2}$, un risultato ottenuto in precedenza nella Sezione 6.3. □

• **Esempio 7.2.7.** (Legge Zeta o di Zipf) Si consideri la v.a. discreta X con f.p. data da

$$p_X(x) = \frac{1}{\zeta(s)}\, x^{-s}\, \mathbf{1}_{\{1,2,\ldots\}}(x)\,,$$

dove $s \in]1,\infty[$, mentre $\zeta(s) = \sum_{x=1}^{\infty} x^{-s}$ rappresenta la funzione Zeta di Riemann. La legge associata a questa v.a. è detta di Zipf in onore del linguista e filologo statunitense George Kingsley Zipf (1902-1950), che si occupò delle applicazioni probabilistiche alle analisi dei testi. Ovviamente, la funzione Zeta prende nome dal matematico tedesco Georg Friedrich Bernhard Riemann (1826-1866). Dal momento che si ha $\zeta(2) = \frac{\pi^2}{6}$, per $s = 2$ la f.p. della v.a. X risulta

$$p_X(x) = \frac{6}{\pi^2}\, x^{-2}\, \mathbf{1}_{\{1,2,\ldots\}}(x)\,.$$

Per questo valore di s, la v.a. X non possiede momenti finiti di alcun ordine, essendo

$$\mathrm{E}[X] = \frac{6}{\pi^2} \sum_{x=1}^{\infty} x^{-1} = \infty\,.$$

In effetti, la f.g. risulta

$$G_X(t) = \frac{6}{\pi^2} \sum_{x=1}^{\infty} t^x x^{-2}\,,$$

da cui, tenendo presente l'espressione della serie logaritmica, si ha

$$G_X^{(1)}(t) = \frac{6}{\pi^2 t} \sum_{x=1}^{\infty} \frac{t^x}{x} = -\frac{6}{\pi^2}\, \frac{\log(1-t)}{t}\,,$$

e quindi

$$\lim_{t \to 1^-} G_X^{(1)}(t) = \infty\,.$$ □

7.3. Funzioni caratteristiche multivariate

Viene considerata di seguito l'estensione multivariata della f.c. quando si dispone di un v.v.a. Formalmente si ha la seguente definizione.

Definizione 7.3.1. Dato il v.v.a. $X = (X_1, \ldots, X_n)^{\mathrm{T}}$ definito sullo spazio probabilizzato $(\Omega, \mathcal{F}, P)$, si dice *funzione caratteristica multivariata* (f.c.m.) del v.v.a. X la funzione $\varphi_X : \mathbb{R}^n \to \mathbb{C}$

$$\varphi_X(t) = \mathrm{E}[\cos(t^{\mathrm{T}} X)] + \mathrm{i}\,\mathrm{E}[\sin(t^{\mathrm{T}} X)] = \mathrm{E}[e^{\mathrm{i}t^{\mathrm{T}} X}]\,,$$

dove $t = (t_1, \ldots, t_n)^{\mathrm{T}} \in \mathbb{R}^n$. □

Dalla Definizione 7.3.1 è ovvio che la f.c.m. possiede la seguente espressione alternativa

$$\varphi_X(t) = \varphi_X(t_1, \ldots, t_n) = \mathrm{E}\left[e^{\sum_{k=1}^{n} \mathrm{i}t_k X_k}\right] = \mathrm{E}\left[\prod_{k=1}^{n} e^{\mathrm{i}t_k X_k}\right]\,.$$

Risulta importante osservare che, se nella f.c.m. relativa al v.v.a. X si pone $t_l = 0$ per ogni $l \neq k = 1, \ldots, n$, si ha

$$\varphi_X(0,\dots,t_k,\dots,0) = \mathrm{E}[e^{\mathrm{i}t_k X_k}] = \varphi_{X_k}(t_k)\,,$$

ovvero si ottiene la f.c. della componente marginale X_k. In modo analogo, ponendo $t_l = 0$ per ogni $l \neq k, j = 1,\dots,n$, si ha

$$\varphi_X(0,\dots,t_k,\dots,t_j,\dots,0) = \mathrm{E}[e^{\mathrm{i}t_k X_k + \mathrm{i}t_j X_j}] = \varphi_{(X_k,X_j)}(t_k,t_j)\,,$$

che costituisce la f.c.m. del v.v.a. $(X_k, X_j)^{\mathrm{T}}$. In generale, azzerando opportunamente alcune componenti del vettore t, si ottengono le f.c.m. di tutte le possibili scelte di componenti marginali del v.v.a. X.

• **Esempio 7.3.1.** Si consideri il v.v.a. X distribuito con legge Normale Multivariata $\mathcal{N}_n(\mu,\Sigma)$. Tenendo presente quanto visto nella Sezione 6.11, in base al cambio di variabile $y = \Sigma^{-\frac{1}{2}}(x-\mu)$ con $x = \mu + \Sigma^{\frac{1}{2}} y$ e tenendo presente le regole di integrazione di una funzione complessa, la f.c.m. relativa al v.v.a. X è data da

$$\begin{aligned}\varphi_X(t) &= \int_{\mathbb{R}^n} e^{\mathrm{i}t^{\mathrm{T}}x} \det(2\pi\Sigma)^{-\frac{1}{2}}\, e^{-\frac{1}{2}(x-\mu)^{\mathrm{T}}\Sigma^{-1}(x-\mu)}\,dx = (2\pi)^{-\frac{1}{2}n} \int_{\mathbb{R}^n} e^{\mathrm{i}t^{\mathrm{T}}\mu + \mathrm{i}t^{\mathrm{T}}\Sigma^{\frac{1}{2}}y - \frac{1}{2}y^{\mathrm{T}}y}\,dy \\ &= (2\pi)^{-\frac{1}{2}n}\, e^{\mathrm{i}\mu^{\mathrm{T}}t - \frac{1}{2}t^{\mathrm{T}}\Sigma t} \int_{\mathbb{R}^n} e^{-\frac{1}{2}(y - \mathrm{i}\,\Sigma^{\frac{1}{2}}t)^{\mathrm{T}}(y - \mathrm{i}\,\Sigma^{\frac{1}{2}}t)}\,dy\,.\end{aligned}$$

Effettuando inoltre il cambio di variabile $z = y - \mathrm{i}\,\Sigma^{\frac{1}{2}}t$, con Jacobiano pari a 1, si ottiene

$$\varphi_X(t) = (2\pi)^{-\frac{1}{2}n}\, e^{\mathrm{i}\mu^{\mathrm{T}}t - \frac{1}{2}t^{\mathrm{T}}\Sigma t} \int_{\mathbb{R}^n} e^{-\frac{1}{2}z^{\mathrm{T}}z}\,dz = e^{\mathrm{i}\mu^{\mathrm{T}}t - \frac{1}{2}t^{\mathrm{T}}\Sigma t}\,,$$

dal momento che

$$(2\pi)^{-\frac{1}{2}n} \int_{\mathbb{R}^n} e^{-\frac{1}{2}z^{\mathrm{T}}z}\,dz = \left(\int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}}\, e^{-\frac{1}{2}u^2}\,du\right)^n = 1\,.$$

Si ha inoltre

$$\varphi_{X_k}(t_k) = \varphi_X(0,\dots,t_k,\dots,0) = e^{\mathrm{i}\mu_{X_k}t - \frac{1}{2}\sigma^2_{X_k}t_k^2}\,,$$

ovvero ogni componente marginale X_k si distribuisce con legge Normale $\mathcal{N}(\mu_{X_k}, \sigma^2_{X_k})$. In modo simile, si dimostra che ogni scelta di componenti marginali del v.v.a. X si distribuisce a sua volta con legge Normale Multivariata. □

La f.c.m. di un v.v.a. possiede proprietà analoghe rispetto alla f.c. di una v.a. Più esattamente, si hanno le seguenti Proposizioni.

Proposizione 7.3.2. *Se $X = (X_1,\dots,X_n)^{\mathrm{T}}$ è un v.v.a. definito sullo spazio probabilizzato $(\Omega,\mathcal{F},P)$ con f.c.m. φ_X, allora si ha $|\varphi_X(t)| \leq 1$ per ogni $t \in \mathbb{R}^n$ e in particolare $\varphi_X(0) = 1$, dove $0 = (0,\dots,0)^{\mathrm{T}}$ rappresenta l'origine di $\mathbb{R}^n$, mentre $\varphi_X(t)$ è uniformemente continua su $\mathbb{R}^n$.*

Dimostrazione. Risulta simile a quella delle Proposizioni 7.1.3. e 7.1.4. □

Proposizione 7.3.3. *Se $X = (X_1,\dots,X_n)^{\mathrm{T}}$ è un v.v.a. definito sullo spazio probabilizzato $(\Omega,\mathcal{F},P)$ con f.c.m. $\varphi_X(t_1,\dots,t_n)$, la f.c. della v.a. trasformata $Y = a + \sum_{k=1}^n b_k X_k$, dove $a \in \mathbb{R}$ e $b = (b_1,\dots,b_n)^{\mathrm{T}} \in \mathbb{R}^n$, risulta*

$$\varphi_Y(t) = e^{\mathrm{i}at}\,\varphi_X(tb_1,\dots,tb_n) = e^{\mathrm{i}at}\,\varphi_X(tb)\,,$$

dove $t \in \mathbb{R}$.

Dimostrazione. Si ha

$$\varphi_Y(t) = \mathrm{E}[e^{\mathrm{i}tY}] = \mathrm{E}\left[e^{\mathrm{i}at+\sum_{k=1}^n \mathrm{i}b_k t X_k}\right] = e^{\mathrm{i}at}\,\mathrm{E}\left[e^{\sum_{k=1}^n \mathrm{i}b_k t X_k}\right] = e^{\mathrm{i}at}\,\varphi_X(b_1 t, \ldots, b_n t) = e^{\mathrm{i}at}\,\varphi_X(tb)\,,$$

che è quanto si voleva dimostrare. □

Come conseguenza della precedente Proposizione si ottiene che la f.c. della v.a. $Y = \sum_{k=1}^n X_k$ è data da

$$\varphi_Y(t) = \varphi_X(t, \ldots, t)\,,$$

un risultato che riveste notevole importanza pratica nel determinare la legge della v.a. Y.

• **Esempio 7.3.2.** Si consideri il v.v.a. X distribuito con legge Normale Multivariata $\mathcal{N}_n(\mu, \Sigma)$. Tenendo presente l'Esempio 7.3.1, per la Proposizione 7.3.3 la f.c. della v.a. trasformata $Y = a + \sum_{k=1}^n b_k X_k$ è data da

$$\varphi_Y(t) = e^{\mathrm{i}at}\, e^{\mathrm{i}(tb)^{\mathrm{T}}\mu - \frac{1}{2}(tb)^{\mathrm{T}}\Sigma(tb)} = e^{\mathrm{i}(a+b^{\mathrm{T}}\mu)t - \frac{1}{2}(b^{\mathrm{T}}\Sigma b)t^2}\,,$$

che risulta essere la f.c. di una v.a. distribuita con legge Normale $\mathcal{N}(a + b^{\mathrm{T}}\mu, b^{\mathrm{T}}\Sigma b)$. Dunque, ogni combinazione lineare (e quindi anche la somma) delle componenti di un v.v.a. con legge Normale Multivariata si distribuisce a sua volta con legge Normale. □

Proposizione 7.3.4. *Dato il v.v.a. $X = (X_1, \ldots, X_n)^{\mathrm{T}}$ definito sullo spazio probabilizzato $(\Omega, \mathcal{F}, P)$ con f.c.m. $\varphi_X(t_1, \ldots, t_n)$, se $\mathrm{E}[X_1^{r_1} \ldots X_n^{r_n}]$ esiste finito, φ_X è derivabile r_1 volte rispetto a t_1, …, r_n volte rispetto a t_n e*

$$\left.\frac{\partial^s \varphi_X(t_1, \ldots, t_n)}{\partial t_1^{s_1} \ldots \partial t_n^{s_n}}\right|_{t_1,\ldots,t_n=0} = \mathrm{i}^s \mathrm{E}[X_1^{s_1} \ldots X_n^{s_n}]\,,$$

dove $s_k = 1, \ldots, r_k$, $k = 1, \ldots, n$, mentre $s = \sum_{k=1}^n s_k$.

Dimostrazione. Risulta analoga a quella della Proposizione 7.1.6. □

Teorema 7.3.5. (Teorema di inversione multivariato) *Dato il v.v.a. $X = (X_1, \ldots, X_n)^{\mathrm{T}}$ definito sullo spazio probabilizzato $(\Omega, \mathcal{F}, P)$ con f.r. F_X e f.c.m. φ_X, se $B =]x_1, y_1] \times \cdots \times]x_n, y_n]$, con $x_k < y_k$, $k = 1, \ldots, n$, è un insieme la cui frontiera è un insieme di continuità di F_X, allora*

$$P(X \in B) = \lim_{T\to\infty} \frac{1}{(2\pi)^n} \int_{[-T,T]^n} \left(\prod_{k=1}^n \frac{e^{\mathrm{i}t_k x_k} - e^{\mathrm{i}t_k y_k}}{\mathrm{i}t}\right) \varphi_X(t)\, dt\,.$$

Dimostrazione. Si veda Billingsley (1995, p.382). □

La seguente Proposizione fornisce un risultato di particolare rilevanza nel caso in cui le componenti del v.v.a. X sono indipendenti.

Proposizione 7.3.6. *Dato il v.v.a. $X = (X_1, \ldots, X_n)^{\mathrm{T}}$ definito sullo spazio probabilizzato $(\Omega, \mathcal{F}, P)$ con f.c.m. $\varphi_X(t_1, \ldots, t_n)$, allora si ha*

$$\varphi_X(t_1, \ldots, t_n) = \prod_{k=1}^n \varphi_{X_k}(t_k)\,,$$

se e solo se $X_1, \ldots, X_n$ sono indipendenti.

Dimostrazione. Se $X_1, \ldots, X_n$ sono v.a. indipendenti, allora dalla Proposizione 4.2.5 si ottiene

$$\varphi_X(t) = \mathrm{E}[\prod_{k=1}^{n} e^{\mathrm{i}t_k X_k}] = \prod_{k=1}^{n} \mathrm{E}[e^{\mathrm{i}t_k X_k}] = \prod_{k=1}^{n} \varphi_{X_k}(t_k) \, .$$

Inversamente, se $\varphi_X(t) = \prod_{k=1}^{n} \varphi_{X_k}(t_k)$ e ponendo $B =]x_1, y_1] \times \cdots \times]x_n, y_n]$, per il Teorema 7.3.5 si ottiene

$$P(X \in B) = \lim_{T \to \infty} \frac{1}{(2\pi)^n} \int_{[-T,T]^n} \left(\prod_{k=1}^{n} \frac{e^{\mathrm{i}tx_k} - e^{\mathrm{i}ty_k}}{\mathrm{i}t} \, \varphi_{X_k}(t_k) \right) dt_1 \ldots dt_n$$

e, tenendo presente i Teoremi A.7 e A.11, si ha

$$P(X \in B) = \prod_{k=1}^{n} \lim_{T \to \infty} \frac{1}{2\pi} \int_{[-T,T]} \frac{e^{\mathrm{i}tx_k} - e^{\mathrm{i}ty_k}}{\mathrm{i}t} \, \varphi_{X_k}(t_k) \, dt_k = \prod_{k=1}^{n} P(X_k \in]x_k, y_k]) \, ,$$

che dalla Definizione 3.6.1 implica l'indipendenza delle v.a. $X_1, \ldots, X_n$. □

Il Teorema 7.3.5 implica la proprietà di unicità della f.c.m., ovvero, dati i v.v.a. X e Y, allora si ha $X \stackrel{\mathcal{L}}{=} Y$ se e solo se $\varphi_X = \varphi_Y$. Inoltre, tenendo presente la Proposizione 7.3.3, se $X_1, \ldots, X_n$ sono v.a. indipendenti, la f.c. della v.a. trasformata $Y = a + \sum_{k=1}^{n} b_k X_k$, dove $a \in \mathbb{R}$ e $b = (b_1, \ldots, b_n)^{\mathrm{T}} \in \mathbb{R}^n$, risulta

$$\varphi_Y(t) = e^{\mathrm{i}at} \, \varphi_X(b_1 t, \ldots, b_n t) = e^{\mathrm{i}at} \prod_{k=1}^{n} \varphi_{X_k}(b_k t) \, .$$

In particolare, se $Y = \sum_{k=1}^{n} X_k$ è la somma di n v.a. indipendenti, la relativa f.c. risulta

$$\varphi_Y(t) = \prod_{k=1}^{n} \varphi_{X_k}(t) \, .$$

Dunque, quando le v.a. $X_1, \ldots, X_n$ si distribuiscono con la stessa legge che coincide con quella della v.a. X, ovvero quando $\varphi_{X_k}(t) = \varphi_X(t)$ per ogni $k = 1, \ldots, n$, si ottiene

$$\varphi_Y(t) = \varphi_X(t)^n \, .$$

In generale, si può dimostrare in modo semplice ma laborioso che se $X_1, \ldots, X_n$ sono v.v.a. indipendenti con k componenti marginali, la f.c.m. del v.v.a. trasformato $Y = \sum_{j=1}^{n} X_j$ risulta

$$\varphi_Y(t_1, \ldots, t_k) = \prod_{j=1}^{n} \varphi_{X_j}(t_1, \ldots, t_k) \, .$$

• **Esempio 7.3.3.** Si consideri la v.a. Z distribuita con legge Esponenziale ridotta $\mathcal{G}(0, 1, 1)$. Dunque, dall'Esempio 7.1.5 la f.c. della v.a. Z è data da

$$\varphi_Z(t) = (1 - \mathrm{i}t)^{-1} \, .$$

Di conseguenza, se $Z_1, \ldots, Z_k$ sono v.a. indipendenti con la stessa legge della v.a. Z, la v.a. $Y = \sum_{j=1}^{k} X_j$ possiede f.c. data da

$$\varphi_Y(t) = ((1 - \mathrm{i}t)^{-1})^k = (1 - \mathrm{i}t)^{-k} \, .$$

Dunque, tenendo ancora presente l'Esempio 7.1.5, la v.a. Y si distribuisce con legge Gamma ridotta $\mathcal{G}(0, 1, k)$. □

• **Esempio 7.3.4.** Si consideri la v.a. X distribuita con legge Chi-quadrato χ^2_1. Dunque, dall'Esempio 7.1.7 la f.c. di X è data da

$$\varphi_X(t) = (1 - 2\mathrm{i}t)^{-\frac{1}{2}} .$$

Di conseguenza, se $X_1, \ldots, X_n$ sono v.a. indipendenti con la stessa legge di X, la v.a. $Y = \sum_{k=1}^n X_k$ possiede f.c. data da

$$\varphi_Y(t) = ((1 - 2\mathrm{i}t)^{-\frac{1}{2}})^n = (1 - 2\mathrm{i}t)^{-\frac{1}{2}n} .$$

Dunque, tenendo presente l'Esempio 7.1.7, la v.a. Y si distribuisce con legge Chi-quadrato χ^2_n. □

• **Esempio 7.3.5.** Si consideri le v.a. indipendenti $X_1, \ldots, X_n$, rispettivamente distribuite con legge Normale $\mathcal{N}(\mu_{X_k}, \sigma^2_{X_k})$ per $k = 1, \ldots, n$. Tenendo presente l'Esempio 7.1.6, la v.a. $Y = \sum_{k=1}^n X_k$ possiede f.c. data da

$$\varphi_Y(t) = \prod_{k=1}^n e^{\mathrm{i}\mu_{X_k}t - \frac{1}{2}\sigma^2_{X_k}t^2} = e^{\mathrm{i}\mu_Y t - \frac{1}{2}\sigma^2_Y t^2} ,$$

dove $\mu_Y = \sum_{k=1}^n \mu_{X_k}$ e $\sigma^2_Y = \sum_{k=1}^n \sigma^2_{X_k}$. Dunque, la v.a. Y si distribuisce con legge Normale $\mathcal{N}(\mu_Y, \sigma^2_Y)$, un risultato che conferma un caso particolare di quanto ottenuto nell'Esempio 7.3.2. □

• **Esempio 7.3.6.** Si consideri il v.v.a. X con legge Normale Multivariata $\mathcal{N}_k(\mu, \Sigma)$. Se $X_1, \ldots, X_n$ sono v.v.a. indipendenti con la stessa legge del v.v.a. X, sulla base dei risultati ottenuti nell'Esempio 7.3.1, il v.v.a. $Y = \sum_{j=1}^n X_j$ possiede f.c.m. data da

$$\varphi_Y(t) = (e^{\mathrm{i}\mu^{\mathrm{T}}t - \frac{1}{2}t^{\mathrm{T}}\Sigma t})^n = e^{\mathrm{i}(n\mu)^{\mathrm{T}}t - \frac{1}{2}t^{\mathrm{T}}(n\Sigma)t} .$$

Dunque, tenendo ancora presente l'Esempio 7.3.1, il v.v.a. Y si distribuisce con legge Normale Multivariata $\mathcal{N}_k(n\mu, n\Sigma)$. □

7.4. Funzioni generatrici multivariate

Quando si dispone di un v.v.a. discreto con n componenti marginali e a valori su $\mathbb{N}^n$ (o su un suo sottoinsieme) si può estendere il concetto di funzione generatrice. Più esattamente, si ha la seguente definizione.

Definizione 7.4.1. Dato il v.v.a. discreto $X = (X_1, \ldots, X_n)^{\mathrm{T}}$ definito sullo spazio probabilizzato $(\Omega, \mathcal{F}, P)$ a valori su $S \subseteq \mathbb{N}^n$ e con f.p.c. p_X, si dice *funzione generatrice multivariata di probabilità* (f.g.m.) (o semplicemente *funzione generatrice multivariata*) del v.v.a. X la funzione $G_X : [0,1]^n \to \mathbb{R}$ tale che

$$G_X(t_1, \ldots, t_n) = \mathrm{E}[\prod_{k=1}^n t_k^{X_k}] = \sum_{(x_1,\ldots,x_n)\in S} \left(\prod_{k=1}^n t_k^{x_k} \right) p_X(x_1, \ldots, x_n) ,$$

con la convenzione che $0^0 = 1$. □

Se nella f.g.m. relativa al v.v.a. X si pone $t_l = 1$ per ogni $l \neq k = 1, \ldots, n$, si ottiene

$$G_X(1, \ldots, t_k, \ldots, 1) = \mathrm{E}[t_k^{X_k}] = G_{X_k}(t_k) ,$$

che costituisce la f.g. della componente marginale X_k. In modo analogo, ponendo $t_l = 1$ per ogni $l \neq k, j = 1, \ldots, n$, si ottiene

$$G_X(1, \ldots, t_k, \ldots, t_j, \ldots, 1) = \mathrm{E}[t_k^{X_k} t_j^{X_j}] = G_{(X_k, X_j)}(t_k, t_j) \,,$$

che costituisce la f.g.m. del v.v.a. (X_k, X_j). In generale, ponendo pari all'unità in modo opportuno alcune componenti del vettore $(t_1, \ldots, t_n)$, si ottengono le f.g.m. di tutte le possibili scelte di componenti marginali del v.v.a. X.

• **Esempio 7.4.1.** Si consideri il v.v.a. X con legge Multinomiale $\mathcal{M}(n, p)$. Dunque, per il Teorema Multinomiale si ha

$$\begin{aligned} G_X(t_1, \ldots, t_k) &= \sum_{(x_1, \ldots, x_k) \in S} \left(\prod_{j=1}^{k} t_j^{x_j} \right) \binom{n}{x_1 \ldots x_k} \prod_{j=1}^{k} p_j^{x_j} \\ &= \sum_{(x_1, \ldots, x_k) \in S} \binom{n}{x_1 \ldots x_k} \prod_{j=1}^{k} (p_j t_j)^{x_j} = \left(\sum_{j=1}^{k} p_j t_j \right)^n \,, \end{aligned}$$

dove $S \subset \mathbb{N}^k$ è definito nella Sezione 6.5. Inoltre, risulta

$$G_{X_j}(t_j) = G_X(1, \ldots, t_j, \ldots, 1) = (q_j + p_j t_j)^n \,,$$

dove $q_j = 1 - p_j = \sum_{l \neq j = 1}^{k} p_l$, ovvero ogni componente marginale X_j si distribuisce con legge Binomiale $\mathcal{B}(n, p_j)$. In modo analogo, si dimostra che ogni scelta di componenti marginali del v.v.a. X si distribuisce ancora con legge Multinomiale. □

La f.g.m. di un v.v.a. possiede proprietà analoghe alla f.g. di una v.a. In particolare, la f.g.m. è definita per ogni $t \in [0, 1]^n$ e risulta $G_X(1, \ldots, 1) = 1$, mentre $G_X(0, \ldots, 0) = p_X(0, \ldots, 0)$. Si ha inoltre la seguente Proposizione.

Proposizione 7.4.2. *Dato il v.v.a. discreto $X = (X_1, \ldots, X_n)^{\mathrm{T}}$ definito sullo spazio probabilizzato $(\Omega, \mathcal{F}, P)$ a valori su $S \subseteq \mathbb{N}^n$ con f.p.c. p_X e f.g.m. G_X, si ha*

$$p_X(x_1, \ldots, x_n) = \frac{1}{x_1! \ldots x_n!} \left. \frac{\partial^{x_1 + \ldots + x_n} G_X(t_1, \ldots, t_n)}{\partial t_1^{x_1} \ldots \partial t_n^{x_n}} \right|_{t_1, \ldots, t_n = 0}$$

dove $(x_1, \ldots, x_n) \in \mathbb{N}^n$.

Dimostrazione. Risulta simile a quella della Proposizione 7.2.2. □

Dalla precedente Proposizione segue anche la proprietà di unicità della f.g.m., ovvero, dati i v.v.a. X e Y, allora si ha $X \overset{\mathcal{L}}{=} Y$ se e solo se $G_X = G_Y$.

Proposizione 7.4.3. *Dato il v.v.a. discreto $X = (X_1, \ldots, X_n)^{\mathrm{T}}$ definito sullo spazio probabilizzato $(\Omega, \mathcal{F}, P)$ a valori su $S \subseteq \mathbb{N}^n$ con f.g.m. G_X, se $\mathrm{E}[X_1^{r_1} \ldots X_n^{r_n}]$ esiste finito, G_X è derivabile r_1 volte rispetto a t_1, …, r_n volte rispetto a t_n e*

$$\left. \frac{\partial^s G_X(t_1, \ldots, t_n)}{\partial t_1^{s_1} \ldots \partial t_n^{s_n}} \right|_{t_1, \ldots, t_n = 1} = \mathrm{E}\left[\prod_{k=1}^{n} X_k (X_k - 1) \ldots (X_k - s_k + 1) \right] \,,$$

dove $s_k = 1, \ldots, r_k$ e $k = 1, \ldots, n$, mentre $s = \sum_{k=1}^{n} s_k$.

Dimostrazione. Risulta simile a quella della Proposizione 7.2.3. □

• **Esempio 7.4.2.** Si consideri il v.v.a. X distribuito con legge Multinomiale $\mathcal{M}(n, p)$. Dunque, dalla Proposizione 7.4.3 si ha

$$\mathrm{E}[X_j] = \frac{\partial}{\partial t_j}\left(\sum_{l=1}^{k} p_l t_l\right)^n \Bigg|_{t_1,\ldots,t_k=1} = np_j \,,$$

per cui il vettore delle medie del v.v.a. X è dato da $\mu = np$. Inoltre, si ha

$$\mathrm{E}[X_j(X_j-1)] = \frac{\partial^2}{\partial t_j^2}\left(\sum_{l=1}^{k} p_l t_l\right)^n \Bigg|_{t_1,\ldots,t_k=1} = n(n-1)p_j^2$$

e

$$\mathrm{E}[X_i X_j] = \frac{\partial^2}{\partial t_i \partial t_j}\left(\sum_{l=1}^{k} p_l t_l\right)^n \Bigg|_{t_1,\ldots,t_k=1} = n(n-1)p_i p_j \,.$$

Dunque, risulta $\mathrm{Var}[X_j] = np_j(1-p_j)$ e $\mathrm{Cov}[X_i, X_j] = -np_i p_j$, e quindi la matrice di varianza-covarianza è data da $\Sigma = n(\mathrm{diag}(p) - pp^{\mathrm{T}})$. □

Proposizione 7.4.4. *Dato il v.v.a. discreto $X = (X_1, \ldots, X_n)^{\mathrm{T}}$ definito sullo spazio probabilizzato $(\Omega, \mathcal{F}, P)$ a valori su $S \subseteq \mathbb{N}^n$ con f.g.m. G_X, la f.g.m. della v.a. trasformata $Y = \sum_{k=1}^n X_k$ risulta*

$$G_Y(t) = G_X(t, \ldots, t) \,,$$

dove $t \in \mathbb{R}$.

Dimostrazione. Si ha

$$G_Y(t) = \mathrm{E}[t^{\sum_{k=1}^n X_k}] = \mathrm{E}[\prod_{k=1}^{n} t^{X_k}] = G_X(t, \ldots, t) \,,$$

che è quanto si voleva dimostrare. □

La seguente Proposizione fornisce l'espressione della f.g.m. nel caso in cui le componenti del v.v.a. X risultano indipendenti.

Proposizione 7.4.5. *Dato il v.v.a. discreto $X = (X_1, \ldots, X_n)^{\mathrm{T}}$ definito sullo spazio probabilizzato $(\Omega, \mathcal{F}, P)$ a valori su $S \subseteq \mathbb{N}^n$ con f.g.m. G_X, allora si ha*

$$G_X(t_1, \ldots, t_n) = \prod_{k=1}^{n} G_{X_k}(t_k) \,,$$

se e solo se $X_1, \ldots, X_n$ sono indipendenti.

Dimostrazione. Se $X_1, \ldots, X_n$ sono v.a. indipendenti, allora dalla Proposizione 4.2.5 si ottiene

$$G_X(t_1, \ldots, t_n) = \mathrm{E}[\prod_{k=1}^{n} t_k^{X_k}] = \prod_{k=1}^{n} \mathrm{E}[t_k^{X_k}] = \prod_{k=1}^{n} G_{X_k}(t_k) \,.$$

Inversamente, se $G_X(t_1, \ldots, t_n) = \prod_{k=1}^n G_{X_k}(t_k)$, dalla Proposizione 7.4.2 si ha

$$p_X(x_1,\dots,x_n) = \frac{1}{x_1!\dots x_n!}\left.\frac{\partial^{x_1+\dots+x_n}\prod_{k=1}^n G_{X_k}(t_k)}{\partial t_1^{x_1}\dots\partial t_n^{x_n}}\right|_{t_1,\dots,t_n=0}$$
$$= \prod_{k=1}^n \frac{1}{x_k!}\left.\frac{\partial^{x_k} G_{X_k}(t_k)}{\partial t_k^{x_k}}\right|_{t_1,\dots,t_n=0} = \prod_{k=1}^n p_{X_k}(x_k)\,,$$

e il Teorema 3.6.1 implica l'indipendenza delle v.a. $X_1,\dots,X_n$. □

Tenendo presente la Proposizione 7.4.4, se $X_1,\dots,X_n$ sono v.a. indipendenti, la f.g. della v.a. trasformata $Y = \sum_{k=1}^n X_k$ risulta

$$G_Y(t) = \prod_{k=1}^n G_{X_k}(t)\,.$$

Dunque, quando le v.a. X_k possiedono la stessa legge che coincide con quella di una data v.a. X, ovvero quando $G_{X_k}(t) = G_X(t)$ per ogni $k = 1,\dots,n$, si ottiene

$$G_Y(t) = G_X(t)^n\,.$$

In generale, si può dimostrare in modo semplice ma laborioso che se $X_1,\dots,X_n$ sono v.v.a. indipendenti con k componenti marginali, la f.g.m. del v.v.a. trasformato $Y = \sum_{j=1}^n X_j$ risulta

$$G_Y(t_1,\dots,t_k) = \prod_{j=1}^n G_{X_j}(t_1,\dots,t_k)\,.$$

• **Esempio 7.4.3.** Si consideri la v.a. X distribuita con legge di Bernoulli $\mathcal{B}(1,p)$. Dunque, dall'Esempio 7.2.1 la f.g. di X è data da

$$G_X(t) = q + pt\,.$$

Di conseguenza, se $X_1,\dots,X_n$ sono v.a. indipendenti con la stessa legge della v.a. X, la v.a. $Y = \sum_{k=1}^n X_k$ possiede f.g. data da

$$G_Y(t) = (q+pt)^n\,.$$

Dunque, tenendo ancora presente l'Esempio 7.2.1, la v.a. Y si distribuisce con legge Binomiale $\mathcal{B}(n,p)$. □

• **Esempio 7.4.4.** Si consideri la v.a. X distribuita con legge Binomiale Negativa $\mathcal{BN}(1,p)$, ovvero con legge Geometrica. Dunque, dall'Esempio 7.2.3 la f.g. della v.a. X è data da

$$G_X(t) = \frac{p}{1-qt}\,.$$

Di conseguenza, se $X_1,\dots,X_k$ sono v.a. indipendenti con la stessa legge della v.a. X, la v.a. $Y = \sum_{k=1}^k X_k$ possiede f.g. data da

$$G_Y(t) = \left(\frac{p}{1-qt}\right)^k\,.$$

Dunque, tenendo ancora presente l'Esempio 7.2.3, la v.a. Y si distribuisce con legge Binomiale Negativa $\mathcal{BN}(k,p)$. □

• **Esempio 7.4.5.** Si consideri il v.v.a. X distribuito con legge Multinomiale $\mathcal{M}(1, p)$. Dunque, dall'Esempio 7.4.1 la f.g.m. del v.v.a. X è data da

$$G_X(t_1, \ldots, t_k) = \sum_{j=1}^{k} p_j t_j \, .$$

Di conseguenza, se $X_1, \ldots, X_n$ sono v.v.a. indipendenti con la stessa legge del v.v.a. X, il v.v.a. $Y = \sum_{j=1}^{k} X_j$ possiede f.g.m. data da

$$G_Y(t_1, \ldots, t_k) = \left(\sum_{j=1}^{k} p_j t_j \right)^n \, .$$

Dunque, tenendo ancora presente l'Esempio 7.4.1, il v.v.a. Y si distribuisce con legge Multinomiale $\mathcal{M}(n, p)$. □

7.5. Leggi additive e infinitamente divisibili

In questa sezione vengono considerate due importanti famiglie di leggi che sono caratterizzate da particolari comportamenti rispetto alla somma di v.a. indipendenti. In particolare, si ha la seguente definizione.

Definizione 7.5.1. Sia data la v.a. X definita sullo spazio probabilizzato $(\Omega, \mathcal{F}, P)$ con f.r. $F_{X,\theta}$ che dipende da un vettore di parametri $\theta \in \Theta \subseteq \mathbb{R}^m$. La v.a. X e la relativa legge si dicono *additive rispetto a* θ se, date le v.a. indipendenti X_1 e X_2 con rispettive f.r. F_{X_1,θ_1} e F_{X_2,θ_2}, la f.r. della v.a. $(X_1 + X_2)$ risulta essere $F_{X_1+X_2,\theta_1+\theta_2}$ per ogni $\theta_1, \theta_2 \in \Theta$. □

Tenendo presente la proprietà di unicità della f.c., la Definizione 7.5.1 può essere anche data in termini della f.c. $\varphi_{X,\theta}$ della v.a. X. Evidentemente, se la v.a. X e la rispettiva legge risultano additive rispetto a θ, si verifica che

$$\varphi_{X_1,\theta_1}(t)\varphi_{X_2,\theta_2}(t) = \varphi_{X_1+X_2,\theta_1+\theta_2}(t) \, ,$$

per ogni $\theta_1, \theta_2 \in \Theta$. Ovviamente, nel caso di una v.a. discreta X a valori su $S \subseteq \mathbb{N}$, una analoga definizione può essere data in termini della corrispondente f.g. $G_{X,\theta}$, ovvero si ha

$$G_{X_1,\theta_1}(t)G_{X_2,\theta_2}(t) = G_{X_1+X_2,\theta_1+\theta_2}(t) \, ,$$

per ogni $\theta_1, \theta_2 \in \Theta$. Si osservi che il vettore di parametri θ può non comprendere l'intero insieme di parametri che caratterizzano la legge in questione, ma soltanto il sottoinsieme di parametri rispetto ai quali si vuole verificare l'additività. In questo caso, i rimanenti parametri devono invece essere costanti.

• **Esempio 7.5.1.** Si consideri la v.a. X distribuita con legge Normale $\mathcal{N}(\mu, \sigma^2)$. In questo caso, si pone $\theta = (\mu, \sigma^2)^{\mathrm{T}}$ con $\Theta = \mathbb{R} \times]0, \infty[$. La legge risulta additiva rispetto al vettore di parametri θ, in quanto dall'Esempio 7.3.5 si ha

$$\varphi_{X_1,\theta_1}(t)\varphi_{X_2,\theta_2}(t) = e^{\mathrm{i}\mu_{X_1}t - \frac{1}{2}\sigma^2_{X_1}t^2} \, e^{\mathrm{i}\mu_{X_2}t - \frac{1}{2}\sigma^2_{X_2}t^2} = e^{\mathrm{i}(\mu_{X_1}+\mu_{X_2})t - \frac{1}{2}(\sigma^2_{X_1}+\sigma^2_{X_2})t^2} = \varphi_{X_1+X_2,\theta_1+\theta_2}(t) \, ,$$

ovvero, in altri termini, la legge della somma di due v.a. indipendenti, rispettivamente con legge Normale $\mathcal{N}(\mu_{X_1}, \sigma^2_{X_1})$ e $\mathcal{N}(\mu_{X_2}, \sigma^2_{X_2})$, è Normale $\mathcal{N}(\mu_{X_1} + \mu_{X_2}, \sigma^2_{X_1} + \sigma^2_{X_2})$. □

• **Esempio 7.5.2.** Si consideri la v.a. X distribuita con legge Gamma $\mathcal{G}(0, b, k)$. In questo caso, si pone $\theta = k$ con $\Theta =]0, \infty[$, ovvero si vuole verificare l'additività rispetto al parametro k e non al parametro b. La legge è additiva rispetto a θ, in quanto tenendo presente l'Esempio 7.1.7 si ha

$$\varphi_{X_1,\theta_1}(t)\varphi_{X_2,\theta_2}(t) = (1 - \mathrm{i}bt)^{-k_1}(1 - \mathrm{i}bt)^{-k_2} = (1 - \mathrm{i}bt)^{-(k_1+k_2)} = \varphi_{X_1+X_2,\theta_1+\theta_2}(t) \, .$$

Dunque, la legge della somma di due v.a. indipendenti, rispettivamente con legge Gamma $\mathcal{G}(0, b, k_1)$ e $\mathcal{G}(0, b, k_2)$, è Gamma $\mathcal{G}(0, b, k_1 + k_2)$. In particolare, se la legge della v.a. X è Chi-quadrato χ^2_n, dalla precedente espressione con $b = 2$ e $k = \frac{n}{2}$, la legge risulta additiva rispetto al parametro n. Dunque, la legge della somma di due v.a. indipendenti, rispettivamente con legge Chi-quadrato $\chi^2_{n_1}$ e $\chi^2_{n_2}$, è Chi-quadrato $\chi^2_{n_1+n_2}$. Si osservi infine che l'additività non vale per il parametro di scala b. In effetti, se si considera il vettore di parametri $\theta = (b, k)^{\mathrm{T}}$ con $\Theta =]0, \infty[\times]0, \infty[$, si ha

$$\varphi_{X_1,\theta_1}(t)\varphi_{X_2,\theta_2}(t) = (1 - \mathrm{i}b_1t)^{-k_1}(1 - \mathrm{i}b_2t)^{-k_2} \neq \varphi_{X_1+X_2,\theta_1+\theta_2}(t) = (1 - \mathrm{i}(b_1 + b_2)t)^{-(k_1+k_2)} \, .$$

In altre parole, la legge della somma di due v.a. indipendenti, distribuite rispettivamente con legge Gamma $\mathcal{G}(0, b_1, k_1)$ e $\mathcal{G}(0, b_2, k_2)$, non è Gamma $\mathcal{G}(0, b_1 + b_2, k_1 + k_2)$. □

• **Esempio 7.5.3.** Si consideri la v.a. X distribuita con legge Binomiale $\mathcal{B}(n, p)$. In questo caso, si pone $\theta = n$ con $\Theta = \{1, 2, \dots\}$, ovvero si vuole verificare l'additività rispetto al parametro n e non al parametro p. La legge risulta additiva rispetto a n, in quanto tenendo presente l'Esempio 7.2.1 si ha

$$G_{X_1,\theta_1}(t)G_{X_2,\theta_2}(t) = (q + pt)^{n_1}(q + pt)^{n_2} = (q + pt)^{n_1+n_2} = G_{X_1+X_2,\theta_1+\theta_2}(t) \, .$$

Quindi, la legge della somma di due v.a. indipendenti, distribuite rispettivamente con legge Binomiale $\mathcal{B}(n_1, p)$ e $\mathcal{B}(n_2, p)$, è Binomiale $\mathcal{B}(n_1 + n_2, p)$. Risulta immediato verificare che l'additività non vale per il parametro p. Dunque, se si considera il vettore di parametri $\theta = (n, p)^{\mathrm{T}}$ con $\Theta = \mathbb{Z}^+ \times]0, 1[$, la legge non risulta additiva rispetto a θ. □

• **Esempio 7.5.4.** Si consideri la v.a. X distribuita con legge di Poisson $\mathcal{P}(\lambda)$. Si pone $\theta = \lambda$ con $\Theta =]0, \infty[$. La legge risulta additiva rispetto al parametro λ, in quanto tenendo presente l'Esempio 7.2.2 si ha

$$G_{X_1,\theta_1}(t)G_{X_2,\theta_2}(t) = e^{\lambda_1(t-1)}\, e^{\lambda_2(t-1)} = e^{(\lambda_1+\lambda_2)(t-1)} = G_{X_1+X_2,\theta_1+\theta_2}(t) \, .$$

Dunque, la legge della somma di due v.a. indipendenti, distribuite rispettivamente con legge di Poisson $\mathcal{P}(\lambda_1)$ e $\mathcal{P}(\lambda_2)$, è Poisson $\mathcal{P}(\lambda_1 + \lambda_2)$. □

• **Esempio 7.5.5.** Si consideri la v.a. X distribuita con legge Binomiale Negativa $\mathcal{BN}(k, p)$. Si pone $\theta = k$ con $\Theta = \mathbb{Z}^+$, ovvero si vuole verificare l'additività rispetto al parametro k e non al parametro p. La legge risulta additiva rispetto a θ, in quanto tenendo presente l'Esempio 7.2.3 si ha

$$G_{X_1,\theta_1}(t)G_{X_2,\theta_2}(t) = \left(\frac{p}{1 - qt}\right)^{k_1}\left(\frac{p}{1 - qt}\right)^{k_2} = \left(\frac{p}{1 - qt}\right)^{k_1+k_2} = G_{X_1+X_2,\theta_1+\theta_2}(t) \, ,$$

ovvero si ha che la legge della somma di due v.a. indipendenti, distribuite rispettivamente con legge Binomiale Negativa $\mathcal{BN}(k_1, p)$ e $\mathcal{BN}(k_2, p)$, è Binomiale Negativa $\mathcal{BN}(k_1 + k_2, p)$. In modo analogo alla legge Binomiale, risulta inoltre immediato verificare che l'additività non vale per il parametro p. Dunque, se si considera il vettore di parametri $\theta = (k, p)^{\mathrm{T}}$ con $\Theta = \mathbb{Z}^+ \times]0, 1[$, la legge non è additiva rispetto a θ. □

Definizione 7.5.2. Si consideri la v.a. X definita sullo spazio probabilizzato $(\Omega, \mathcal{F}, P)$. La v.a. X e la rispettiva legge sono dette *infinitamente divisibili* se per ogni $n = 1, 2, \dots$ si ha $X \stackrel{\mathcal{L}}{=} \sum_{k=1}^{n} X_k$, dove $(X_1, \dots, X_n)^{\mathrm{T}}$ è un v.v.a. indipendenti e ugualmente distribuite. □

Tenendo presente la proprietà di unicità della f.c., la Definizione 7.5.2 può essere data in modo alternativo utilizzando la f.c. In questo caso, se la v.a. X possiede f.c. φ_X, mentre φ_{X_n} è la f.c. comune alle componenti marginali del v.v.a. $(X_1, \ldots, X_n)^{\mathrm{T}}$, allora la v.a. X è infinitamente divisibile se per ogni $n = 1, 2, \ldots$ si ha

$$\varphi_X(t) = \varphi_{X_n}(t)^n \, .$$

Nel caso in cui la v.a. X è discreta a valori su $S \subseteq \mathbb{N}$, una analoga definizione può essere data in termini della f.g., nel senso che la precedente condizione è equivalente a

$$G_X(t) = G_{X_n}(t)^n \, .$$

Per una v.a. infinitamente divisibile vale inoltre la seguente Proposizione.

Proposizione 7.5.3. *Data la v.a. X definita sullo spazio probabilizzato $(\Omega, \mathcal{F}, P)$, se la v.a. X è infinitamente divisibile, anche la trasformata $Y = a + bX$, dove $a, b \in \mathbb{R}$, è infinitamente divisibile.*

Dimostrazione. Tenendo presente la Proposizione 7.3.3, si ha

$$\varphi_Y(t) = e^{\mathrm{i}at} \, \varphi_X(bt) = e^{\mathrm{i}at} \, \varphi_{X_n}(bt)^n = (e^{\frac{\mathrm{i}at}{n}} \, \varphi_{X_n}(bt))^n \, ,$$

da cui segue che $Y \stackrel{\mathcal{L}}{=} \sum_{k=1}^n (\frac{a}{n} + bX_k)$, ovvero la v.a. Y è infinitamente divisibile. □

In generale, dalla precedente Proposizione, è immediato verificare che, se una legge ridotta è infinitamente divisibile, lo è anche la relativa legge con parametro di posizione e di scala. Dunque, è sufficiente verificare che legge ridotta sia infinitamente divisibile per concludere che la legge con parametro di posizione e di scala è a sua volta infinitamente divisibile.

• **Esempio 7.5.6.** Si consideri la v.a. X distribuita con legge Normale $\mathcal{N}(0, 1)$. Dal momento che per ogni $n = 1, 2, \ldots$ si ha

$$\varphi_X(t) = e^{-\frac{1}{2}t^2} = \left(e^{-\frac{1}{2n}t^2}\right)^n = \varphi_{X_n}(t)^n \, ,$$

la legge è infinitamente divisibile. In altri termini, una v.a. X con legge Normale $\mathcal{N}(0, 1)$ è equivalente in legge alla somma di n v.a. indipendenti con legge Normale $\mathcal{N}(0, n^{-1})$ per ogni $n = 1, 2, \ldots$. In base alla Proposizione 7.5.3 è inoltre sufficiente verificare che legge Normale ridotta $\mathcal{N}(0, 1)$ è infinitamente divisibile per concludere che la legge Normale $\mathcal{N}(\mu, \sigma^2)$ è a sua volta infinitamente divisibile. □

• **Esempio 7.5.7.** Si consideri la v.a. X distribuita con legge Gamma ridotta $\mathcal{G}(0, 1, k)$. Dal momento che per ogni $n = 1, 2, \ldots$ si ha

$$\varphi_X(t) = (1 - \mathrm{i}t)^{-k} = ((1 - \mathrm{i}t)^{-\frac{k}{n}})^n = \varphi_{X_n}(t)^n \, ,$$

la legge è infinitamente divisibile. In altri termini, una v.a. X con legge Gamma ridotta $\mathcal{G}(0, 1, k)$ è equivalente in legge alla somma di n v.a. indipendenti con legge Gamma $\mathcal{G}(0, 1, \frac{k}{n})$ per ogni $n = 1, 2, \ldots$. □

• **Esempio 7.5.8.** (Legge Stabile) Si consideri una v.a. X con f.c.

$$\varphi_X(t) = e^{-|t|^\alpha} \, ,$$

dove $\alpha \in \,]0, 2]$. La legge associata alla v.a. X è detta *legge Stabile simmetrica* di parametro α e, ovviamente, contiene come casi particolari la legge $\mathcal{N}(0, 2)$ per $\alpha = 2$ (si veda l'Esempio 7.1.6) e la legge di Cauchy ridotta per $\alpha = 1$ (si veda l'esempio 7.1.9). Dal momento che per $\alpha \in \,]0, 2]$ si ha

$$\int_{-\infty}^{\infty} |\varphi_X(t)|\, dt = \int_{-\infty}^{\infty} e^{-|t|^\alpha}\, dt = 2\,\Gamma\left(1+\frac{1}{\alpha}\right) < \infty\,,$$

sulla base del Teorema 7.1.10 la v.a. X è assolutamente continua. La d.p. della v.a. X non può essere ottenuta in forma chiusa, eccetto che per i due casi particolari precedentemente citati. Inoltre, tenendo presente la Proposizione 7.1.6, $\mathrm{E}[X]$ non è definito se $\alpha \in]0,1]$, mentre esiste finito $\mathrm{E}[X]$ ma non $\mathrm{E}[X^2]$ se $\alpha \in]1,2[$. Ovviamente, i momenti di ogni ordine esistono finiti se $\alpha = 2$. Dal momento che per ogni legge della famiglia si ha

$$\varphi_X(t) = e^{-|t|^\alpha} = (e^{-|n^{-\frac{1}{\alpha}}t|^\alpha})^n = \varphi_{X_n}(t)^n\,,$$

la legge Stabile simmetrica è infinitamente divisibile. In altri termini, una v.a. X con legge Stabile simmetrica di parametro α è equivalente in legge alla somma di n v.a. indipendenti del tipo $n^{-\frac{1}{\alpha}}X_k$, $k = 1, \ldots, n$, dove la legge di ogni v.a. X_k è ancora Stabile simmetrica di parametro α. □

• **Esempio 7.5.9.** Si consideri la v.a. X distribuita con legge Binomiale $\mathcal{B}(m,p)$. Dal momento che si ha

$$G_X(t) = (q+pt)^m = ((q+pt)^{\frac{m}{n}})^n\,,$$

la legge è infinitamente divisibile solamente se m è multiplo di n. Dunque, la legge Binomiale non è infinitamente divisibile. □

• **Esempio 7.5.10.** Si consideri la v.a. X distribuita con legge di Poisson $\mathcal{P}(\lambda)$. Dal momento che si ha

$$G_X(t) = e^{\lambda(t-1)} = (e^{\frac{\lambda}{n}(t-1)})^n = G_{X_n}(t)^n\,,$$

la legge è infinitamente divisibile. In altri termini, una v.a. X distribuita con legge di Poisson $\mathcal{P}(\lambda)$ è equivalente in legge alla somma di n v.a. indipendenti con legge di Poisson $\mathcal{P}(\frac{\lambda}{n})$ per ogni $n = 1, 2, \ldots$. □

7.6. Riferimenti bibliografici

Lukacs (1970) è una monografia sulle funzioni caratteristiche. I classici testi di Feller (1968, 1971) trattano estesamente le funzioni generatrici e caratteristiche. Un testo più recente sulle funzioni caratteristiche è Sasvári (2013). Wilf (2006) è una monografia espressamente dedicata alle funzioni generatrici. Sull'argomento si veda anche Petkovšek *et al.* (1996) e Spivey (2019). Una trattazione esauriente delle leggi infinitamente divisibili è data in Sato (1999) e Steutel e van Harn (2004). Le leggi stabili sono estesamente considerate in Uchaikin e Zolotarev (1999) e Zolotarev (1986).

7.7. Esercizi svolti

• **Nota.** Al fine di evitare notazioni ridondanti e ripetizioni, nei seguenti esercizi si assume l'esistenza di uno spazio probabilizzato $(\Omega, \mathcal{F}, P)$, a meno che non venga specificato diversamente.

Sezione 7.1

• **Esercizio 7.1.1.** Si consideri la v.a. Z con legge Normale ridotta $\mathcal{N}(0,1)$ e se ne determini la f.c.

Soluzione. Tenendo presente che la v.a. Z è simmetrica rispetto all'origine e che la funzione seno è dispari, dalla definizione di f.c. si ha

$$\varphi_Z(t) = \mathrm{E}[\cos(tZ)] = \int_{-\infty}^{\infty} \cos(tz)\, \phi(z)\, dz\ .$$

Differenziando i termini della precedente relazione ed effettuando una integrazione per parti tenendo presente che

$$\frac{d}{dz}\, \phi(z) = \ -z\phi(z)\ ,$$

si ottiene

$$\frac{d}{dt}\, \varphi_Z(t) = \ -\int_{-\infty}^{\infty} \sin(tz)\, z\phi(z)\, dz = \sin(tz)\, \phi(z)|_{-\infty}^{\infty} - t\int_{-\infty}^{\infty} \cos(tz)\, \phi(z)\, dz = \ -t\varphi_Z(t)\ .$$

Dunque, la precedente espressione fornisce una equazione differenziale con condizione iniziale data da $\varphi_Z(0) = 1$, di cui si determina facilmente la soluzione $\varphi_Z(t) = e^{-\frac{1}{2}t^2}$. □

• **Esercizio 7.1.2.** Si consideri la v.a. assolutamente continua X con f.c.

$$\varphi_X(t) = \left(1 - \frac{|t|}{\pi}\right) \mathbf{1}_{[-\pi,\pi]}(t)\ .$$

Si determini la d.p. f_X.

Soluzione. Si ha

$$\int_{-\infty}^{\infty} |\varphi_X(t)|\, dt = \ \int_{-\pi}^{\pi} \left(1 - \frac{|t|}{\pi}\right) dt = \pi < \infty\ .$$

Dal momento che la f.c. φ_X è una funzione pari, dal Teorema di Inversione di Lévy si ottiene

$$f_X(x) = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{-\mathrm{i}tx} \left(1 - \frac{|t|}{\pi}\right) dt = \frac{1}{2\pi}\int_{-\pi}^{\pi} \cos(tx) \left(1 - \frac{|t|}{\pi}\right) dt = \frac{1 - \cos(\pi x)}{\pi^2 x^2}\ .$$

La d.p. f_X è in effetti tale, in quanto $f_X(x) \geq 0$ per ogni $x \in \mathbb{R}$ ed inoltre

$$\int_{-\infty}^{\infty} f_X(x)\, dx = \int_{-\infty}^{\infty} \frac{1 - \cos(\pi x)}{\pi^2 x^2}\, dx = \int_{-\infty}^{\infty} \frac{2\sin^2(\frac{\pi x}{2})}{\pi^2 x^2}\, dx = \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{\sin^2(x)}{x^2}\, dx = 1\ .$$

Per una verifica probabilistica dell'integrale $\int_{-\infty}^{\infty} (\frac{\sin(x)}{x})^2 dx = \pi$ mediante l'uso delle f.c. si veda l'Esercizio 7.3.5. □

• **Esercizio 7.1.3.** Si considerino le v.a. indipendenti X_1 e X_2 con legge Normale ridotta $\mathcal{N}(0,1)$ e si determini la f.c. della v.a. $Y = X_1 X_2$.

Soluzione. Tenendo presente l'espressione della f.c. di una v.a. con legge Normale, si ha

$$\varphi_Y(t) = \mathrm{E}[e^{\mathrm{i}tX_1X_2}] = \mathrm{E}[\mathrm{E}[e^{\mathrm{i}tX_1X_2} \mid X_2]] = \mathrm{E}[e^{-\frac{1}{2}t^2X_2^2}]\ .$$

Dunque, risulta

$$\varphi_Y(t) = \int_{-\infty}^{\infty} e^{-\frac{1}{2}t^2x_2^2}\, \frac{1}{\sqrt{2\pi}}\, e^{-\frac{1}{2}x_2^2}\, dx_2 = \frac{1}{\sqrt{1+t^2}}\int_{-\infty}^{\infty} \frac{\sqrt{1+t^2}}{\sqrt{2\pi}}\, e^{-\frac{1}{2}(1+t^2)x_2^2}\, dx_2 = \frac{1}{\sqrt{1+t^2}}\ ,$$

dove si è tenuto presente che la funzione integranda nell'ultimo integrale è la d.p. relativa ad una v.a. con legge Normale $\mathcal{N}(0, (1+t^2)^{-1})$. □

Sezione 7.2

• **Esercizio 7.2.1.** (Valori record, terza parte) Tenendo presente gli Esercizi 2.7.3 e 3.3.2, si consideri la v.a. discreta X con f.p. data da

$$p_X(x) = \frac{1}{x(x+1)}\,\mathbf{1}_{\{1,2,\ldots\}}(x)$$

e si determini la f.g. G_X. Si calcoli inoltre $\mathrm{E}[X]$ mediante la f.g.

Soluzione. Tenendo presente la serie logaritmica, si ha

$$\begin{aligned} G_X(t) &= \sum_{x=1}^{\infty} t^x\,\frac{1}{x(x+1)} = \sum_{x=1}^{\infty} t^x\left(\frac{1}{x}-\frac{1}{x+1}\right) = \sum_{x=1}^{\infty}\frac{t^x}{x} - \frac{1}{t}\sum_{x=1}^{\infty}\frac{t^{x+1}}{x+1} \\ &= -\log(1-t) - \frac{1}{t}\,(-t-\log(1-t)) = 1+\frac{1-t}{t}\log(1-t)\,. \end{aligned}$$

Inoltre, risulta

$$G_X^{(1)}(t) = -\,\frac{t+\log(1-t)}{t^2}\,,$$

da cui $\mathrm{E}[X] = \lim_{t\to 1^-} G_X^{(1)}(t) = \infty$. □

• **Esercizio 7.2.2.** (Legge Uniforme discreta) Si consideri la v.a. discreta X con f.p. data da

$$p_X(x) = \frac{1}{N}\,\mathbf{1}_{\{1,\ldots,N\}}(x)\,,$$

dove $N \in \mathbb{Z}^+$, ovvero la v.a. X si distribuisce con legge Uniforme discreta su $\{1,\ldots,N\}$. Si determini la f.g. G_X della v.a. X. Si calcoli inoltre $\mathrm{E}[X]$ e $\mathrm{Var}[X]$ mediante la f.g.

Soluzione. Tenendo presente le proprietà della serie geometrica, per $t \neq 1$ la f.g. della v.a. X può essere espressa come

$$G_X(t) = \frac{1}{N}\sum_{x=1}^{N} t^x = \frac{t(1-t^N)}{N(1-t)}\,,$$

mentre $G_X(1) = 1$. Dunque, si ha

$$G_X^{(1)}(t) = \frac{1-(N+1)t^N+Nt^{N+1}}{N(1-t)^2}\,,$$

e quindi

$$\mathrm{E}[X] = \lim_{t\to 1^-} G_X^{(1)}(t) = \frac{N+1}{2}\,.$$

Inoltre, risulta

$$G_X^{(2)}(t) = \frac{2-N(N+1)t^{N-1}+2(N^2-1)t^N-N(N-1)t^{N+1}}{N(1-t)^3}\,,$$

e quindi

$$\mathrm{E}[X(X-1)] = \lim_{t\to 1^-} G_X^{(1)}(t) = \frac{N^2-1}{3}\,,$$

da cui

$$\mathrm{Var}[X] = \frac{N^2 - 1}{12} \,.$$

In questo esercizio sono state implicitamente dimostrate le due relazioni $\sum_{k=1}^{n} k = \frac{1}{2} n(n+1)$ e $\sum_{k=1}^{n} k^2 = \frac{1}{6} n(n+1)(2n+1)$. Alla stessa maniera, con calcoli evidentemente sempre più laboriosi, si può ottenere le espressioni di $\sum_{k=1}^{n} k^s$ per $s = 3, 4, \ldots$. □

• **Esercizio 7.2.3.** (Mistura di leggi) Dato il v.v.a. $(X, Y)^{\mathrm{T}}$, si assuma che la legge condizionata della v.a. X all'evento $\{Y = y\}$ sia la legge di Poisson $\mathcal{P}(y)$. Inoltre, si assuma che la v.a. Y si distribuisca con legge Esponenziale ridotta $\mathcal{G}(0, 1, 1)$. Si determini la legge della v.a. X.

Soluzione. Tenendo presente l'espressione della f.g. di una v.a. con legge di Poisson, la f.g. della v.a. X è data da

$$G_X(t) = \mathrm{E}[\mathrm{E}[t^X \mid Y]] = \mathrm{E}[e^{Y(t-1)}] = \int_0^\infty e^{y(t-1)}\, e^{-y}\, dy = \frac{1}{2-t} = \frac{\frac{1}{2}}{1 - \frac{1}{2}\, t} \,.$$

Dunque, la v.a. X si distribuisce con legge Geometrica di parametro $p = \frac{1}{2}$. In modo alternativo, la f.p. della v.a. X può essere ottenuta come

$$p_X(x) = \int_0^\infty e^{-y}\, \frac{y^x}{x!}\, \mathbf{1}_{\mathbb{N}}(x)\, e^{-y}\, dy = \frac{1}{x!}\, 2^{-x-1}\, \mathbf{1}_{\mathbb{N}}(x) \int_0^\infty z^x\, e^{-z}\, dz = 2^{-x-1}\, \mathbf{1}_{\mathbb{N}}(x) \,,$$

ovvero si ha una conferma del precedente risultato. In generale, una *mistura di leggi* viene definita come una legge con un parametro che a sua volta è una v.a. con una ulteriore legge. Nel presente caso, la legge Geometrica è una mistura di leggi di Poisson il cui parametro si distribuisce con legge Esponenziale ridotta. □

Sezione 7.3

• **Esercizio 7.3.1.** Si considerino le v.a. indipendenti X_1 e X_2 che si distribuiscono con la stessa legge della v.a. X. Se φ_X è la f.c. della v.a. X, si verifichi che la f.c. della v.a. $Y = X_1 - X_2$ è data da $\varphi_Y(t) = |\varphi_X(t)|^2$.

Soluzione. La f.c. della v.a. Y risulta

$$\varphi_Y(t) = \mathrm{E}[e^{\mathrm{i}t(X_1 - X_2)}] = \mathrm{E}[e^{\mathrm{i}tX_1}]\mathrm{E}[e^{-\mathrm{i}tX_2}] = \varphi_X(t)\varphi_X(-t) \,.$$

Dalla definizione di f.c. si ha inoltre

$$\begin{aligned} \varphi_X(t)\varphi_X(-t) &= (\mathrm{E}[\cos(tX)] + \mathrm{i}\,\mathrm{E}[\sin(tX)])(\mathrm{E}[\cos(-tX)] + \mathrm{i}\,\mathrm{E}[\sin(-tX)]) \\ &= (\mathrm{E}[\cos(tX)] + \mathrm{i}\,\mathrm{E}[\sin(tX)])(\mathrm{E}[\cos(tX)] - \mathrm{i}\,\mathrm{E}[\sin(tX)]) \\ &= \mathrm{E}[\cos(tX)]^2 + \mathrm{E}[\sin(tX)]^2 = |\varphi_X(t)|^2 \,. \end{aligned}$$ □

• **Esercizio 7.3.2.** Si consideri la v.a. X distribuita con legge di Cauchy ridotta, ovvero la v.a. X ammette d.p. data da

$$f_X(x) = \frac{1}{\pi(1 + x^2)} \,.$$

Si determini la f.c. φ_X della v.a. X.

Soluzione. Dal momento che la v.a. X è simmetrica rispetto all'origine, la f.c. può essere espressa come

$$\varphi_X(t) = \mathrm{E}[\cos(tZ)] = \int_{-\infty}^{\infty} \cos(tx)\, \frac{1}{\pi(1+x^2)}\, dx\ .$$

Al fine di calcolare il precedente integrale, si considerino due v.a. Y_1 e Y_2 indipendenti con legge Esponenziale ridotta e la v.a. differenza $Y = Y_1 - Y_2$. Tenendo presente l'Esercizio 6.8.1, la v.a. Y si distribuisce con legge di Laplace ridotta e dunque si ha

$$f_Y(y) = \frac{1}{2}\, e^{-|y|}\ .$$

Inoltre, risulta $\varphi_{Y_1}(t) = \varphi_{Y_2}(t) = (1 - \mathrm{i}t)^{-1}$ e, sulla base dell'Esercizio 7.3.1, la f.c. della v.a. Y è data da $\varphi_Y(t) = (1+t^2)^{-1}$. Dal momento che

$$\int_{-\infty}^{\infty} |\varphi_Y(t)|\, dt = \int_{-\infty}^{\infty} \frac{1}{1+t^2}\, dt = \pi < \infty\ ,$$

dal Teorema di Inversione si ottiene un'ulteriore espressione per f_Y data da

$$f_Y(y) = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-\mathrm{i}ty}\, \frac{1}{1+t^2}\, dt = \frac{1}{2\pi} \int_{-\pi}^{\pi} \cos(ty)\, \frac{1}{1+t^2}\, dt\ ,$$

ovvero si ha

$$\int_{-\infty}^{\infty} \cos(ty)\, \frac{1}{\pi(1+t^2)}\, dt = e^{-|y|}\ .$$

Quindi, si deve concludere che $\varphi_X(t) = e^{-|t|}$. □

• **Esercizio 7.3.3.** Si consideri la v.a. X distribuita con legge di Cauchy ridotta descritta nell'Esercizio 7.3.2. Se le v.a. X_1 e X_2 sono distribuite con la stessa legge di X e sono v.a. tali che $\{X_1 = X_2\}$ $q.c.$, si verifichi che

$$\varphi_{X_1+X_2}(t) = \varphi_{X_1}(t)\varphi_{X_2}(t)\ .$$

Soluzione. Ovviamente, le v.a. X_1 e X_2 non sono indipendenti. Dall'Esercizio 7.3.2 la f.c. della v.a. X risulta $\varphi_X(t) = e^{-|t|}$. Dal momento che $X_1 + X_2 \stackrel{\mathcal{L}}{=} Y$ dove $Y = 2X$, si ha

$$\varphi_{X_1+X_2}(t) = \varphi_Y(t) = \varphi_X(2t) = e^{-2|t|} = \varphi_{X_1}(t)\varphi_{X_2}(t)\ .$$

Dunque, la f.c. della somma di due v.a. può essere identica al prodotto delle f.c. delle v.a., anche se queste non sono indipendenti. □

• **Esercizio 7.3.4.** Si consideri la v.a. X con legge di Laplace ridotta descritta nell'Esercizio 6.8.1. Se inoltre $(X_1, \ldots, X_n)^{\mathrm{T}}$ è un v.v.a. con componenti indipendenti e con la stessa legge di X, si determini la legge della v.a. $Y = \sum_{k=1}^{n} |X_k|$.

Soluzione. Posto $Z = |X|$, dal momento che la v.a. X è simmetrica rispetto all'origine, allora risulta

$$f_Z(z) = 2 f_X(z)\, \mathbf{1}_{[0,\infty[}(z) = e^{-z}\, \mathbf{1}_{[0,\infty[}(z)\ ,$$

ovvero la v.a. Z si distribuisce con legge Esponenziale ridotta. Dal momento che si ha $\varphi_Z(t) = (1 - \mathrm{i}t)^{-1}$, allora

$$\varphi_Y(t) = \varphi_Z(t)^n = (1 - \mathrm{i}t)^{-n}\ ,$$

ovvero la v.a. Y si distribuisce con legge Gamma ridotta $\mathcal{G}(0, 1, n)$. □

• **Esercizio 7.3.5.** (Legge Triangolare, terza parte) Si consideri la v.a. X distribuita con la legge Triangolare descritta nell'Esercizio 3.7.1. Sulla base delle proprietà di questa legge e mediante l'uso della f.c. si verifichi che

$$\int_{-\infty}^{\infty} \frac{\sin^2(x)}{x^2}\, dx = \pi\,.$$

Soluzione. Tenendo presente l'Esercizio 3.7.2, si ha $X = \frac{1}{2}Y$, dove Y è una v.a. data dalla somma di due v.a. che si distribuiscono come una v.a. Z con legge Uniforme su $]-1,1[$. Dal momento che

$$\varphi_Z(t) = \int_{-1}^{1} e^{\mathrm{i}tz}\,\frac{1}{2}\, dz = \frac{1}{2}\int_{-1}^{1} \cos(tz)\, dz = \frac{\sin(t)}{t}\,,$$

allora si ha $\varphi_Y(t) = \frac{\sin^2(t)}{t^2}$ e quindi

$$\varphi_X(t) = \varphi_Y\left(\frac{t}{2}\right) = \frac{\sin^2(\frac{t}{2})}{(\frac{t}{2})^2}\,.$$

Inoltre, essendo

$$\int_{-\infty}^{\infty} |\varphi_X(t)|\, dt = \int_{-\infty}^{\infty} \frac{\sin^2(\frac{t}{2})}{(\frac{t}{2})^2}\, dt = 2\int_{-\infty}^{\infty} \frac{\sin^2(t)}{t^2}\, dt < 2\int_{-\infty}^{\infty} \min(1, t^{-2})\, dt = 8 < \infty\,,$$

dal Teorema di Inversione si ottiene un'ulteriore espressione per f_X data da

$$f_X(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-\mathrm{i}tx}\,\frac{\sin^2(\frac{t}{2})}{(\frac{t}{2})^2}\, dt = \frac{1}{2\pi}\int_{-\infty}^{\infty} \cos(tx)\,\frac{\sin^2(\frac{t}{2})}{(\frac{t}{2})^2}\, dt = \frac{1}{\pi}\int_{-\infty}^{\infty} \cos(2tx)\,\frac{\sin^2(t)}{t^2}\, dt\,,$$

ovvero, tenendo presente l'Esercizio 3.7.1, si ha

$$(1-|x|)\,\mathbf{1}_{]-1,1[}(x) = \frac{1}{\pi}\int_{-\infty}^{\infty} \cos(2tx)\,\frac{\sin^2(t)}{t^2}\, dt\,.$$

Ponendo $x = 0$ nella precedente espressione si ha l'integrale desiderato. □

• **Esercizio 7.3.6.** Si considerino le v.a. indipendenti X_1 e X_2, tali che X_1 si distribuisce con legge Uniforme su $[-1,1]$, mentre X_2 è una v.a. discreta con f.p. data da

$$p_{X_2}(x_2) = \frac{1}{2}\,\mathbf{1}_{\{-1,1\}}(x_2)\,.$$

Si determini la legge della v.a. $Y = X_1 + X_2$.

Soluzione. La f.c. della v.a. X_1 è data da $\varphi_{X_1}(t) = \frac{\sin(t)}{t}$, mentre la f.c. della v.a. X_2 è data da $\varphi_{X_2}(t) = \cos(t)$. Dunque, la f.c. della v.a. Y risulta

$$\varphi_Y(t) = \varphi_{X_1}(t)\varphi_{X_2}(t) = \frac{\sin(t)\cos(t)}{t} = \frac{\sin(2t)}{2t}\,,$$

da cui risulta $Y \stackrel{\mathcal{L}}{=} 2X_1$. Dunque, la v.a. Y si distribuisce con legge Uniforme su $[-2,2]$. □

• **Esercizio 7.3.7.** (Identità di Wald) Si consideri la v.a. $X = \sum_{k=1}^{Y} Z_k$, dove Y è una v.a. discreta a valori su $\mathbb{N}$ tale che $\mathrm{E}[Y] < \infty$, mentre le Z_k sono v.a. indipendenti con la medesima legge della v.a. Z, dove $\mathrm{E}[Z] < \infty$. Inoltre le v.a. Z_k sono indipendenti dalla v.a. Y. Si verifichi che risulta $\mathrm{E}[X] = \mathrm{E}[Y]\mathrm{E}[Z]$.

Soluzione. Si ha

$$\varphi_X(t) = \mathrm{E}[e^{\mathrm{i}t\sum_{k=1}^{Y} Z_k}] = \mathrm{E}[\mathrm{E}[e^{\mathrm{i}t\sum_{k=1}^{Y} Z_k} \mid Y]] = \mathrm{E}[\varphi_Z(t)^Y] = G_Y(\varphi_Z(t)) \,.$$

Dunque, dalla precedente relazione risulta

$$\varphi_X^{(1)}(t) = G_Y^{(1)}(\varphi_Z(t))\varphi_Z^{(1)}(t) \,,$$

da cui

$$\mathrm{E}[X] = \mathrm{i}^{-1}\varphi_X^{(1)}(0) = \mathrm{i}^{-1}G_Y^{(1)}(\varphi_Z(0))\varphi_Z^{(1)}(0) = G_Y^{(1)}(1)\,\mathrm{i}^{-1}\varphi_Z^{(1)}(0) = \mathrm{E}[Y]\mathrm{E}[Z] \,.$$

L'identità prende il nome dal matematico e probabilista ungherese Abraham Wald (1902-1950), che ha introdotto le tecniche di statistica sequenziale. □

Sezione 7.4

• **Esercizio 7.4.1.** (Legge di Poisson composta) Si consideri la v.a. Y con legge di Poisson $\mathcal{P}(\lambda)$ e la v.a. discreta Z con f.p.

$$p_Z(z) = \, -\frac{1}{\log(1-q)}\,\frac{q^z}{z}\,\mathbf{1}_{\{1,2,\ldots\}}(z) \,,$$

dove $q \in \,]0,1[$. Si determini la legge relativa alla v.a. $X = \sum_{k=1}^{Y} Z_k$, dove le Z_k sono v.a. indipendenti con la medesima legge della v.a. Z.

Soluzione. Tenendo presente l'Esempio 3.3.1, si osservi che la v.a. Z si distribuisce con legge Logaritmica di parametro q. La f.g. della v.a. Z è quindi data da

$$G_Z(t) = \sum_{z=1}^{\infty} t^z \left(-\frac{1}{\log(1-q)}\,\frac{q^z}{z}\right) = \, -\frac{1}{\log(1-q)}\sum_{z=1}^{\infty}\frac{(qt)^z}{z} = \frac{\log(1-qt)}{\log(1-q)} \,.$$

Dal momento che X è la somma di v.a. indipendenti con la medesima legge e un numero aleatorio di addendi distribuito con legge di Poisson, la f.g. della v.a. X è data da

$$G_X(t) = \mathrm{E}[t^{\sum_{k=1}^{Y} Z_k}] = \mathrm{E}[\mathrm{E}[t^{\sum_{k=1}^{Y} Z_k} \mid Y]] = \mathrm{E}[G_Z(t)^Y] = G_Y(G_Z(t)) \,,$$

da cui

$$G_X(t) = e^{\lambda\left(\frac{\log(1-qt)}{\log(1-q)}-1\right)} = \left(\frac{1-q}{1-qt}\right)^{-\frac{\lambda}{\log(1-q)}} = \left(\frac{p}{1-qt}\right)^k \,,$$

dove si è posto $p = 1-q$ e $k = \, -\frac{\lambda}{\log(p)}$. Si deve dunque concludere che la v.a. X si distribuisce con legge Binomiale Negativa $\mathcal{BN}(k,p)$. La legge Binomiale Negativa è quindi una distribuzione di Poisson composta, ovvero è la legge della somma di v.a. indipendenti e ugualmente distribuite con legge Logaritmica e con un numero aleatorio di addendi distribuito con legge di Poisson. □

• **Esercizio 7.4.2.** (Legge di Poisson composta, seconda parte) Si consideri la v.a. X tale che $X = \sum_{k=1}^{Y} Z_k$, dove la legge della v.a. Y è Poisson $\mathcal{P}(\lambda)$, mentre le Z_k sono v.a. indipendenti con legge di Bernoulli $\mathcal{B}(1,p)$. Si determini la legge della v.a. X.

Soluzione. In modo simile all'Esercizio 7.4.1, tenendo presente l'espressione della f.g. di una v.a. con legge di Bernoulli $\mathcal{B}(1,p)$, si ha

$$G_X(t) = G_Y(G_Z(t)) = G_Y(1-p+pt) = e^{p\lambda(t-1)} \,.$$

Dunque, si deve concludere che la v.a. X si distribuisce con legge di Poisson $\mathcal{P}(p\lambda)$. □

• **Esercizio 7.4.3.** (Formula di Waring) Si dimostri mediante l'uso delle funzioni generatrici la formula di Waring introdotta nell'Esercizio 2.2.1.

Soluzione. Assumendo le notazioni dell'Esercizio 2.2.1, sia $Y = \sum_{k=1}^{n} X_k$, dove $X_k = \mathbf{1}_{E_k}$. Evidentemente, il v.v.a. $(X_1, \ldots, X_n)^{\mathrm{T}}$ possiede componenti marginali che sono v.a. non indipendenti con legge di Bernoulli di parametro $P(E_k) = \mathrm{E}[\mathbf{1}_{E_k}] = \mathrm{E}[X_k]$. La f.g. della v.a. X_k è quindi data da

$$G_{X_k}(t) = 1 - \mathrm{E}[X_k] + \mathrm{E}[X_k]t = \mathrm{E}[1 + (t-1)X_k]$$

e quindi la f.g. della v.a. Y risulta

$$G_Y(t) = \mathrm{E}[\prod_{k=1}^{n} (1 + (t-1)X_k)] = \mathrm{E}[1 + \sum_{k=1}^{n} \sum_{1 \le j_1 < \ldots < j_k \le n} (t-1)^k \, X_{j_1} \ldots X_{j_k}] \, .$$

Tenendo presente che $\mathrm{E}[X_{j_1} \ldots X_{j_k}] = P(E_{j_1} \cap \ldots \cap E_{j_k})$, si ha dunque

$$G_Y(t) = 1 + \sum_{k=1}^{n} (t-1)^k \sum_{1 \le j_1 < \ldots < j_k \le n} P(E_{j_1} \cap \ldots \cap E_{j_k}) = 1 + \sum_{k=1}^{n} (t-1)^k \, S_{k,n} \, .$$

Dal momento che

$$G_Y^{(m)}(t) = \sum_{k=m}^{n} \frac{k!}{(k-m)!} \, (t-1)^{k-m} S_{k,m} \, ,$$

per le proprietà della f.g. si ha

$$p_Y(m) = \frac{G_Y^{(m)}(0)}{m!} = \sum_{k=m}^{n} \binom{k}{m} (-1)^{k-m} S_{k,n} \, .$$

Si osservi che

$$P\left(\bigcup_{k=1}^{n} E_k \right) = 1 - p_Y(0) = \sum_{k=1}^{n} (-1)^{k+1} S_{k,n} \, ,$$

ovvero si è anche ottenuto una ulteriore dimostrazione della Formula di Inclusione ed Esclusione. □

Sezione 7.5

• **Esercizio 7.5.1.** Si consideri la v.a. X con legge di Cauchy descritta nell'Esercizio 7.3.2 e si verifichi che è infinitamente divisibile.

Soluzione. Tenendo presente l'Esercizio 7.3.2, per ogni $n = 1, 2, \ldots$ si ha

$$\varphi_X(t) = e^{-|t|} = \left(e^{-\frac{1}{n}|t|} \right)^n = \varphi_{X_n}(t)^n \, ,$$

dove per le proprietà della f.c. la v.a. X_n ammette d.p. data da

$$f_{X_n}(x) = \frac{n}{\pi(1 + n^2 x^2)} \, .$$

Dunque, la legge è infinitamente divisibile. In altri termini, una v.a. X con legge di Cauchy è equivalente in legge alla somma di n v.a. indipendenti con legge di Cauchy con parametro di scala pari a n^{-1} per ogni $n = 1, 2, \ldots$. □

• **Esercizio 7.5.2.** (Legge di Poisson composta, terza parte) Si consideri la v.a. X tale che $X = \sum_{k=1}^{Y} Z_k$, dove la v.a. Y si distribuisce con legge di Poisson $\mathcal{P}(\lambda)$, mentre le Z_k sono v.a. indipendenti con la stessa legge della v.a. Z. Si verifichi che la v.a. X è infinitamente divisibile.

Soluzione. Tenendo presente l'espressione della f.g. della v.a. Y, la f.c. della v.a. X è data da

$$\varphi_X(t) = \mathrm{E}[e^{\mathrm{i}t\sum_{k=1}^{Y} Z_k}] = \mathrm{E}[\mathrm{E}[e^{\mathrm{i}t\sum_{k=1}^{Y} Z_k} \mid Y]] = \mathrm{E}[\varphi_Z(t)^Y] = G_Y(\varphi_Z(t)) = e^{\lambda(\varphi_Z(t)-1)} .$$

Dunque, la v.a. è infinitamente divisibile dal momento che per ogni $n = 1, 2, \ldots$ si ha

$$\varphi_X(t) = \left(e^{\frac{\lambda}{n}(\varphi_Z(t)-1)}\right)^n = \varphi_{X_n}(t)^n ,$$

dove $X_n = \sum_{k=1}^{Y_n} Z_k$ e la legge della v.a. Y_n è Poisson $\mathcal{P}(\frac{\lambda}{n})$. □

Pagina intenzionalmente vuota

Capitolo 8

Convergenze

8.1. Convergenza in legge

Un insieme numerabile di v.a. $(X_1, X_2, \dots)$ è detta *successione di v.a.* e viene indicata con la notazione $(X_n)_{n\geq 1}$. Se la successione viene indicizzata come $(X_0, X_1, \dots)$, si adotta eventualmente la notazione $(X_n)_{n\geq 0}$. Dal momento che alla successione $(X_n)_{n\geq 1}$ è associata la corrispondente successione di f.r. $(F_{X_n})_{n\geq 1}$, risulta interessante definire il concetto di convergenza per una successione di f.r. Formalmente, la cosiddetta *convergenza in legge* è data nella seguente definizione.

Definizione 8.1.1. Se $(X_n)_{n\geq 1}$ è una successione di v.a., a cui corrisponde la successione di f.r. $(F_{X_n})_{n\geq 1}$, si dice che la successione di v.a. *converge in legge* (o *in distribuzione*) alla v.a. X con f.r. F_X se per ogni x tale che $\triangle F_X(x) = 0$ si ha

$$\lim_n F_{X_n}(x) = F_X(x)$$

e si scrive $X_n \xrightarrow{\mathcal{L}} X$ per $n \to \infty$. □

In pratica, la definizione di convergenza in legge implica che la f.r. della v.a. X_n converga puntualmente a quella della v.a. X per ogni x per cui F_X è continua. Si osservi inoltre che la convergenza in legge non richiede che le v.a. della successione e la v.a. X siano definite su uno stesso spazio probabilizzato. In effetti, la condizione della Definizione 8.1.1 riguarda solamente le f.r. e introduce la cosiddetta *convergenza debole* della successione di f.r. $(F_{X_n})_{n\geq 1}$ alla f.r. F_X. Data la biunivocità esistente fra legge e f.r. di una v.a., se $(P_{X_n})_{n\geq 1}$ denota la successione di leggi corrispondente alla successione di f.r. $(F_{X_n})_{n\geq 1}$ e se P_X è la legge relativa alla f.r. F_X, la convergenza debole può essere anche definita mediante la condizione

$$\lim_n P_{X_n}(B) = P_X(B)$$

per ogni $B \in \mathcal{B}(\mathbb{R})$ la cui frontiera (ovvero l'insieme di punti di discontinuità di $\mathbf{1}_B$) ha probabilità nulla secondo P_X.

Nel caso in cui la v.a. X sia degenere e tale che $P(X = c) = 1$, allora si ha $F_X(x) = \mathbf{1}_{[c,\infty[}(x)$ e F_X è continua per $x \neq c$. Dunque, la condizione della Definizione 8.1.1 si riduce a

$$\lim_n F_{X_n}(x) = \mathbf{1}_{[c,\infty[}(x)$$

per ogni $x \neq c$. In questo caso, con un leggero abuso in notazione generalmente adottato nei testi di Teoria della Probabilità, la convergenza in legge viene denotata come $X_n \xrightarrow{\mathcal{L}} c$ per $n \to \infty$.

• **Esempio 8.1.1.** Si consideri la successione di v.a. $(X_n)_{n\geq 1}$ tale che la v.a. X_n si distribuisce con legge Beta ridotta $\mathcal{BE}(0, 1, 1 + n^{-1}, 1)$. Tenendo presente quanto detto nella Sezione 6.9, si ha

$$F_{X_n}(x) = x^{1+\frac{1}{n}}\, \mathbf{1}_{]0,1[}(x) + \mathbf{1}_{[1,\infty[}(x)\,,$$

da cui

$$\lim_n F_{X_n}(x) = x\,\mathbf{1}_{]0,1[}(x) + \mathbf{1}_{[1,\infty[}(x)\,.$$

Dal momento che la precedente espressione fornisce la f.r. di una v.a. X con legge Uniforme su $]0,1[$, allora si ha $X_n \xrightarrow{\mathcal{L}} X$ per $n \to \infty$. □

• **Esempio 8.1.2.** Si consideri la successione di v.a. $(X_n)_{n\geq 1}$ tale che la v.a. X_n si distribuisce con legge Uniforme su $[-n^{-1}, n^{-1}]$. Tenendo presente quanto detto nella Sezione 6.9, si ha

$$F_{X_n}(x) = \frac{nx+1}{2}\,\mathbf{1}_{[-n^{-1},n^{-1}]}(x) + \mathbf{1}_{]n^{-1},\infty[}(x)\,,$$

da cui

$$\lim_n F_{X_n}(x) = \frac{1}{2}\,\mathbf{1}_{\{0\}}(x) + \mathbf{1}_{]0,\infty[}(x)\,.$$

Risulta immediato verificare che la precedente espressione non costituisce una f.r. Tuttavia, si consideri la v.a. degenere X concentrata sullo 0 e che possiede f.r. data da $F_X(x) = \mathbf{1}_{[0,\infty[}(x)$. Dal momento che $\lim_n F_{X_n}(x)$ e $F_X(x)$ coincidono per ogni $x \neq 0$, allora si ha $X_n \xrightarrow{\mathcal{L}} 0$ per $n \to \infty$. □

• **Esempio 8.1.3.** Sia $(Z_n)_{n\geq 1}$ una successione di v.a. indipendenti definite sullo stesso spazio probabilizzato $(\Omega, \mathcal{F}, P)$ e tale che la v.a. Z_n si distribuisce con legge Esponenziale ridotta $\mathcal{G}(0,1,1)$ per ogni n. Inoltre, si consideri la successione di v.a. $(Y_n)_{n\geq 1}$, dove $Y_n = \max_{1\leq k\leq n} Z_k$. Tenendo presente i risultati della Sezione 6.8, dall'assunzione di indipendenza si ha

$$\begin{aligned} F_{Y_n}(x) &= P(\max_{1\leq k\leq n} Z_k \leq x) = P(Z_1 \leq x, \ldots, Z_n \leq x) = \prod_{k=1}^{n} P(Z_k \leq x) \\ &= \prod_{k=1}^{n} F_{Z_k}(x) = \prod_{k=1}^{n} (1-e^{-x})\,\mathbf{1}_{]0,\infty[}(x) = (1-e^{-x})^n\,\mathbf{1}_{]0,\infty[}(x)\,. \end{aligned}$$

Si consideri l'ulteriore successione di v.a. $(X_n)_{n\geq 1}$, con $X_n = Y_n - \log(n)$. In questo caso, si ha

$$\begin{aligned} F_{X_n}(x) &= P(X_n \leq x) = P(Y_n - \log(n) \leq x) = P(Y_n \leq x + \log(n)) \\ &= F_{Y_n}(x + \log(n)) = (1 - e^{-x-\log(n)})^n\,\mathbf{1}_{]-\log(n),\infty[}(x) = \left(1 - \frac{e^{-x}}{n}\right)^n \mathbf{1}_{]-\log(n),\infty[}(x)\,, \end{aligned}$$

da cui

$$\lim_n F_{X_n}(x) = \lim_n \left(1 - \frac{e^{-x}}{n}\right)^n \mathbf{1}_{]-\log(n),\infty[}(x) = e^{-e^{-x}}\,.$$

La precedente espressione fornisce la f.r. di una v.a. X con legge di Gumbel ridotta (si veda l'Esempio 6.6.1) e quindi si ha che $X_n \xrightarrow{\mathcal{L}} X$ per $n \to \infty$. In altre parole, la successione delle f.r. relative ai massimi ("centrati" con un opportuno parametro di posizione) di n v.a. indipendenti con legge Esponenziale ridotta converge alla f.r. relativa ad una legge di Gumbel ridotta. □

I prossimi Teoremi risultano fondamentali per ottenere la convergenza in legge mediante la convergenza delle f.c. o delle f.g. In effetti, dal momento che la f.c. o la f.g. determinano in maniera univoca la f.r. di una v.a., è naturale aspettarsi che la convergenza in legge si possa stabilire tramite la successione delle f.c. o delle f.g.

Teorema 8.1.2. (Teorema di Lévy) *Se $(X_n)_{n\geq 1}$ è una successione di v.a. a cui corrisponde la successione di f.c. $(\varphi_{X_n})_{n\geq 1}$ e se X è una v.a. con f.c. φ_X, allora $X_n \xrightarrow{\mathcal{L}} X$ per $n \to \infty$ se e solo se*

$$\lim_n \varphi_{X_n}(t) = \varphi_X(t)$$

per ogni $t \in \mathbb{R}$.

Dimostrazione. Si veda Billingsley (1995, p.349). □

Si osservi che, nel caso in cui la v.a. X sia degenere e concentrata su c, allora si ha $\varphi_X(t) = e^{\mathrm{i}ct}$, e la condizione del Teorema di Lévy si riduce alla condizione

$$\lim_n \varphi_{X_n}(t) = e^{\mathrm{i}ct}$$

per ogni $t \in \mathbb{R}$.

• **Esempio 8.1.4.** Sia $(Z_n)_{n\geq 1}$ una successione di v.a. tale che la v.a. Z_n si distribuisce con legge Chi-quadrato χ^2_n. Tenendo presente che dalla Sezione 6.8 risulta $\mathrm{E}[Z_n] = n$ e $\mathrm{Var}[Z_n] = 2n$, si consideri la successione di v.a. standardizzate $(X_n)_{n\geq 1}$, dove $X_n = \frac{Z_n - n}{\sqrt{2n}}$. Dunque, dall'Esempio 7.1.7 e dalla Proposizione 7.1.5 si ha

$$\log(\varphi_{X_n}(t)) = -\sqrt{\frac{n}{2}}\,\mathrm{i}t - \frac{n}{2}\log\left(1 - \sqrt{\frac{2}{n}}\,\mathrm{i}t\right).$$

Inoltre, dalle proprietà dell'espansione in serie della funzione logaritmica in campo complesso, per ogni $c \in \mathbb{C}$ tale che $|c| < 1$ risulta

$$\log(1+c) = c - \frac{1}{2}c^2 + o(c^2),$$

da cui

$$\log\left(1 - \sqrt{\frac{2}{n}}\,\mathrm{i}t\right) = -\sqrt{\frac{2}{n}}\,\mathrm{i}t + \frac{t^2}{n} + o(n^{-1}).$$

Dunque, si ha

$$\log(\varphi_{X_n}(t)) = -\frac{t^2}{2} + no(n^{-1}),$$

da cui

$$\lim_n \varphi_{X_n}(t) = e^{-\frac{1}{2}t^2}.$$

Risulta immediato riconoscere che la precedente espressione è la f.c. relativa a una v.a. X con legge Normale ridotta $\mathcal{N}(0,1)$ e perciò si ottiene infine che $X_n \xrightarrow{\mathcal{L}} X$ per $n \to \infty$. □

• **Esempio 8.1.5.** Analogamente all'Esempio 8.1.4, si consideri la successione di v.a. $(Z_n)_{n\geq 1}$, tale che la v.a. Z_n si distribuisce con legge Chi-quadrato χ^2_n. Si consideri inoltre la successione di v.a. $(X_n)_{n\geq 1}$, dove $X_n = n^{-1}Z_n$. Tenendo presente l'Esempio 7.1.7, si ottiene

$$\varphi_{X_n}(t) = \left(1 - \frac{2\mathrm{i}t}{n}\right)^{-\frac{1}{2}n}.$$

Dunque, si ha

$$\lim_n \varphi_{X_n}(t) = e^{\mathrm{i}t} .$$

Dal momento che la precedente espressione fornisce la f.c. di una v.a. degenere X concentrata sul valore 1, risulta dunque $X_n \xrightarrow{\mathcal{L}} 1$ per $n \to \infty$. □

• **Esempio 8.1.6.** Si consideri la successione di v.a. $(Z_n)_{n\geq 1}$ tale che la v.a. Z_n si distribuisce con legge di Poisson $\mathcal{P}(\lambda_n)$. Inoltre, $(\lambda_n)_{n\geq 1}$ è una successione per cui $\lambda_n \to \infty$ per $n \to \infty$. Dalla Sezione 6.2 si ha $\mathrm{E}[Z_n] = \lambda_n$ e $\mathrm{Var}[Z_n] = \lambda_n$, e quindi si consideri la successione di v.a. standardizzate $(X_n)_{n\geq 1}$, dove $X_n = \frac{Z_n - \lambda_n}{\sqrt{\lambda_n}}$. Dunque, dall'Esempio 7.1.2 e dalla Proposizione 7.1.5 si ha

$$\log(\varphi_{X_n}(t)) = \lambda_n\, e^{\frac{1}{\sqrt{\lambda_n}}\mathrm{i}t} - \lambda_n - \sqrt{\lambda_n}\,\mathrm{i}t .$$

Inoltre, dalle proprietà dell'espansione in serie della funzione esponenziale in campo complesso, per ogni $c \in \mathbb{C}$ risulta

$$e^c = 1 + c + \frac{1}{2}\,c^2 + o(c^2) ,$$

da cui

$$e^{\frac{1}{\sqrt{\lambda_n}}\mathrm{i}t} = 1 + \frac{\mathrm{i}t}{\sqrt{\lambda_n}} - \frac{t^2}{2\lambda_n} + o(\lambda_n^{-1}) .$$

Dunque, si ha

$$\log(\varphi_{X_n}(t)) = -\frac{t^2}{2} + \lambda_n o(\lambda_n^{-1}) ,$$

da cui

$$\lim_n \varphi_{X_n}(t) = e^{-\frac{1}{2}t^2} .$$

La precedente espressione fornisce la f.c. relativa a una v.a. X con legge Normale ridotta $\mathcal{N}(0,1)$ e quindi si ha infine che $X_n \xrightarrow{\mathcal{L}} X$ per $n \to \infty$. □

Teorema 8.1.3. *Se $(X_n)_{n\geq 1}$ è una successione di v.a. discrete a valori su $S_n \subseteq \mathbb{N}$ a cui corrisponde la successione di f.g. $(G_{X_n})_{n\geq 1}$ e se X è un'ulteriore v.a. discreta a valori su $S \subseteq \mathbb{N}$ con f.g. G_X, allora $X_n \xrightarrow{\mathcal{L}} X$ per $n \to \infty$ se e solo se*

$$\lim_n G_{X_n}(t) = G_X(t)$$

per ogni $t \in [0,1]$.

Dimostrazione. Si veda Gut (2005, p.240). □

• **Esempio 8.1.7.** Si consideri la successione di v.a. $(X_n)_{n\geq 1}$ tale che la v.a. X_n si distribuisce con legge Binomiale $\mathcal{B}(n, p_n)$. Inoltre, la successione $(p_n)_{n\geq 1}$ è tale che l'ulteriore successione $(np_n)_{n\geq 1}$ converge alla costante $\lambda \in]0, \infty[$. Tenendo presente l'Esempio 7.2.1, si ha

$$G_{X_n}(t) = (1 - p_n + p_n t)^n = \left(1 + \frac{np_n(t-1)}{n}\right)^n .$$

Dunque, si ha

$$\lim_n G_{X_n}(t) = e^{\lambda(t-1)} .$$

La precedente espressione fornisce la f.g. di una v.a. X distribuita con legge di Poisson $\mathcal{P}(\lambda)$ e dunque $X_n \xrightarrow{\mathcal{L}} X$ per $n \to \infty$. Questo risultato è in effetti già noto dalla Sezione 6.2. □

• **Esempio 8.1.8.** Si consideri la successione di v.a. $(X_n)_{n\geq 1}$ tale che la v.a. X_n si distribuisce con legge Binomiale Negativa $\mathcal{BN}(n, p_n)$. Inoltre, la successione $(p_n)_{n\geq 1}$ è tale che l'ulteriore successione $(n\frac{q_n}{p_n})_{n\geq 1}$, dove $q_n = 1 - p_n$, converge alla costante $\lambda \in]0, \infty[$. Tenendo presente l'Esempio 7.2.3, si ha

$$G_{X_n}(t) = \left(\frac{p_n}{1 - q_n t}\right)^n = \left(1 + \frac{n\frac{q_n}{p_n}(1-t)}{n}\right)^{-n} .$$

Dunque, si ha

$$\lim_n G_{X_n}(t) = e^{\lambda(t-1)} .$$

Dal momento che la precedente espressione fornisce la f.g. di una v.a. X con legge di Poisson $\mathcal{P}(\lambda)$, si ha $X_n \xrightarrow{\mathcal{L}} X$ per $n \to \infty$. Questo risultato è in effetti già noto dalla Sezione 6.3. □

Teorema 8.1.4. (Teorema della trasformata continua) *Sia $(X_n)_{n\geq 1}$ una successione di v.a. e X una v.a. tale che $X_n \xrightarrow{\mathcal{L}} X$ per $n \to \infty$. Se $g : \mathbb{R} \to \mathbb{R}$ è una funzione continua, allora si ha $g(X_n) \xrightarrow{\mathcal{L}} g(X)$ per $n \to \infty$.*

Dimostrazione. Posto $Y_n = g(X_n)$ e $Y = g(X)$, tenendo presente il Lemma di Fatou (Teorema A.7), si ha

$$\lim_n \varphi_{Y_n}(t) = \lim_n \mathrm{E}[e^{\mathrm{i}tY_n}] = \mathrm{E}[\lim_n e^{\mathrm{i}tY_n}] = \mathrm{E}[e^{\mathrm{i}tY}] = \varphi_Y(t)$$

e la tesi segue dal Teorema 8.1.2. □

• **Esempio 8.1.9.** Analogamente all'Esempio 8.1.4, si consideri la successione di v.a. standardizzate $(X_n)_{n\geq 1}$ dove $X_n = \frac{Z_n - n}{\sqrt{2n}}$, mentre la v.a. Z_n si distribuisce con legge Chi-quadrato χ^2_n. Nel medesimo Esempio è stato verificato che $X_n \xrightarrow{\mathcal{L}} X$ per $n \to \infty$, dove la v.a. X si distribuisce con legge Normale ridotta $\mathcal{N}(0, 1)$. Tenendo presente quanto detto nella Sezione 6.8, se si considera la funzione continua $g : x \mapsto x^2$, allora risulta $X_n^2 \xrightarrow{\mathcal{L}} Y$ per $n \to \infty$, dove la v.a. trasformata $Y = X^2$ si distribuisce con legge Chi-quadrato χ^2_1. □

8.2. Convergenza in probabilità

Un secondo concetto di convergenza, detto *convergenza in probabilità*, viene introdotto nella seguente definizione.

Definizione 8.2.1. Se $(X_n)_{n\geq 1}$ è una successione di v.a. definite sullo spazio probabilizzato $(\Omega, \mathcal{F}, P)$, allora si dice che la successione *converge in probabilità* alla v.a. X, definita sul medesimo spazio probabilizzato, se per ogni $\varepsilon > 0$ si ha

$$\lim_n P(|X_n - X| > \varepsilon) = 0$$

e si scrive $X_n \xrightarrow{P} X$ per $n \to \infty$. □

In pratica, la definizione di convergenza in probabilità richiede che la probabilità dell'evento che si realizza quando le v.a. X_n e X differiscono tenda a 0 per $n \to \infty$. Nel caso in cui la v.a. X sia degenere e tale che $P(X = c) = 1$, allora la condizione della Definizione 8.2.1 si riduce a

$$\lim_n P(|X_n - c| > \varepsilon) = 0$$

e con un abuso in notazione quasi universalmente adottato nei testi di Teoria della Probabilità, la convergenza in probabilità viene denotata con $X_n \xrightarrow{P} c$ per $n \to \infty$. Infine, si osservi che nel linguaggio della Teoria della Misura questo tipo di convergenza è in effetti la cosiddetta convergenza in misura.

• **Esempio 8.2.1.** Data la v.a. X con legge Normale ridotta $\mathcal{N}(0, 1)$, si consideri la successione di v.a. $(X_n)_{n\geq 1}$ tale che $X_n = (1 - n^{-1})X$. Si vuole verificare che la successione di v.a. converge in probabilità alla v.a. X. Dal momento che

$$\begin{aligned} P(|X_n - X| > \varepsilon) &= 1 - P(|X_n - X| \leq \varepsilon) = 1 - P(|X| \leq n\varepsilon) \\ &= 1 - \int_{-n\varepsilon}^{n\varepsilon} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}\, dz = 2 - 2\Phi(n\varepsilon)\,, \end{aligned}$$

allora si ha

$$\lim_n P(|X_n - X| > \varepsilon) = 2 - 2\lim_n \Phi(n\varepsilon) = 0$$

e quindi $X_n \xrightarrow{P} X$ per $n \to \infty$. □

• **Esempio 8.2.2.** Data la v.a. Z con legge ridotta di Cauchy (si veda l'Esempio 4.1.4), si consideri la successione di v.a. $(X_n)_{n\geq 1}$ tale che $X_n = n^{-1}Z$. Si vuole verificare che la successione di v.a. converge in probabilità ad una v.a. degenere concentrata su 0. In effetti, si ha

$$\begin{aligned} P(|X_n| > \varepsilon) &= 1 - P(|X_n| \leq \varepsilon) = 1 - P(|Z| \leq n\varepsilon) \\ &= 1 - \int_{-n\varepsilon}^{n\varepsilon} \frac{1}{\pi(1 + z^2)}\, dz = 1 - \frac{2}{\pi}\arctan(n\varepsilon)\,, \end{aligned}$$

da cui

$$\lim_n P(|X_n| > \varepsilon) = 1 - \frac{2}{\pi}\lim_n \arctan(n\varepsilon) = 0$$

e quindi $X_n \xrightarrow{P} 0$ per $n \to \infty$. □

La seguente Proposizione fornisce la condizione sufficiente per la convergenza in probabilità ad una v.a. degenere.

Proposizione 8.2.2. *Sia $(X_n)_{n\geq 1}$ una successione di v.a. definite sullo spazio probabilizzato $(\Omega, \mathcal{F}, P)$. Se esiste un $\delta > 0$ per cui si ha*

$$\lim_n \mathrm{E}[|X_n - c|^{\delta}] = 0\,,$$

allora $X_n \xrightarrow{P} c$ *per* $n \to \infty$.

Dimostrazione. Dalla disuguaglianza di Markov (Teorema 4.3.6), per $\varepsilon > 0$ si ha

$$P(|X_n - c| > \varepsilon) = P(|X_n - c|^\delta > \varepsilon^\delta) \leq \frac{\mathrm{E}[|X_n - c|^\delta]}{\varepsilon^\delta}$$

e quindi dall'ipotesi fatta risulta $\lim_n P(|X_n - c| > \varepsilon) = 0$. □

Dal momento che $\mathrm{E}[(X_n - c)^2] = \mathrm{Var}[X_n] + (\mathrm{E}[X_n] - c)^2$, si osservi che per $\delta = 2$ la condizione data nella Proposizione 8.2.2 è equivalente alle condizioni $\lim_n \mathrm{E}[X_n] = c$ e $\lim_n \mathrm{Var}[X_n] = 0$.

• **Esempio 8.2.3.** (Legge dei Grandi Numeri di Bernoulli) Si consideri la successione di v.a. indipendenti $(X_n)_{n\geq 1}$ tale che la v.a. X_n si distribuisce con legge Binomiale $\mathcal{B}(1, p)$ per ogni n. Inoltre, si consideri la successione di v.a. $(Z_n)_{n\geq 1}$, dove $Z_n = n^{-1}S_n$ e $S_n = \sum_{k=1}^n X_k$. Essendo

$$\mathrm{E}[Z_n] = \mathrm{E}[X_n] = p\,,$$

allora

$$\mathrm{E}[(Z_n - p)^2] = \mathrm{Var}[Z_n] = \frac{1}{n}\,\mathrm{Var}[X_n] = \frac{pq}{n}\,,$$

da cui $\lim_n \mathrm{E}[(Z_n - p)^2] = 0$. Dunque, si può concludere che $Z_n \xrightarrow{P} p$ per $n \to \infty$. Dal momento che $n^{-1}S_n$ rappresenta la percentuale di successi in n esperimenti indipendenti dicotomici con probabilità di successo pari a p (ovvero quando si considera lo schema delle prove ripetute, vedi Esempio 1.5.2), allora il precedente risultato stabilisce che la percentuale di successi converge alla probabilità di successo nel singolo esperimento. Questa particolare convergenza in probabilità costituisce in effetti la Legge Debole dei Grandi Numeri di Jakob Bernoulli e sarà analizzata nella sua forma generale in questa Sezione. □

La condizione della Proposizione 8.2.2 è sufficiente, ma non necessaria, per la convergenza in probabilità verso una v.a. degenere. In effetti, tale convergenza si può verificare senza che $\mathrm{E}[(X_n - c)^2]$ converga verso zero. Ad esempio, questo può accadere con successioni di v.a. che non possiedono primo o secondo momento finito.

• **Esempio 8.2.4.** Si consideri la successione di v.a. discrete $(X_n)_{n\geq 1}$, tale che la v.a. X_n è a valori sull'insieme $S = \{0, n\}$ con legge essenziale $P(X_n = 0) = 1 - n^{-1}$ e $P(X_n = n) = n^{-1}$. La successione di v.a. $(X_n)_{n\geq 1}$ converge in probabilità ad una v.a. degenere concentrata sullo 0. In effetti, per $\varepsilon \geq n$ risulta $P(|X_n| > \varepsilon) = 0$, mentre per $\varepsilon < n$ si ha

$$P(|X_n| > \varepsilon) = 1 - P(X_n = 0) = \frac{1}{n}\,.$$

Dunque, si ha $\lim_n P(|X_n| > \varepsilon) = 0$, da cui $X_n \xrightarrow{P} 0$ per $n \to \infty$. Tuttavia, è immediato verificare che $\mathrm{E}[X_n^2] = n$ e quindi $\lim_n \mathrm{E}[X_n^2] = \infty$. Inoltre, si ha $\mathrm{E}[X_n] = 1$, da cui ovviamente segue che

$$\lim_n \mathrm{E}[X_n] = 1\,,$$

ovvero la successione delle medie $(\mathrm{E}[X_n])_{n\geq 1}$ non converge verso 0. □

• **Esempio 8.2.5.** (Legge di Pareto) Si consideri la v.a. assolutamente continua Z che ammette d.p. data da

$$f_Z(x) = \alpha z^{-\alpha-1}\mathbf{1}_{[1,\infty[}(z)\,,$$

dove $\alpha \in]0, \infty[$. Si consideri inoltre la successione di v.a. $(X_n)_{n\geq 1}$, dove $X_n = n^{-1}Z$. Si vuole verificare che per ogni α la successione di v.a. converge in probabilità ad una v.a. degenere concentrata sullo 0. In effetti, si ha

$$P(|X_n| > \varepsilon) = 1 - P(|X_n| \leq \varepsilon) = 1 - P(|Z| \leq n\varepsilon) = 1 - \int_1^{n\varepsilon} \alpha z^{-\alpha-1}\, dz = (n\varepsilon)^{-\alpha}\,,$$

da cui

$$\lim_n P(|X_n| > \varepsilon) = \lim_n (n\varepsilon)^{-\alpha} = 0$$

e quindi $X_n \xrightarrow{P} 0$ per $n \to \infty$. Tuttavia, per $\alpha \in]0, 2]$ risulta $\lim_n \mathrm{E}[X_n^2] = \infty$. La legge è stata introdotta dall'economista Vilfredo Pareto (1848-1923). □

Le prossime Proposizioni riguardano il legame esistente tra la convergenza in probabilità e la convergenza in distribuzione. Più esattamente, si evidenzia che la convergenza in probabilità implica la convergenza in distribuzione, mentre la proposizione inversa è vera quando si ha la convergenza in distribuzione verso una v.a. degenere.

Proposizione 8.2.3. *Date una v.a. X e una successione di v.a. $(X_n)_{n\geq 1}$ definite sullo stesso spazio probabilizzato $(\Omega, \mathcal{F}, P)$, se $X_n \xrightarrow{P} X$ per $n \to \infty$ allora si ha $X_n \xrightarrow{\mathcal{L}} X$ per $n \to \infty$.*

Dimostrazione. Se $\varepsilon > 0$, allora risulta

$$\begin{aligned} F_{X_n}(x) = P(X_n \leq x) &= P(\{X_n \leq x\} \cap \{|X_n - X| \leq \varepsilon\}) + P(\{X_n \leq x\} \cap \{|X_n - X| > \varepsilon\}) \\ &\leq P(\{X \leq x + \varepsilon\} \cap \{|X_n - X| \leq \varepsilon\}) + P(|X_n - X| > \varepsilon) \\ &\leq P(X \leq x + \varepsilon) + P(|X_n - X| > \varepsilon)\,. \end{aligned}$$

Dalla definizione di convergenza in probabilità si ha $\lim_n P(|X_n - X| > \varepsilon) = 0$ e dunque

$$\limsup_n F_{X_n}(x) \leq F_X(x + \varepsilon)\,.$$

In modo analogo si ottiene

$$\liminf_n F_{X_n}(x) \geq F_X(x - \varepsilon)\,.$$

Dal momento che ε è arbitrario, per ogni x per cui F_X è continua si ha

$$F_X(x) = \lim_{\varepsilon \to 0^+} F_X(x - \varepsilon) \leq \liminf_n F_{X_n}(x) \leq \limsup_n F_{X_n}(x) \leq \lim_{\varepsilon \to 0^+} F_X(x + \varepsilon) = F_X(x)\,,$$

da cui segue $\lim_n F_{X_n}(x) = F_X(x)$, ovvero $X_n \xrightarrow{\mathcal{L}} X$ per $n \to \infty$. □

Proposizione 8.2.4. *Data una successione di v.a. $(X_n)_{n\geq 1}$ definite sullo spazio probabilizzato $(\Omega, \mathcal{F}, P)$, allora $X_n \xrightarrow{P} c$ per $n \to \infty$ se e solo se $X_n \xrightarrow{\mathcal{L}} c$ per $n \to \infty$.*

Dimostrazione. Sulla base della Proposizione 8.2.3 si deve dimostrare solo l'implicazione inversa. Se $\varepsilon > 0$, si ha

$$\begin{aligned} P(|X_n - c| > \varepsilon) &= 1 - P(|X_n - c| \leq \varepsilon) = 1 - F_{X_n}(c + \varepsilon) + F_{X_n}(c - \varepsilon) - P(X_n = c - \varepsilon) \\ &\leq 1 - F_{X_n}(c + \varepsilon) + F_{X_n}(c - \varepsilon)\,. \end{aligned}$$

Dal momento che $X_n \xrightarrow{\mathcal{L}} c$ per $n \to \infty$, ovvero $\lim_n F_{X_n}(x) = \mathbf{1}_{[c,\infty[}(x)$ per ogni $x \neq c$, allora si ha $\lim_n F_{X_n}(c + \varepsilon) = 1$ e $\lim_n F_{X_n}(c - \varepsilon) = 0$. Dunque, risulta

$$\lim_n (1 - F_{X_n}(c+\varepsilon) + F_{X_n}(c-\varepsilon)) = 0\,,$$

da cui segue che $X_n \xrightarrow{P} c$ per $n \to \infty$. □

• **Esempio 8.2.6.** Analogamente all'Esempio 8.1.5, si consideri la successione di v.a. $(Z_n)_{n\geq 1}$, tale che la v.a. Z_n si distribuisce con legge Chi-quadrato χ^2_n, e la successione di v.a. $(X_n)_{n\geq 1}$, dove $X_n = n^{-1}Z_n$. Nel medesimo esempio è stato verificato che $X_n \xrightarrow{\mathcal{L}} 1$ per $n \to \infty$ e dunque dalla Proposizione 8.2.4 si ha anche $X_n \xrightarrow{P} 1$ per $n \to \infty$. □

Si introducono di seguito alcuni risultati sulla convergenza per successioni di trasformate di v.a. In particolare, viene dato il cosiddetto Teorema di Cramér-Slutsky, che prende nome dallo statistico matematico svedese Harald Cramér (1893-1985) e dal probabilista ed economista russo Evgeny Evgenievich Slutsky (1880-1948).

Teorema 8.2.5. (Teorema della trasformata continua) *Date una v.a. X e una successione di v.a. $(X_n)_{n\geq 1}$ definite sullo stesso spazio probabilizzato $(\Omega, \mathcal{F}, P)$, se $X_n \xrightarrow{P} X$ per $n \to \infty$ e $g : \mathbb{R} \to \mathbb{R}$ è una funzione continua, allora si ha $g(X_n) \xrightarrow{P} g(X)$ per $n \to \infty$.*

Dimostrazione. Sia $c > 0$ una costante per cui si ha $P(|X| > c) < \frac{1}{2}\delta$ con $\delta > 0$. Inoltre, dal momento che g è continua, allora per n abbastanza elevato esiste un $\delta = \delta(\varepsilon)$ tale che

$$\begin{aligned} P(|g(X_n) - g(X)| > \varepsilon) &= P(\{|g(X_n) - g(X)| > \varepsilon\} \cap \{|X| \leq c\}) \\ &\quad + P(\{|g(X_n) - g(X)| > \varepsilon\} \cap \{|X| > c\}) \\ &\leq P(|X_n - X| > \delta) + P(|X| > c) \leq \frac{\delta}{2} + \frac{\delta}{2} = \delta\,, \end{aligned}$$

da cui segue che $g(X_n) \xrightarrow{P} g(X)$ per $n \to \infty$. □

Teorema 8.2.6. (Teorema di Cramér-Slutsky) *Si considerino due successioni di v.a. $(X_n)_{n\geq 1}$ e $(Y_n)_{n\geq 1}$ definite sullo stesso spazio probabilizzato $(\Omega, \mathcal{F}, P)$ e la v.a. X definita sul medesimo spazio probabilizzato. Se $X_n \xrightarrow{\mathcal{L}} X$ e $Y_n \xrightarrow{P} c$ per $n \to \infty$, allora si ha*

$$X_n + Y_n \xrightarrow{\mathcal{L}} X + c\,,$$

$$X_n - Y_n \xrightarrow{\mathcal{L}} X - c\,,$$

$$X_n Y_n \xrightarrow{\mathcal{L}} cX\,,$$

$$\frac{X_n}{Y_n} \xrightarrow{\mathcal{L}} \frac{X}{c}\,,\ c \neq 0\,,$$

per $n \to \infty$.

Dimostrazione. Si dimostra la prima relazione. Per $\varepsilon > 0$, si ha

$$\begin{aligned} F_{X_n+Y_n}(x) &= P(X_n + Y_n \leq x) \\ &= P(\{X_n + Y_n \leq x\} \cap \{|Y_n - c| \leq \varepsilon\}) + P(\{X_n + Y_n \leq x\} \cap \{|Y_n - c| > \varepsilon\}) \\ &\leq P(\{X_n \leq x - c + \varepsilon\} \cap \{|Y_n - c| \leq \varepsilon\}) + P(|Y_n - c| > \varepsilon) \\ &\leq P(X_n \leq x - c + \varepsilon) + P(|Y_n - c| > \varepsilon)\,. \end{aligned}$$

Dalla definizione di convergenza in probabilità si ha $\lim_n P(|Y_n - c| > \varepsilon) = 0$ e dal momento che $X_n \xrightarrow{\mathcal{L}} X$, se $(x - c + \varepsilon)$ è un punto di continuità per F_X, si ha

$$\limsup_n F_{X_n+Y_n}(x) \leq F_X(x - c + \varepsilon) .$$

In modo analogo, se $(x - c - \varepsilon)$ è un punto di continuità per F_X, si ottiene

$$\liminf_n F_{X_n+Y_n}(x) \geq F_X(x - c - \varepsilon) .$$

Dal momento che ε può essere scelto in modo arbitrario e che l'insieme dei punti di discontinuità di F_X è al più numerabile (Proposizione 3.2.6), per ogni $(x - c)$ per cui F_X è continua si ha

$$\lim_n F_{X_n+Y_n}(x) = F_X(x - c) ,$$

ovvero $X_n + Y_n \xrightarrow{\mathcal{L}} X + c$ per $n \to \infty$. Le altre relazioni si verificano in modo analogo. □

• **Esempio 8.2.7.** Si consideri la successione di v.a. (Z_n), dove la v.a. Z_n si distribuisce con legge t di Student con n gradi di libertà. Tenendo presente la Sezione 6.10, la v.a. Z_n può essere rappresentata come $Z_n = \frac{X_n}{\sqrt{n^{-1}Y_n}}$, dove la v.a. X_n si distribuisce con legge Normale ridotta $\mathcal{N}(0, 1)$, mentre la v.a. Y_n si distribuisce con legge Chi-quadrato χ^2_n. Dall'Esempio 8.2.6 si ha che $n^{-1}Y_n \xrightarrow{P} 1$ per $n \to \infty$. Inoltre, dal momento che la funzione $g : x \mapsto \sqrt{x}$ è continua, dal Teorema 8.2.5 si ha $\sqrt{n^{-1}Y_n} \xrightarrow{P} 1$ per $n \to \infty$. Infine, tenendo presente che banalmente si ha $X_n \xrightarrow{\mathcal{L}} Z$, dove la v.a. Z si distribuisce con legge Normale ridotta $\mathcal{N}(0, 1)$, dal Teorema 8.2.6 si ha infine che $Z_n \xrightarrow{\mathcal{L}} Z$ per $n \to \infty$. □

Il seguente Teorema fornisce un celebre risultato, ovvero la cosiddetta Legge Debole dei Grandi Numeri, nella sua forma più comune.

Teorema 8.2.7. (Legge Debole dei Grandi Numeri) *Sia $(X_n)_{n\geq 1}$ una successione di v.a. non correlate definite sullo spazio probabilizzato $(\Omega, \mathcal{F}, P)$, tali che per ogni n si ha $\mathrm{E}[X_n] = \mu$ e $\mathrm{Var}[X_n] < c < \infty$. Data l'ulteriore successione di v.a. $(Z_n)_{n\geq 1}$, dove $Z_n = n^{-1}S_n$ e $S_n = \sum_{k=1}^n X_k$, allora si ha $Z_n \xrightarrow{P} \mu$ per $n \to \infty$.*

Dimostrazione. Dal momento che

$$\mathrm{E}[Z_n] = \frac{1}{n}\sum_{k=1}^n \mathrm{E}[X_k] = \mu$$

e, dal momento che le v.a. della successione sono non correlate, si ha

$$\mathrm{Var}[Z_n] = \frac{1}{n^2}\sum_{k=1}^n \mathrm{Var}[X_k] < \frac{c}{n} .$$

Inoltre, dalla disuguaglianza di Chebyshev (Teorema 4.3.7), per $\varepsilon > 0$ si ha

$$P(|Z_n - \mu| > \varepsilon) \leq \frac{\mathrm{Var}[Z_n]}{\varepsilon^2} < \frac{c}{n\varepsilon^2} .$$

Dunque, risulta

$$\lim_n P(|Z_n - \mu| > \varepsilon) = 0 ,$$

ovvero $Z_n \xrightarrow{P} \mu$ per $n \to \infty$. □

Un'ulteriore versione fondamentale della Legge Debole dei Grandi Numeri è stata ottenuta in un approccio generale dal probabilista russo Aleksandr Yakovlevich Khinchin (1894-1959) e viene data di seguito.

Teorema 8.2.8. (Legge Debole dei Grandi Numeri di Khinchin) *Sia $(X_n)_{n\geq 1}$ una successione di v.a. indipendenti con la medesima legge definite sullo spazio probabilizzato $(\Omega, \mathcal{F}, P)$, tali che per ogni n si ha $\mathrm{E}[X_n] = \mu$. Data l'ulteriore successione di v.a. $(Z_n)_{n\geq 1}$, dove $Z_n = n^{-1}S_n$ e $S_n = \sum_{k=1}^{n} X_k$, allora si ha $Z_n \xrightarrow{P} \mu$ per $n \to \infty$.*

Dimostrazione. Dalla Proposizione 7.1.7 si ha

$$\varphi_{X_n}(t) = 1 + \mathrm{i}\mu t + o(t) \, .$$

Inoltre, dalle Proposizioni 7.3.6 e 7.1.5 si ha

$$\varphi_{Z_n}(t) = \left(1 + \frac{\mathrm{i}\mu t}{n} + o(n^{-1})\right)^n \, ,$$

da cui

$$\lim_n \varphi_{X_n}(t) = e^{\mathrm{i}\mu t} \, .$$

Dunque, tenendo presente il Teorema 8.1.2 si ha $Z_n \xrightarrow{\mathcal{L}} \mu$ per $n \to \infty$, ovvero dalla Proposizione 8.2.4 si ottiene infine che $Z_n \xrightarrow{P} \mu$ per $n \to \infty$. □

8.3. Convergenza quasi certa

Un terzo concetto di convergenza, detto *convergenza quasi certa*, viene introdotto nella seguente definizione.

Definizione 8.3.1. Se $(X_n)_{n\geq 1}$ è una successione di v.a. definite sullo spazio probabilizzato $(\Omega, \mathcal{F}, P)$, allora si dice che la successione *converge quasi certamente* alla v.a. X, definita sul medesimo spazio probabilizzato, se l'evento

$$\{\omega \in \Omega : \lim_n X_n(\omega) = X(\omega)\} = \{\lim_n X_n = X\}$$

si verifica $q.c.$, ovvero se

$$P(\lim_n X_n = X) = 1 \, ,$$

e si scrive $X_n \xrightarrow{q.c.} X$ per $n \to \infty$. □

Nel caso in cui la v.a. X sia degenere e tale che $P(X = c) = 1$, allora la condizione della Definizione 8.3.1 si riduce a

$$P(\lim_n X_n = c) = 1 \, ,$$

e con il solito abuso in notazione, la convergenza quasi certa viene denotata con $X_n \xrightarrow{q.c.} c$ per $n \to \infty$. Nel linguaggio della Teoria della Misura questo tipo di convergenza è in effetti la cosiddetta convergenza puntuale. Dal momento che la misura dello spazio fondamentale è finita, in Teoria della Probabilità la convergenza puntuale coincide con la convergenza uniforme.

Risulta inoltre interessante notare che, dalla definizione di limite, l'evento $\{\lim_n X_n = X\}$ si verifica se per ogni intero m esiste un intero n tale che per ogni $k \geq n$ si verifica l'evento

$\{|X_k - X| \le m^{-1}\}$. Dunque, anche tenendo presente la definizione di limite superiore di una successione di eventi, si ha

$$\{\lim_n X_n = X\} = \bigcap_{m=1}^{\infty} \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} \{|X_k - X| \le m^{-1}\} = \lim_m \limsup_n \{|X_n - X| \le m^{-1}\} ,$$

ovvero, dal momento che $(\limsup_n \{|X_n - X| \le m^{-1}\})_{m\ge 1}$ è una successione decrescente di eventi e in base al Teorema 2.2.10, la condizione della Definizione 8.3.1 è equivalente a

$$\begin{aligned} P(\lim_n X_n = X) &= P\left(\bigcap_{m=1}^{\infty} \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} \{|X_k - X| \le m^{-1}\} \right) \\ &= \lim_m P\big(\limsup_n \{|X_n - X| \le m^{-1}\}\big) = 1 . \end{aligned}$$

La seguente Proposizione stabilisce una condizione necessaria e sufficiente affinchè si verifichi la convergenza quasi certa.

Proposizione 8.3.2. *Date una v.a. X e una successione di v.a. $(X_n)_{n\ge 1}$ definite sullo stesso spazio probabilizzato $(\Omega, \mathcal{F}, P)$, allora $X_n \overset{q.c.}{\to} X$ per $n \to \infty$ se e solo se per ogni $\varepsilon > 0$ si ha*

$$\lim_n P\left(\bigcap_{k=n}^{\infty} \{|X_k - X| \le \varepsilon\} \right) = 1 ,$$

o equivalentemente che

$$\lim_n P\left(\bigcup_{k=n}^{\infty} \{|X_k - X| > \varepsilon\} \right) = 0 .$$

Dimostrazione. Posto

$$F_{n,m} = \bigcap_{k=n}^{\infty} \{|X_k - X| \le m^{-1}\} ,$$

si osservi che la successione di eventi $(F_{n,m})_{n\ge 1}$ è crescente per ogni intero m. Inoltre, la condizione della Definizione 8.3.1 è verificata se e solo se per ogni intero m si ha

$$P\left(\bigcup_{n=1}^{\infty} F_{n,m} \right) = 1 .$$

Tenendo presente la definizione di limite di successione crescente di eventi, dal Teorema 2.2.10 si ottiene inoltre

$$P\left(\bigcup_{n=1}^{\infty} F_{n,m} \right) = P(\lim_n F_{n,m}) = \lim_n P(F_{n,m}) ,$$

da cui segue la prima parte. Inoltre, dalla relazione di De Morgan si ha

$$F_{n,m} = \left(\bigcup_{k=n}^{\infty} \{|X_k - X| > m^{-1}\} \right)^c ,$$

da cui segue la seconda parte. □

• **Esempio 8.3.1.** Si consideri la successione di v.a. $(X_n)_{n\geq 1}$ dell'Esempio 8.2.1. Dal momento che si ha

$$P\left(\bigcup_{k=n}^{\infty}\{|X_k - X| > \varepsilon\}\right) = P\left(\bigcup_{k=n}^{\infty}\{|X| > k\varepsilon\}\right) = P(|X| > n\varepsilon) = 2 - 2\Phi(n\varepsilon)$$

allora risulta

$$\lim_n P\left(\bigcup_{k=n}^{\infty}\{|X_k - X| > \varepsilon\}\right) = 2 - 2\lim_n \Phi(n\varepsilon) = 0$$

e quindi $X_n \overset{q.c.}{\to} X$ per $n \to \infty$. □

• **Esempio 8.3.2.** Si consideri la successione di v.a. $(X_n)_{n\geq 1}$ dell'Esempio 8.2.2. Dal momento che si ha

$$P\left(\bigcup_{k=n}^{\infty}\{|X_k| > \varepsilon\}\right) = P\left(\bigcup_{k=n}^{\infty}\{|Z| > k\varepsilon\}\right) = P(|Z| > n\varepsilon) = 1 - \frac{2}{\pi}\arctan(n\varepsilon)\,,$$

allora risulta

$$\lim_n P\left(\bigcup_{k=n}^{\infty}\{|X_k| > \varepsilon\}\right) = 1 - \frac{2}{\pi}\lim_n \arctan(n\varepsilon) = 0$$

e quindi $X_n \overset{q.c.}{\to} 0$ per $n \to \infty$. □

Dalla Proposizione 8.3.2 si intuisce come la convergenza quasi certa implichi una condizione più forte di quella richiesta nella convergenza in probabilità. In effetti, vale la seguente Proposizione.

Proposizione 8.3.3. *Date una v.a. X e una successione di v.a. $(X_n)_{n\geq 1}$ definite sullo stesso spazio probabilizzato $(\Omega, \mathcal{F}, P)$, se $X_n \overset{q.c.}{\to} X$ per $n \to \infty$ allora si ha $X_n \overset{P}{\to} X$ per $n \to \infty$.*

Dimostrazione. Dato $\varepsilon > 0$, per ogni n si ha

$$P(|X_n - X| > \varepsilon) \leq P\left(\bigcup_{k=n}^{\infty}\{|X_k - X| > \varepsilon\}\right).$$

Dunque, se $X_n \overset{q.c.}{\to} X$ per $n \to \infty$, dalla Proposizione 8.3.2 e dalla precedente relazione si ha $\lim_n P(|X_n - X| > \varepsilon) = 0$, ovvero $X_n \overset{P}{\to} X$ per $n \to \infty$. □

• **Esempio 8.3.3.** Si consideri la successione di v.a. discrete $(X_n)_{n\geq 1}$ dell'Esempio 8.2.4, assumendo l'indipendenza delle v.a. Evidentemente, anche in questo caso si ha $X_n \overset{P}{\to} 0$ per $n \to \infty$. Dunque, in base alla Proposizione 8.3.3, se vi è convergenza quasi certa la successione deve necessariamente convergere verso una v.a. degenere concentrata sullo 0. Tuttavia, per ogni $\epsilon > 0$ si ha $P(|X_n| > \varepsilon) = n^{-1}$, e dal momento che la serie armonica non converge risulta

$$P\left(\bigcup_{n=1}^{\infty}\{|X_n| > \varepsilon\}\right) = \sum_{n=1}^{\infty} P(|X_n| > \varepsilon) = \sum_{n=1}^{\infty}\frac{1}{n} = \infty\,.$$

Dunque, dal Lemma di Borel-Cantelli (Teorema 2.7.1) si ottiene che $P(\limsup_n |X_n| > \varepsilon) = 1$, ovvero la successione non può converge quasi certamente a 0. □

La seguente Proposizione fornisce condizioni sufficienti per la convergenza quasi certa verso una v.a. degenere.

Proposizione 8.3.4. *Data una successione di v.a.* $(X_n)_{n\geq 1}$ *definite sullo stesso spazio probabilizzato* $(\Omega, \mathcal{F}, P)$*, allora* $X_n \overset{q.c.}{\to} c$ *per* $n \to \infty$ *se per ogni* ε *si ha*

$$\sum_{n=1}^{\infty} P(|X_n - c| > \varepsilon) < \infty \,.$$

Alternativamente, $X_n \overset{q.c.}{\to} c$ *per* $n \to \infty$ *se esiste un* $\delta > 0$ *per cui si ha*

$$\sum_{n=1}^{\infty} \mathrm{E}[|X_n - c|^{\delta}] < \infty \,.$$

Dimostrazione. Dalla disuguaglianza di Bonferroni (Teorema 2.2.8), segue che

$$P\left(\bigcup_{k=n}^{\infty} \{|X_k - c| > \varepsilon\}\right) \leq \sum_{k=n}^{\infty} P(|X_k - c| > \varepsilon) \,.$$

Dal momento che l'ipotesi di convergenza della serie $\sum_{n=1}^{\infty} P(|X_n - c| > \varepsilon)$ implica

$$\lim_n \sum_{k=n}^{\infty} P(|X_k - c| > \varepsilon) = 0 \,,$$

segue la prima parte. Tenendo presente la disuguaglianza di Markov (Teorema 4.3.6), si ha inoltre

$$\sum_{k=n}^{\infty} P(|X_k - c| > \varepsilon) = \sum_{k=n}^{\infty} P(|X_k - c|^{\delta} > \varepsilon^{\delta}) \leq \frac{1}{\varepsilon^{\delta}} \sum_{k=n}^{\infty} \mathrm{E}[|X_k - c|^{\delta}] \,.$$

Dal momento che l'ipotesi di convergenza della serie $\sum_{n=1}^{\infty} \mathrm{E}[|X_n - c|^{\delta}]$ implica

$$\lim_n \sum_{k=n}^{\infty} \mathrm{E}[|X_k - c|^{\delta}] = 0 \,,$$

allora si ha la seconda parte. □

• **Esempio 8.3.4.** In modo simile all'Esempio 8.2.3, si consideri la successione di v.a. indipendenti $(X_n)_{n\geq 1}$ tale che la v.a. X_n si distribuisce con legge Binomiale $\mathcal{B}(1, p)$ per ogni n. Inoltre, si consideri l'ulteriore successione di v.a. $(Z_n)_{n\geq 1}$, dove $Z_n = n^{-1} S_n$ e $S_n = \sum_{k=1}^{n} X_k$. Tenendo presente la seconda condizione della Proposizione 8.3.4, posto $\delta = 4$, dal momento che S_n si distribuisce con legge Binomiale $\mathcal{B}(n, p)$ si ha

$$\mathrm{E}[(Z_n - p)^4] = \frac{1}{n^4} \mathrm{E}[(S_n - np)^4] = \frac{3p^2q^2}{n^2} + \frac{pq(1 - 6pq)}{n^3} \,,$$

dove l'ultima relazione è stata ottenuta attraverso l'applicazione laboriosa, ma ovvia, della Proposizione 7.1.6. Dunque, dalla definizione della funzione Zeta di Riemann (si veda Esempio 7.2.7), risulta

$$\sum_{n=1}^{\infty} \mathrm{E}[(Z_n - p)^4] = 3p^2q^2\zeta(2) + pq(1 - 6pq)\zeta(3) < \infty \,,$$

dove $\zeta(2) = \frac{\pi^2}{6}$ e $\zeta(3) < \zeta(4) = \frac{\pi^4}{90}$. Dunque, si può concludere che $Z_n \overset{q.c.}{\to} p$ per $n \to \infty$. Questa particolare convergenza quasi certa costituisce un caso particolare della Legge Forte dei Grandi Numeri. □

Il seguente Teorema fornisce la cosiddetta Legge Forte dei Grandi Numeri, nella forma introdotta da Andrej Kolmogorov.

Teorema 8.3.5. (Legge Forte dei Grandi Numeri di Kolmogorov) *Sia $(X_n)_{n\geq 1}$ una successione di v.a. indipendenti con la medesima legge definite sullo spazio probabilizzato $(\Omega, \mathcal{F}, P)$, tali che per ogni n si ha $\mathrm{E}[X_n] = \mu$. Data l'ulteriore successione di v.a. $(Z_n)_{n\geq 1}$, dove $Z_n = n^{-1}S_n$ e $S_n = \sum_{k=1}^{n} X_k$, allora si ha $Z_n \overset{q.c.}{\to} \mu$ per $n \to \infty$.*

Dimostrazione. Si veda Billingsley (1995, p.85). □

8.4. Convergenza in media

Un ulteriore concetto di convergenza, che è usualmente detto *convergenza in media di ordine p* (o semplicemente *in media quadratica* quando $p = 2$), viene definito di seguito. La convergenza in media sarà centrale nello sviluppo della teoria dell'integrazione stocastica presentata nel Capitolo 10.

Definizione 8.4.1. Se $(X_n)_{n\geq 1}$ è una successione di v.a. definite sullo spazio probabilizzato $(\Omega, \mathcal{F}, P)$ e se $p \in]0, \infty[$, allora si dice che la successione *converge in media di ordine p* alla v.a. X, definita sul medesimo spazio probabilizzato, se si ha

$$\lim_n \mathrm{E}[|X_n - X|^p] = 0$$

e si scrive $X_n \overset{L^p}{\to} X$ per $n \to \infty$. □

La convergenza in media di ordine p risulta effettivamente equivalente alla convergenza in norma di ordine p considerata nella Teoria della Misura. Questo è il motivo per cui questo tipo di convergenza viene indicata con il simbolo L^p (si veda l'Appendice A). Nel caso in cui la v.a. X sia degenere e tale che $P(X = c) = 1$, allora la condizione della Definizione 8.4.1 si riduce a

$$\lim_n \mathrm{E}[|X_n - c|^p] = 0$$

e si scrive $X_n \overset{L^p}{\to} c$ per $n \to \infty$ con il solito abuso in notazione.

• **Esempio 8.4.1.** Si consideri una successione $(X_n)_{n\geq 1}$ di v.a. indipendenti con la medesima legge e tali che per ogni n si ha $\mathrm{E}[X_n] = \mu$ e $\mathrm{Var}[X_n] = \sigma^2 < \infty$. Si consideri inoltre l'ulteriore successione di v.a. $(Z_n)_{n\geq 1}$ con $Z_n = n^{-1}S_n$ e $S_n = \sum_{k=1}^{n} X_k$. Dal momento che $\mathrm{E}[S_n] = n\mu$ e $\mathrm{Var}[S_n] = n\sigma^2$, allora si ha

$$\lim_n \mathrm{E}[(Z_n - \mu)^2] = \lim_n \frac{\mathrm{E}[(S_n - n\mu)^2]}{n^2} = \lim_n \frac{\mathrm{Var}[S_n]}{n^2} = \lim_n \frac{\sigma^2}{n} = 0\,.$$

Dunque, risulta $Z_n \overset{L^2}{\to} \mu$ per $n \to \infty$. Inoltre, dal momento che si ha

$$\mathrm{E}[|Z_n - \mu|] \leq \mathrm{E}[(Z_n - \mu)^2]^{\frac{1}{2}}$$

per la disuguaglianza di Lyapunov (Teorema 4.3.2), risulta anche $Z_n \overset{L^1}{\to} \mu$ per $n \to \infty$. □

La convergenza in media di ordine p implica la convergenza in probabilità. Questo risultato era già stato in effetti evidenziato nella Proposizione 8.2.2 nel caso di convergenza ad una variabile degenere e viene generalizzato nella seguente Proposizione.

Proposizione 8.4.2. *Date una v.a. X e una successione di v.a. $(X_n)_{n\geq 1}$ definite sullo stesso spazio probabilizzato $(\Omega, \mathcal{F}, P)$, se esiste un $p \in]0,\infty[$ per cui si ha $X_n \xrightarrow{L^p} X$ per $n \to \infty$ allora risulta $X_n \xrightarrow{P} X$ per $n \to \infty$.*

Dimostrazione. Dalla disuguaglianza di Markov (Teorema 4.3.6), per $\varepsilon > 0$ si ha

$$P(|X_n - X| > \varepsilon) = P(|X_n - X|^p > \varepsilon^p) \leq \frac{\mathrm{E}[|X_n - X|^p]}{\varepsilon^p}$$

e, dal momento che dalle ipotesi fatte si ha $\lim_n \mathrm{E}[|X_n - X|^p] = 0$, allora risulta anche $\lim_n P(|X_n - X| > \varepsilon) = 0$, ovvero $X_n \xrightarrow{P} X$ per $n \to \infty$. □

Per quanto riguarda la relazione fra convergenza quasi certa e convergenza in media di ordine p, in generale non si possono fare affermazioni. In effetti, ci sono casi in cui vi è convergenza quasi certa, ma non quella in media di ordine p e viceversa. I seguenti esempi evidenziano queste situazioni.

• **Esempio 8.4.2.** Si consideri una v.a. Z con legge Uniforme su $]0,1[$ e la successione $(X_n)_{n\geq 1}$ dove $X_n = 2^n \mathbf{1}_{[0,\frac{1}{n}[}(Z)$. Evidentemente la v.a. X_n è discreta ed è definita su $\{0, 2^n\}$ con probabilità $P(X_n = 0) = 1 - n^{-1}$ e $P(X_n = 2^n) = n^{-1}$. Tenendo presente la Proposizione 8.3.2 e dal momento che per ogni $k > n$ si ha $\{|X_k| > \varepsilon\} \subset \{|X_n| > \varepsilon\}$ *q.c.*, allora per $\varepsilon > 0$ risulta

$$\lim_n P\left(\bigcup_{k=n}^{\infty} \{|X_k| > \varepsilon\}\right) = \lim_n P(X_n > \varepsilon) = \lim_n \frac{1}{n} = 0\,,$$

ovvero $X_n \xrightarrow{q.c.} 0$. Tuttavia, dal momento che

$$\lim_n \mathrm{E}[|X_n|^p] = \lim_n \frac{2^{np}}{n} = \infty\,,$$

non vi è convergenza in media per nessun valore di p. □

• **Esempio 8.4.3.** Si consideri una v.a. Z con legge Uniforme su $]0,1[$ e sia data inoltre la successione $(X_n)_{n\geq 1}$ dove $X_n = \mathbf{1}_{[j2^{-m},(j+1)2^{-m}[}(Z)$ con $n = 2^m + j$, $j = \{0, 1, \ldots, 2^m - 1\}$ e $m \in \mathbb{N}$. Dunque, $X_1 = \mathbf{1}_{[0,1[}(Z)$, $X_2 = \mathbf{1}_{[0,\frac{1}{2}[}(Z)$, $X_3 = \mathbf{1}_{[\frac{1}{2},1[}(Z)$, $X_4 = \mathbf{1}_{[0,\frac{1}{4}[}(Z)$, $X_5 = \mathbf{1}_{[\frac{1}{4},\frac{1}{2}[}(Z)$, $X_6 = \mathbf{1}_{[\frac{1}{2},\frac{3}{4}[}(Z)$, $X_7 = \mathbf{1}_{[\frac{3}{4},1[}(Z)$ e così via. In questo caso, X_1 è una v.a. degenere tale che $P(X_1 = 1) = 1$, mentre X_n è una v.a. di Bernoulli di parametro $P(X_n = 1) = 2^{-m}$ per $n = 2^m + j > 1$. Dal momento che

$$\lim_n \mathrm{E}[|X_n|^p] = \lim_m 2^{-mp} = 0\,,$$

allora $X_n \xrightarrow{L^p} 0$ per ogni valore di p. Tuttavia, dalla Proposizione 8.3.2, si ha

$$\lim_n P\left(\bigcup_{k=n}^{\infty} \{|X_k| > \varepsilon\}\right) = \lim_n P(X_1 = 1) = 1\,,$$

ovvero la successione non converge quasi certamente a 0. □

La prossima Proposizione fornisce una condizione per la quale la convergenza quasi certa implica la convergenza in media di ordine p.

Proposizione 8.4.3. *Date una v.a. X e una successione di v.a. $(X_n)_{n\geq 1}$ definite sullo stesso spazio probabilizzato $(\Omega, \mathcal{F}, P)$, se $X_n \overset{q.c.}{\to} X$ per $n \to \infty$ e $\{|X_n| \leq Y\}$ $q.c.$ per una v.a. Y tale che $\mathrm{E}[Y^p] < \infty$, allora si ha $X_n \overset{L^p}{\to} X$ per $n \to \infty$.*

Dimostrazione. Dalle assunzioni fatte si ha $\mathrm{E}[|X_n|^p] \leq \mathrm{E}[Y^p] < \infty$. Inoltre, dal momento che $X_n \overset{q.c.}{\to} X$, allora risulta anche $\{|X| \leq Y\}$ $q.c.$ Dal momento che $\{|X_n - X| \leq |X_n| + |X|\}$ $q.c.$, si ha anche $\{|X_n - X| \leq 2Y\}$ $q.c.$ Quindi, per ogni $\varepsilon > 0$ si ottiene

$$\begin{aligned}\mathrm{E}[|X_n - X|^p] &= \mathrm{E}[|X_n - X|^p \, \mathbf{1}_{[0,\varepsilon[}(|X_n - X|)] + \mathrm{E}[|X_n - X|^p \, \mathbf{1}_{[\varepsilon,\infty[}(|X_n - X|)] \\ &\leq \varepsilon^p + 2^p \, \mathrm{E}[Y^p \mathbf{1}_{[\varepsilon,\infty[}(|X_n - X|)] \, .\end{aligned}$$

Tenendo ancora presente che $X_n \overset{q.c.}{\to} X$, allora $\lim_n \mathrm{E}[Y^p \mathbf{1}_{[\varepsilon,\infty[}(|X_n - X|)] = 0$. Dunque, essendo ε arbitrario, dalla precedente relazione segue che $\lim_n \mathrm{E}[|X_n - X|^p] = 0$. □

8.5. Teoremi limite

Si considerano di seguito alcuni teoremi di convergenza che hanno importanti aspetti applicativi in ambiti come la statistica inferenziale. Questi classici risultati sono comunemente denominati come Teoremi Centrali del Limite (dove centrale va inteso nel senso di fondamentale) sulla base di una traduzione leggermente impropria dalla terminologia inglese, ma di uso comune nella letteratura italiana.

Teorema 8.5.1. (Teorema Centrale del Limite di Lindeberg-Lévy) *Sia $(X_n)_{n\geq 1}$ una successione di v.a. indipendenti con la medesima legge e definite sullo spazio probabilizzato $(\Omega, \mathcal{F}, P)$, tali che per ogni n si ha $\mathrm{E}[X_n] = \mu$ e $\mathrm{Var}[X_n] = \sigma^2 < \infty$. Si consideri inoltre la successione di v.a. $(Z_n)_{n\geq 1}$ con*

$$Z_n = \frac{S_n - \mathrm{E}[S_n]}{\mathrm{Var}[S_n]} = \frac{S_n - n\mu}{\sigma\sqrt{n}}$$

e $S_n = \sum_{k=1}^n X_k$. Se la v.a. Z si distribuisce con legge Normale $\mathcal{N}(0,1)$, allora $Z_n \overset{\mathcal{L}}{\to} Z$ per $n \to \infty$.

Dimostrazione. Posto $Y_k = \frac{X_k - \mu}{\sigma}$, si ha $\mathrm{E}[Y_k] = 0$ e $\mathrm{Var}[Y_k] = 1$. Sulla base della Proposizione 7.1.6, per $t \to 0$ la f.c. della v.a. Y_k può essere espressa come

$$\varphi_{Y_k}(t) = 1 + \mathrm{i}t \, \mathrm{E}[Y_k] + \frac{(\mathrm{i}t)^2}{2} \, \mathrm{E}[Y_k^2] + o(t^2) = 1 - \frac{t^2}{2} + o(t^2) \, .$$

Dal momento che

$$Z_n = \frac{1}{\sqrt{n}} \sum_{k=1}^n Y_k \, ,$$

allora si ottiene

$$\varphi_{Z_n}(t) = \varphi_{Y_k}\left(\frac{t}{\sqrt{n}}\right)^n = \left(1 - \frac{t^2}{2n} + o(n^{-1})\right)^n \, .$$

Dunque, si ha

$$\lim_n \varphi_{Z_n}(t) = e^{-\frac{1}{2}t^2} \, ,$$

ovvero dal Teorema di Lévy (Teorema 8.1.2) si ottiene la tesi. □

Questa versione di Teorema Centrale del Limite è stata appunto introdotta da Paul Lévy e dal probabilista finnico Jarl Waldemar Lindeberg (1876-1932). Il Teorema 8.5.1 contiene come caso particolare la versione primitiva del Teorema Centrale del Limite data da Abraham de Moivre, nella quale si assume che $(X_n)_{n\geq 1}$ sia una successione di v.a. indipendenti con la medesima legge e tale che ogni v.a. X_n si distribuisce con legge di Bernoulli $\mathcal{B}(1,p)$. Dal momento che $S_n = \sum_{k=1}^n X_k$ si distribuisce con legge Binomiale $\mathcal{B}(n,p)$ (si veda l'Esempio 7.4.3), il Teorema Centrale del Limite dato di De Moivre fornisce la convergenza in legge della successione di v.a. standardizzate $(Z_n)_{n\geq 1}$, dove $Z_n = \frac{S_n - np}{\sqrt{npq}}$, ad una v.a. con legge $\mathcal{N}(0,1)$ per $n \to \infty$.

La seguente versione di Teorema Centrale del Limite può essere applicata a successioni di v.a. indipendenti con differenti medie e varianze. Il Teorema, oltre che da Paul Lévy e da Jarl Lindeberg, prende nome dal probabilista croato Vilibald (William) Srecko Feller (1906-1970).

Figure 8.5.1. Vilibald (William) Srecko Feller (1906-1970).

Teorema 8.5.2. (Teorema Centrale del Limite di Lindeberg-Lévy-Feller) *Sia $(X_n)_{n\geq 1}$ una successione di v.a. indipendenti definite sullo spazio probabilizzato $(\Omega, \mathcal{F}, P)$, tali che $\mathrm{E}[X_n] = \mu_n$ e $\mathrm{Var}[X_n] = \sigma_n^2 < \infty$ per ogni n. Posto $v_n^2 = \sum_{k=1}^n \sigma_k^2$, si consideri inoltre la successione di v.a. $(Z_n)_{n\geq 1}$ con*

$$Z_n = \frac{1}{v_n} \sum_{k=1}^n (X_k - \mu_k) \, .$$

Se per ogni $\varepsilon > 0$ si ha

$$\lim_n \frac{1}{v_n^2} \sum_{k=1}^n \mathrm{E}[(X_k - \mu_k)^2 \, \mathbf{1}_{]\varepsilon v_n, \infty[}(|X_k - \mu_k|)] = 0 \, ,$$

allora $Z_n \xrightarrow{\mathcal{L}} Z$ per $n \to \infty$, dove la v.a. Z si distribuisce con legge Normale $\mathcal{N}(0,1)$.

Dimostrazione. Si veda Billingsley (1995, p.359). □

La prossima versione di Teorema Centrale del Limite è dovuta a Lyapunov. Nelle medesime ipotesi del Teorema Centrale del Limite dato da Lindeberg-Lévy-Feller, questa versione richiede una condizione più forte, ma spesso più facile da verificare. Si tenga comunque presente che la condizione del Teorema Centrale del Limite dato da Lindeberg-Lévy-Feller è necessaria e sufficiente.

Teorema 8.5.3. (Teorema Centrale del Limite di Lyapunov) *Sia $(X_n)_{n\geq 1}$ una successione di v.a. indipendenti definite sullo spazio probabilizzato $(\Omega, \mathcal{F}, P)$, tali che $\mathrm{E}[X_n] = \mu_n$ e $\mathrm{Var}[X_n] = \sigma_n^2 < \infty$ per ogni n. Posto $v_n^2 = \sum_{k=1}^n \sigma_k^2$, si consideri inoltre la successione di v.a. $(Z_n)_{n\geq 1}$ con*

$$Z_n = \frac{1}{v_n} \sum_{k=1}^{n} (X_k - \mu_k) \, .$$

Se esiste un $\delta > 2$ *tale che* $\mathrm{E}[|X_n|^\delta] < \infty$ *per ogni* n *e che*

$$\lim_n \frac{1}{v_n^{\frac{1}{2}\delta}} \sum_{k=1}^{n} \mathrm{E}[|X_k - \mu_k|^\delta] = 0 \, ,$$

allora $Z_n \xrightarrow{\mathcal{L}} Z$ *per* $n \to \infty$*, dove la v.a.* Z *si distribuisce con legge Normale* $\mathcal{N}(0,1)$.

Dimostrazione. Si veda Billingsley (1995, p.362). □

• **Esempio 8.5.1.** Si consideri la successione di v.a. indipendenti $(X_n)_{n\geq 1}$ tale che la v.a. X_n si distribuisce con legge Binomiale $\mathcal{B}(1, p_n)$. Risulta dunque $v_n^2 = \sum_{k=1}^{n} p_k q_k$. Inoltre, dal momento che $\{|X_k - p_k| \leq 1\}$ $q.c.$, segue che

$$\mathrm{E}[|X_k - p_k|^3] = \mathrm{E}[(X_k - p_k)^2 |X_k - p_k|] \leq \mathrm{E}[(X_k - p_k)^2] = p_k q_k \, .$$

Dunque, considerando il Teorema 8.5.3 con $\delta = 3$, si ha

$$\frac{1}{\sqrt{v_n^3}} \sum_{k=1}^{n} \mathrm{E}[|X_k - \mu_k|^3] \leq \frac{1}{\sqrt{v_n^3}} \sum_{k=1}^{n} p_k q_k = \frac{1}{\sqrt{v_n}}$$

e la relativa condizione è dunque soddisfatta se $\lim_n v_n = \infty$. In questo caso, se si considera la successione di v.a. $(Z_n)_{n\geq 1}$, con

$$Z_n = \frac{1}{v_n} \sum_{k=1}^{n} (X_k - p_k) = \frac{S_n - \sum_{k=1}^{n} p_k}{\sqrt{\sum_{k=1}^{n} p_k q_k}}$$

e $S_n = \sum_{k=1}^{n} X_k$, risulta $Z_n \xrightarrow{\mathcal{L}} Z$ per $n \to \infty$ dove la v.a. Z si distribuisce con legge Normale $\mathcal{N}(0,1)$. Questo risultato costituisce in effetti una estensione del Teorema Centrale del Limite dato da De Moivre. In particolare, se esiste $c \in]0, 1[$ tale che $p_k \in]c, 1-c[$, allora si ha anche $q_k \in]c, 1-c[$ e da cui segue $v_n > nc^2$, ovvero la condizione di Lyapunov è soddisfatta. La condizione di Lyapunov può essere soddisfatta perfino se $\lim_n p_n = 0$. Ad esempio, se $p_n = n^{-1}$, allora si ha

$$v_n = \sum_{k=1}^{n} \frac{1}{k} - \sum_{k=1}^{n} \frac{1}{k^2}$$

e quindi la condizione è verificata dal momento che la serie armonica $\sum_{n=1}^{\infty} n^{-1}$ non converge, mentre risulta $\sum_{n=1}^{\infty} n^{-2} = \zeta(2) = \frac{\pi^2}{6}$. □

Si considera infine un Teorema di convergenza, comunemente detto metodo Delta, che è spesso di notevole utilità pratica per lo studio di successioni di v.a.

Teorema 8.5.4. (Metodo Delta) *Sia* $(X_n)_{n\geq 1}$ *una successione di v.a. indipendenti con la medesima legge e definite sullo spazio probabilizzato* $(\Omega, \mathcal{F}, P)$*. Sia inoltre* $(Z_n)_{n\geq 1}$ *la successione di v.a. con*

$$Z_n = \sqrt{n} \, \frac{X_n - \theta}{\psi} \, ,$$

dove θ *e* ψ *sono opportune costanti, tale che* $Z_n \xrightarrow{\mathcal{L}} Z$ *per* $n \to \infty$ *e dove la v.a.* Z *è distribuita con legge Normale* $\mathcal{N}(0,1)$*. Se* $g : \mathbb{R} \to \mathbb{R}$ *è una funzione continua e differenziabile tale che* $g'(\theta) \neq 0$*, data la successione di v.a.* $(Y_n)_{n\geq 1}$ *con*

$$Y_n = \sqrt{n}\,\frac{g(X_n) - g(\theta)}{|g'(\theta)|\psi}\,,$$

si ha $Y_n \xrightarrow{\mathcal{L}} Z$ *per* $n \to \infty$.

Dimostrazione. Si veda Billingsley (1995). □

• **Esempio 8.5.2.** Sia $(Z_n)_{n\geq 1}$, dove $Z_n = n^{-1}S_n$, una successione di v.a. indipendenti e con la medesima legge tale che ogni v.a. S_n è distribuita con legge Binomiale $\mathcal{B}(n,p)$. Tenendo presente il Teorema Centrale del Limite di De Moivre, la successione di v.a. $(V_n)_{n\geq 1}$ con

$$V_n = \sqrt{n}\,\frac{Z_n - p}{\sqrt{p(1-p)}}$$

è tale che $V_n \xrightarrow{\mathcal{L}} V$ per $n \to \infty$, dove la v.a. V si distribuisce con legge Normale $\mathcal{N}(0,1)$. Dunque, se $g : \mathbb{R} \to \mathbb{R}$ è una funzione continua e differenziabile tale che $g'(p) \neq 0$, per la successione di v.a. $(Y_n)_{n\geq 1}$ con

$$Y_n = \sqrt{n}\,\frac{g(Z_n) - g(p)}{|g'(p)|\sqrt{p(1-p)}}\,,$$

risulta $Y_n \xrightarrow{\mathcal{L}} V$ per $n \to \infty$. In particolare, se si richiede che $|g'(p)|\sqrt{p(1-p)} = 1$, allora deve essere $g(p) = 2\text{arcsin}(\sqrt{p})$, ovvero in questo caso si ha

$$Y_n = 2\sqrt{n}\,(\text{arcsin}(\sqrt{Z_n}) - \text{arcsin}(\sqrt{p}))\,.$$

Questa particolare scelta della funzione g è detta trasformazione stabilizzatrice, in quanto permette di ottenere una costante al denominatore della precedente definizione della v.a. Y_n. Questo risultato è spesso utile nella statistica inferenziale. □

8.6. Convergenza di vettori aleatori

I concetti di convergenza possono essere estesi al caso di v.v.a. Per quanto riguarda la notazione adottata in questo capitolo, $\|\cdot\|$ denota come al solito la distanza euclidea in $\mathbb{R}^k$. Inoltre, se F_X rappresenta la f.r.c. di un v.v.a. X con k componenti marginali, $x = (x_1, \ldots, x_k)^{\mathrm{T}}$ è un punto di continuità di F_X se la frontiera dell'insieme $]-\infty, x_1] \times \cdots \times]-\infty, x_k]$ ha probabilità nulla.

Definizione 8.6.1. Se $(X_n)_{n\geq 1}$ è una successione di v.v.a. (ognuno dei quali possiede k componenti marginali) a cui corrisponde la successione di f.r.c. $(F_{X_n})_{n\geq 1}$, si dice che la successione di v.v.a. *converge in legge* al v.v.a. X con f.r.c. F_X se per ogni punto di continuità $x \in \mathbb{R}^k$ di F_X si ha

$$\lim_n F_{X_n}(x) = F_X(x)$$

e si scrive $X_n \xrightarrow{\mathcal{L}} X$ per $n \to \infty$. Se $(X_n)_{n\geq 1}$ è una successione di v.v.a. definiti sullo spazio probabilizzato $(\Omega, \mathcal{F}, P)$, si dice che la successione di v.v.a. *converge in probabilità* al v.v.a. X, definito sul medesimo spazio probabilizzato, se per ogni $\varepsilon > 0$ si ha

$$\lim_n P(\|X_n - X\| > \varepsilon) = 0$$

e si scrive $X_n \xrightarrow{P} X$ per $n \to \infty$. Se $(X_n)_{n\geq 1}$ è una successione di v.v.a. definiti sullo spazio probabilizzato $(\Omega, \mathcal{F}, P)$, si dice che la successione di v.v.a. *converge quasi certamente* al v.v.a. X, definito sul medesimo spazio probabilizzato, se l'evento

$$\{\omega \in \Omega : \lim_n X_n(\omega) = X(\omega)\} = \{\lim_n X_n = X\}$$

si verifica $q.c.$ e si scrive $X_n \xrightarrow{q.c.} X$ per $n \to \infty$. Se $(X_n)_{n\geq 1}$ è una successione di v.v.a. definiti sullo spazio probabilizzato $(\Omega, \mathcal{F}, P)$, si dice che la successione di v.v.a. *converge in media di ordine* p al v.v.a. X, definito sul medesimo spazio probabilizzato, se

$$\lim_n \mathrm{E}(\|X_n - X\|^p) = 0$$

per $p \in]0, \infty[$ e si scrive $X_n \xrightarrow{L^p} X$ per $n \to \infty$. □

La gran parte dei risultati ottenuti per successioni di v.a. possono essere opportunamente estesi al caso di v.v.a. Di conseguenza, vengono considerati nel seguito solo alcuni Teoremi che hanno interesse specifico per v.v.a.

Il prossimo Teorema consente di determinare la convergenza in legge di una successione di v.v.a. attraverso la convergenza in legge di combinazioni lineari degli elementi della successione ed è dovuto a Harald Cramér e allo statistico e matematico norvegese Herman Ole Andreas Wold (1908-1992).

Teorema 8.6.2. (Teorema di Cramér-Wold) *Data una successione di v.v.a.* $(X_n)_{n\geq 1}$ *(ognuno dei quali possiede* k *componenti marginali), allora* $X_n \xrightarrow{\mathcal{L}} X$ *per* $n \to \infty$ *se e solo se* $c^{\mathrm{T}} X_n \xrightarrow{\mathcal{L}} c^{\mathrm{T}} X$ *per ogni* $c \in \mathbb{R}^k$.

Dimostrazione. Si veda Billingsley (1995, p.383). □

• **Esempio 8.6.1.** Si consideri la successione di v.v.a. $(Z_n)_{n\geq 1}$ tale che ogni v.v.a. Z_n è distribuito con legge Normale Multivariata $\mathcal{N}_k(\mu, \Sigma)$ e l'ulteriore successione di v.v.a. $(X_n)_{n\geq 1}$ dove $X_n = n^{-1} Z_n$. Dall'Esempio 7.3.2 è noto che la v.a. $c^{\mathrm{T}} Z_n$ si distribuisce con legge Normale $\mathcal{N}(c^{\mathrm{T}}\mu, c^{\mathrm{T}}\Sigma c)$. Dunque, la v.a. $c^{\mathrm{T}} X_n$ è tale che

$$\mathrm{E}[c^{\mathrm{T}} X_n] = \frac{1}{n}\, c^{\mathrm{T}}\mu$$

e

$$\mathrm{Var}[c^{\mathrm{T}} X_n] = \frac{1}{n^2}\, c^{\mathrm{T}}\Sigma c\,.$$

Di conseguenza, dalla Proposizione 8.2.2 segue che $c^{\mathrm{T}} X_n \xrightarrow{P} 0$ per $n \to \infty$ e quindi dalla Proposizione 8.2.4 si ottiene che $c^{\mathrm{T}} X_n \xrightarrow{\mathcal{L}} 0$ per $n \to \infty$. Dunque, dal Teorema 8.6.2 risulta infine che $X_n \xrightarrow{\mathcal{L}} 0$ per $n \to \infty$, ovvero la successione di v.v.a. converge in legge ad un v.v.a. degenere concentrato sul vettore a componenti nulle. □

Teorema 8.6.3. *Data una successione di v.v.a.* $(X_n)_{n\geq 1}$ *(ognuno dei quali possiede* k *componenti marginali) e se il v.v.a.* X *è degenere e tale che* $P(X = c) = 1$ *con* $c \in \mathbb{R}^k$*, allora* $X_n \xrightarrow{P} X$ *per* $n \to \infty$ *se e solo se ogni componente marginale del v.v.a.* X_n *converge in probabilità alla rispettiva componente marginale del v.v.a.* X*. Analogamente,* $X_n \xrightarrow{q.c.} X$ *per* $n \to \infty$ *se e solo se ogni componente marginale del v.v.a.* X_n *converge quasi certamente alla rispettiva componente marginale del v.v.a.* X.

Dimostrazione. Si veda Billingsley (1995, p.378). □

• **Esempio 8.6.2.** Si consideri la successione di v.v.a. $(Z_n)_{n\geq 1}$ tale che il v.v.a. Z_n è distribuito con legge Multinomiale $\mathcal{M}(n, p)$ dove $p = (p_1, \ldots, p_k)^{\mathrm{T}}$ e l'ulteriore successione di v.v.a. $(X_n)_{n\geq 1}$ dove $X_n = n^{-1}Z_n$. Dall'Esempio 7.4.1 è noto che la j-esima componente marginale del v.v.a. Z_n si distribuisce con legge Binomiale $\mathcal{B}(n, p_j)$ e dunque dall'Esempio 8.3.4 si ottiene che la j-esima componente marginale del v.v.a. X_n converge quasi certamente ad una v.a. degenere concentrata su p_j. Quindi, dal Teorema 8.6.3 si ha infine che $X_n \xrightarrow{q.c.} p$ per $n \to \infty$, ovvero la successione di v.v.a. converge quasi certamente al v.v.a. degenere concentrato sul vettore p. □

Teorema 8.6.4. (Teorema Centrale Multivariato del Limite) *Sia $(X_n)_{n\geq 1}$ una successione di v.v.a. indipendenti con la medesima legge (ognuno dei quali possiede k componenti marginali) e definiti sullo spazio probabilizzato $(\Omega, \mathcal{F}, P)$, tali che per ogni n si ha $\mathrm{E}[X_n] = \mu$ e $\mathrm{Var}[X_n] = \Sigma$ dove $\det(\Sigma) < \infty$. Si consideri inoltre la successione di v.v.a. $(Z_n)_{n\geq 1}$ con*

$$Z_n = \frac{1}{\sqrt{n}}\,(S_n - n\mu)$$

e $S_n = \sum_{k=1}^{n} X_k$. Se il v.v.a. Z è distribuito con legge Normale Multivariata $\mathcal{N}_k(0, \Sigma)$, allora $Z_n \xrightarrow{\mathcal{L}} Z$ per $n \to \infty$.

Dimostrazione. Si veda Billingsley (1995, p.385). □

• **Esempio 8.6.3.** Si consideri la successione di v.v.a. $(X_n)_{n\geq 1}$ tale che ogni v.v.a. Z_n è distribuito con legge Multinomiale $\mathcal{M}(1, p)$. Inoltre, dall'Esempio 7.4.5 risulta che la legge del v.v.a. S_n è Multinomiale $\mathcal{M}(n, p)$. Dunque, il Teorema Centrale Multivariato del Limite fornisce la convergenza in legge della successione di v.v.a. $(Z_n)_{n\geq 1}$, dove $Z_n = n^{-\frac{1}{2}}(S_n - np)$, ad un v.v.a. con legge $\mathcal{N}_k(0, \mathrm{diag}(p) - pp^{\mathrm{T}})$ per $n \to \infty$. □

Teorema 8.6.5. (Metodo Delta Multivariato) *Sia $(X_n)_{n\geq 1}$ una successione di v.v.a. indipendenti con la medesima legge (ognuno dei quali possiede k componenti marginali) e definiti sullo spazio probabilizzato $(\Omega, \mathcal{F}, P)$. Sia inoltre $(Z_n)_{n\geq 1}$ la successione di v.v.a. con $Z_n = \sqrt{n}(X_n - \theta)$, tale che $Z_n \xrightarrow{\mathcal{L}} Z$ per $n \to \infty$, dove il v.v.a. Z è distribuito con legge Normale Multivariata $\mathcal{N}_k(0, \Psi)$, mentre $\theta \in \mathbb{R}^k$ e Ψ è una matrice simmetrica definita positiva. Se $g : \mathbb{R}^k \to \mathbb{R}^h$ è una funzione continua e differenziabile tale che $A = \frac{\partial}{\partial x}\, g(x)\big|_{x=\theta}$ è una matrice non nulla, data la successione di v.v.a. $(Y_n)_{n\geq 1}$ con $Y_n = \sqrt{n}(g(X_n) - g(\theta))$, si ha $Y_n \xrightarrow{\mathcal{L}} Y$ per $n \to \infty$, dove il v.v.a. Z è distribuito con legge Normale Multivariata $\mathcal{N}_h(0, A^{\mathrm{T}}\Psi A)$.*

Dimostrazione. Si veda Billingsley (1995, p.385). □

8.7. Riferimenti bibliografici

I testi di Ferguson (1996), Serfling (1980) e van der Vaart (1998) considerano estesamente gli argomenti relativi alle convergenze probabilistiche, con speciale attenzione alle applicazioni statistiche. Il testo di Petrov (1995) è espressamente dedicato ai teoremi limite. I testi con approccio alla probabilità basato sulla Teoria della Misura hanno ampie parti dedicati alle convergenze e ai teoremi limite, come ad esempio Ash e Doléans-Dade (2000), Billingsley (1995), Gut (2005), Resnick (2014) e Shiryaev (2019).

8.8. Esercizi svolti

• **Nota.** Al fine di evitare notazioni ridondanti e ripetizioni, nei seguenti esercizi si assume l'esistenza di uno spazio probabilizzato $(\Omega, \mathcal{F}, P)$, a meno che non venga specificato diversamente.

Sezione 8.1

• **Esercizio 8.1.1.** Si consideri la successione di v.a. $(X_n)_{n\geq 1}$ tale che la v.a. X_n si distribuisce con legge Normale $\mathcal{N}(\mu_n, \sigma_n^2)$. Sia inoltre $\lim_n \mu_n = \mu$ e $\lim_n \sigma_n^2 = \sigma^2$. Si determini la legge della v.a. X tale che $X_n \xrightarrow{\mathcal{L}} X$ per $n \to \infty$.

Soluzione. La f.c. della v.a. X_n è data da

$$\varphi_{X_n}(t) = e^{\mathrm{i}\mu_n t - \frac{1}{2}\sigma_n^2 t^2} ,$$

e quindi

$$\lim_n \varphi_{X_n}(t) = e^{\mathrm{i}\mu t - \frac{1}{2}\sigma^2 t^2} .$$

Dunque, per il Teorema di Lévy la v.a. X si distribuisce con legge Normale $\mathcal{N}(\mu, \sigma^2)$. □

• **Esercizio 8.1.2.** Si consideri la successione di v.a. indipendenti $(X_n)_{n\geq 1}$ tale che la f.p. della v.a. X_n è data da

$$p_{X_n}(x) = \frac{1}{2}\,\mathbf{1}_{\{-2^{-n}, 2^{-n}\}}(x) .$$

Inoltre, si consideri l'ulteriore successione di v.a. $(S_n)_{n\geq 1}$, dove $S_n = \sum_{k=1}^n X_k$. Se la v.a. X si distribuisce con legge Uniforme su $]-1, 1[$, si verifichi che $S_n \xrightarrow{\mathcal{L}} X$ per $n \to \infty$.

Soluzione. Dal momento che la v.a. X_n è simmetrica rispetto all'origine, la f.c. è data da

$$\varphi_{X_n}(t) = \mathrm{E}[\cos(tX_n)] = \cos(2^{-n}t)$$

e quindi

$$\varphi_{S_n}(t) = \prod_{k=1}^n \cos(2^{-k}t) .$$

Evidentemente, per $t = 0$ risulta $\varphi_{S_n}(0) = 1$. Inoltre, tenendo presente la formula di duplicazione $\sin(2\theta) = 2\sin(\theta)\cos(\theta)$, per $t \neq 0$ si ha

$$\varphi_{S_n}(t) = \prod_{k=1}^n \frac{2\sin(2^{-k}t)\cos(2^{-k}t)}{2\sin(2^{-k}t)} = \prod_{k=1}^n \frac{\sin(2^{-k+1}t)}{2\sin(2^{-k}t)} = \frac{\sin(t)}{2^n \sin(2^{-n}t)} ,$$

dal momento che il prodotto è telescopico. Quindi, risulta

$$\lim_n \varphi_{S_n}(t) = \lim_n \frac{\sin(t)}{t}\frac{2^{-n}t}{\sin(2^{-n}t)} = \frac{\sin(t)}{t} .$$

Inoltre, la v.a. X è simmetrica rispetto all'origine e la corrispondente f.c. è data da

$$\varphi_X(t) = \mathrm{E}[\cos(tX)] = \int_{-1}^1 \cos(tx)\,\frac{1}{2}\,dx = \frac{\sin(t)}{t} .$$

Dunque, per il Teorema di Lévy si ha $S_n \xrightarrow{\mathcal{L}} X$ per $n \to \infty$. □

• **Esercizio 8.1.3.** Si consideri la successione $(X_n)_{n\geq 1}$ di v.a. indipendenti e con la medesima legge, tale che la f.p. della v.a. discreta X_n è data da

$$p_{X_n}(x) = \frac{1}{10}\,\mathbf{1}_{\{0,1,\ldots,9\}}(x)\,,$$

ovvero la v.a. X_n è distribuita con legge Uniforme discreta su $\{0,1,\ldots,9\}$. Inoltre, si consideri la successione di v.a. $(S_n)_{n\geq 1}$, dove $S_n = \sum_{k=1}^{n} 10^{-k} X_k$. Se la v.a. X è distribuita con legge Uniforme su $]0,1[$, si verifichi che $S_n \xrightarrow{\mathcal{L}} X$ per $n \to \infty$.

Soluzione. Dal momento che per ogni $a \in \mathbb{C}$ e tale che $a \neq 1$ si ha

$$\sum_{k=0}^{n} a^k = \frac{1-a^{n+1}}{1-a}\,,$$

per $t \neq 0$ la f.c. della v.a. X_n è data da

$$\varphi_{X_n}(t) = \sum_{x=0}^{9} \frac{1}{10}\, e^{\mathrm{i}tx} = \frac{1}{10}\,\frac{1-e^{10\mathrm{i}t}}{1-e^{\mathrm{i}t}}\,,$$

mentre $\varphi_{X_n}(0) = 1$. Quindi, risulta

$$\varphi_{S_n}(t) = \prod_{k=1}^{n} \varphi_{X_n}(10^{-k}t) = \prod_{k=1}^{n} \frac{1}{10}\,\frac{1-e^{10^{-k+1}\mathrm{i}t}}{1-e^{10^{-k}\mathrm{i}t}} = \frac{1}{10^n}\,\frac{1-e^{\mathrm{i}t}}{1-e^{10^{-n}\mathrm{i}t}}\,,$$

dal momento che il prodotto è telescopico. Dalla precedente espressione si ha

$$\lim_n \varphi_{S_n}(t) = \lim_n \frac{e^{\mathrm{i}t}-1}{\mathrm{i}t}\,\frac{10^{-n}\mathrm{i}t}{e^{10^{-n}\mathrm{i}t}-1} = \frac{e^{\mathrm{i}t}-1}{\mathrm{i}t}\,.$$

Inoltre, la f.c. della v.a. X è data da

$$\varphi_X(t) = \int_0^1 \cos(tx)\,dx + \mathrm{i}\int_0^1 \sin(tx)\,dx = \frac{e^{\mathrm{i}t}-1}{\mathrm{i}t}\,.$$

Dunque, per il Teorema di Lévy risulta $S_n \xrightarrow{\mathcal{L}} X$ per $n \to \infty$. □

• **Esercizio 8.1.4.** (Legge di Fréchet) Si consideri una successione $(Y_n)_{n\geq 1}$ di v.a. indipendenti tale che la v.a. Y_n è distribuita con legge di Pareto ridotta con parametro di forma pari ad α, ovvero la f.r. di Y_n è data da

$$F_{Y_n}(y) = (1 - y^{-\alpha})\,\mathbf{1}_{]1,\infty[}(y)$$

con $\alpha > 0$ (vedi Esempio 8.2.5). Data la successione di v.a. $(X_n)_{n\geq 1}$, dove $X_n = n^{-\frac{1}{\alpha}} \max_{1\leq k\leq n} Y_k$, si determini la legge della v.a. X tale che $X_n \xrightarrow{\mathcal{L}} X$ per $n \to \infty$.

Soluzione. Considerata la successione $(Z_n)_{n\geq 1}$, dove $Z_n = \max_{1\leq k\leq n} Y_k$, dall'assunzione di indipendenza si ha

$$F_{Z_n}(z) = P(Y_1 \leq z, \ldots, Y_n \leq z) = \prod_{k=1}^{n} F_{Y_k}(z) = (1-z^{-\alpha})^n\,\mathbf{1}_{]1,\infty[}(z)\,.$$

Dunque, risulta

$$F_{X_n}(x) = P(X_n \leq x) = P(Z_n \leq n^{\frac{1}{\alpha}}x) = \left(1 - \frac{x^{-\alpha}}{n}\right)^n \mathbf{1}_{]n^{-\frac{1}{\alpha}},\infty[}(x)\,,$$

da cui

$$\lim_n F_{X_n}(x) = \lim_n \left(1 - \frac{x^{-\alpha}}{n}\right)^n \mathbf{1}_{]n^{-\frac{1}{\alpha}},\infty[}(x) = e^{-x^{-\alpha}}\,\mathbf{1}_{]0,\infty[}(x)\,.$$

La precedente espressione fornisce la f.r. di una v.a. X con legge di Fréchet ridotta con parametro di forma pari ad α. Si deve concludere che $X_n \xrightarrow{\mathcal{L}} X$ per $n \to \infty$. La legge prende il nome dal matematico francese Maurice René Fréchet (1878-1973). □

• **Esercizio 8.1.5.** (Legge Zeta, seconda parte) Si consideri la v.a. X con legge Zeta (si veda l'Esercizio 4.1.2 e l'Esempio 7.2.7). Tenendo presente la formula di Eulero

$$\zeta(s) = \prod_{p\in\mathbb{P}} \frac{1}{1-p^{-s}}\,,$$

dove $\mathbb{P}$ rappresenta l'insieme dei numeri primi, e considerata la successione di v.a. indipendenti $(X_p)_{p\in\mathbb{P}}$, dove la v.a. X_p è distribuita con legge Geometrica di parametro $(1-p^{-s})$, si verifichi che

$$X = \prod_{p\in\mathbb{P}} p^{X_p}\,.$$

Soluzione. Posto $Y = -\log(X)$ e $Y_p = -\log(p)X_p$, la relazione da verificare è equivalente all'ulteriore relazione

$$Y = \sum_{p\in\mathbb{P}} Y_p\,.$$

Si osservi che la f.c. della v.a. Y è data da

$$\varphi_Y(t) = \mathrm{E}[e^{-\mathrm{i}t\log(X)}] = \mathrm{E}[X^{-\mathrm{i}t}] = \sum_{x=1}^{\infty} x^{-\mathrm{i}t}\,\frac{1}{\zeta(s)}\,x^{-s} = \frac{1}{\zeta(s)}\sum_{x=1}^{\infty} x^{-(s+\mathrm{i}t)} = \frac{\zeta(s+\mathrm{i}t)}{\zeta(s)}\,.$$

Tenendo presente l'espressione della f.c. di una v.a. con legge Geometrica, la f.c. della v.a. Y_p risulta

$$\varphi_{Y_p}(t) = \mathrm{E}[e^{-\mathrm{i}t\log(p)X_p}] = \frac{1-p^{-s}}{1-p^{-s}e^{-\mathrm{i}t\log(p)}} = \frac{1-p^{-s}}{1-p^{-(s+\mathrm{i}t)}}\,.$$

Si consideri inoltre la successione di v.a. $(S_n)_{n\geq 1}$, dove $S_n = \sum_{p\leq n, p\in\mathbb{P}} Y_p$. Dal momento che le v.a. Y_p sono indipendenti, in quanto trasformate di v.a. indipendenti, la f.c. della v.a. S_n risulta

$$\varphi_{S_n}(t) = \prod_{p\leq n, p\in\mathbb{P}} \varphi_{Y_p}(t) = \prod_{p\leq n, p\in\mathbb{P}} \frac{1-p^{-s}}{1-p^{-(s+\mathrm{i}t)}}\,.$$

Dunque, essendo $\frac{1}{\zeta(s)} = \prod_{p\in\mathbb{P}}(1-p^{-s})$ per la formula di Eulero, si ha

$$\lim_n \varphi_{S_n}(t) = \prod_{p\in\mathbb{P}} \frac{1-p^{-s}}{1-p^{-(s+\mathrm{i}t)}} = \frac{\prod_{p\in\mathbb{P}}(1-p^{-s})}{\prod_{p\in\mathbb{P}}(1-p^{-(s+\mathrm{i}t)})} = \frac{\zeta(s+\mathrm{i}t)}{\zeta(s)}$$

e per il Teorema di Lévy si ottiene la relazione richiesta. □

Sezione 8.2

• **Esercizio 8.2.1.** Si consideri la successione $(X_n)_{n\geq 1}$ di v.a. indipendenti con legge Uniforme su $]0,1[$ e la successione $(Y_n)_{n\geq 1}$ tale che

$$Y_n = \left(\prod_{k=1}^{n} X_k \right)^{\frac{1}{n}} .$$

Si verifichi che $Y_n \xrightarrow{P} e^{-1}$ per $n \to \infty$.

Soluzione. Si consideri la successione di v.a. $(Z_n)_{n\geq 1}$ tale che

$$Z_n = \log(Y_n) = \frac{1}{n} \sum_{k=1}^{n} \log(X_k) .$$

Si ha

$$\mathrm{E}[\log(X_n)] = \int_0^1 \log(x)\, dx = -1 .$$

Quindi, sulla base della Legge Debole dei Grandi Numeri di Khinchin, si ha $Z_n \xrightarrow{P} -1$ per $n \to \infty$. Dal momento che la funzione esponenziale è continua, risulta infine $Y_n = e^{Z_n} \xrightarrow{P} e^{-1}$ per $n \to \infty$. □

• **Esercizio 8.2.2.** Si consideri la successione di v.a. indipendenti $(X_n)_{n\geq 1}$ tale che la f.p. della v.a. X_n è data da

$$p_{X_n}(x) = \frac{1}{2}\, \mathbf{1}_{\{-n^a, n^a\}}(x) .$$

Inoltre, si consideri l'ulteriore successione di v.a. $(Z_n)_{n\geq 1}$ con $Z_n = n^{-1} \sum_{k=1}^{n} X_k$. Si verifichi che $Z_n \xrightarrow{P} 0$ per $n \to \infty$ se $a < \frac{1}{2}$.

Soluzione. Si ha $\mathrm{E}[X_n] = 0$ e $\mathrm{Var}[X_n] = n^{2a}$, da cui $\mathrm{E}[Z_n] = 0$ e

$$\mathrm{Var}[Z_n] = \frac{1}{n^2} \sum_{k=1}^{n} k^{2a} .$$

Tenendo presente che $\sum_{k=1}^{n} k^c = O(n^{c+1})$ se $c > -1$, allora risulta anche $\mathrm{Var}[Z_n] = O(n^{2a-1})$. Inoltre, dalla disuguaglianza di Chebyshev, per $\varepsilon > 0$ si ha

$$P(|Z_n| > \varepsilon) \leq \frac{\mathrm{Var}[Z_n]}{\varepsilon^2} .$$

Dunque, si ha $Z_n \xrightarrow{P} 0$ per $n \to \infty$ se $\lim_n \mathrm{Var}[Z_n] = 0$, ovvero deve sussistere $a < \frac{1}{2}$. □

Sezione 8.3

• **Esercizio 8.3.1.** Si consideri la successione $(X_n)_{n\geq 1}$ di v.a. indipendenti con legge Uniforme su $[0, 1]$ e l'ulteriore successione di v.a. $(Z_n)_{n\geq 1}$ dove $Z_n = \max(X_1, \ldots, X_n)$. Si verifichi che $Z_n \xrightarrow{q.c.} 1$ per $n \to \infty$.

Soluzione. La f.r. della v.a. Z_n è data da

$$F_{Z_n}(z) = \prod_{k=1}^{n} F_{X_n}(z) = z^n\, \mathbf{1}_{[0,1[}(z) + \mathbf{1}_{[1,\infty[}(z) .$$

Tenendo presente che se si verifica l'evento $\{Z_n > z\}$ si verificano anche tutti gli eventi $\{Z_k > z\}$ per $k \geq n$, se $\varepsilon \in]0, 1]$ risulta

$$\lim_n P\left(\bigcap_{k=n}^{\infty}\{|Z_k-1|\le\varepsilon\}\right)=\lim_n P\left(\bigcap_{k=n}^{\infty}\{Z_k>1-\varepsilon\}\right)=\lim_n P(Z_n>1-\varepsilon)$$
$$=\lim_n \min(1,1-(1-\varepsilon)^n)=1\,,$$

ovvero $Z_n \overset{q.c.}{\to} 1$ per $n\to\infty$. □

• **Esercizio 8.3.2.** Si consideri la successione $(X_n)_{n\ge1}$ di v.a. indipendenti tale che la v.a. X_n è distribuita con legge Normale $\mathcal{N}(n^{-1},n^{-1})$. Si verifichi che $X_n \overset{q.c.}{\to} 0$ per $n\to\infty$.

Soluzione. Dal momento che $X_n \overset{\mathcal{L}}{=} n^{-\frac{1}{2}}Z+n^{-1}$, dove la v.a. Z si distribuisce con legge Normale $\mathcal{N}(0,1)$, allora si ha

$$\mathrm{E}[|X_n|^4]=\frac{1}{n^4}\,\mathrm{E}[(n^{\frac{1}{2}}Z+1)^4]=\frac{1}{n^4}\,\mathrm{E}[n^2Z^4+4n^{\frac{3}{2}}Z^3+6nZ^2+4n^{\frac{1}{2}}Z+1]$$
$$=\frac{1}{n^2}\,\mathrm{E}[Z^4]+\frac{6}{n^3}\,\mathrm{E}[Z^2]+\frac{1}{n^4}=\frac{3}{n^2}+\frac{6}{n^3}+\frac{1}{n^4}$$

da cui

$$\sum_{n=1}^{\infty}\mathrm{E}[|X_n|^4]=3\zeta(2)+6\zeta(3)+\zeta(4)<\infty\,.$$

Dunque, si deve concludere che $X_n \overset{q.c.}{\to} 0$ per $n\to\infty$. □

Sezione 8.4

• **Esercizio 8.4.1.** Data la v.a. assolutamente continua Z che ammette d.p.

$$f_Z(z)=\frac{\log(2)}{z\log^2(z)}\,\mathbf{1}_{[2,\infty[}(z)\,,$$

si consideri la successione di v.a. $(X_n)_{n\ge1}$ tale che $X_n=n^{-1}Z$. Si verifichi che la successione di v.a. converge in probabilità ad una v.a. degenere concentrata su 0, ma non converge in media.

Soluzione. La f.r. della v.a. Z è data da

$$F_Z(z)=1-\frac{\log(2)}{\log(z)}\,\mathbf{1}_{[2,\infty[}(z)\,,$$

da cui si ottiene

$$\lim_n P(|X_n|>\varepsilon)=\lim_n P(Z>n\varepsilon)=\lim_n\frac{\log(2)}{\log(n\varepsilon)}=0\,,$$

ovvero $X_n \overset{P}{\to} 0$ per $n\to\infty$. Tuttavia, per ogni $p>0$ si ha

$$\mathrm{E}[|X_n|^p]=\frac{1}{n^p}\,\mathrm{E}[Z^p]=\frac{\log(2)}{n^p}\int_2^{\infty}\frac{z^{p-1}}{\log^2(z)}=\infty$$

e quindi la successione non converge in media ad una v.a. degenere concentrata su 0. □

• **Esercizio 8.4.2.** Si consideri la successione di v.a. indipendenti $(X_n)_{n\ge1}$ tale che X_n è una v.a. con f.p. data da

$$p_{X_n}(x) = (1 - n^{-a})\,\mathbf{1}_{\{0\}}(x) + n^{-a}\,\mathbf{1}_{\{n\}}(x)\,,$$

dove $a > 0$. Si verifichi che $X_n \xrightarrow{L^p} 0$ per $n \to \infty$ se $a > p$. Si commenti inoltre la convergenza quasi certa e in probabilità della successione $(X_n)_{n\geq 1}$.

Soluzione. Si osservi che $\mathrm{E}[|X_n|^p] = n^{p-a}$. Quindi, se $a > p$ si ha

$$\lim_n \mathrm{E}[|X_n|^p] = \lim_n n^{p-a} = 0\,,$$

ovvero $X_n \xrightarrow{L^p} 0$ per $n \to \infty$. Per quanto riguarda la convergenza quasi certa, se $a > 1$ si ha

$$\sum_{n=1}^{\infty} P(|X_n| > \varepsilon) = \sum_{n=1}^{\infty} n^{-a} = \zeta(a) < \infty\,,$$

ovvero $X_n \xrightarrow{q.c.} 0$ per $n \to \infty$. Dunque, quando $a = 2$ si ottiene un esempio di una successione di v.a. che converge quasi certamente, ma non converge in media quadratica. Infine, se $a > 0$ si ha

$$\lim_n P(|X_n| > \varepsilon) = \lim_n n^{-a} = 0\,,$$

ovvero $X_n \xrightarrow{P} 0$ per $n \to \infty$. Dunque, quando $a = \frac{1}{2}$ si ottiene un esempio di una successione di v.a. che converge in probabilità, ma non converge quasi certamente e in media quadratica. □

Sezione 8.5

• **Esercizio 8.5.1.** Si consideri la successione $(X_n)_{n\geq 1}$ di v.a. indipendenti discrete, tale che la f.p. della v.a. X_n è data da

$$p_{X_n}(x) = \frac{1}{2}\,\mathbf{1}_{\{-an,an\}}(x)\,,$$

dove $a > 0$. Si consideri inoltre la successione di v.a. $(S_n)_{n\geq 1}$, dove $S_n = \sum_{k=1}^n X_k$. Si verifichi la condizione del Teorema Centrale del Limite di Lindeberg-Lévy-Feller.

Soluzione. Si ha $\mathrm{E}[X_n] = 0$ e $\mathrm{Var}[X_n] = a^2n^2$. Dunque, risulta

$$v_n^2 = \sum_{k=1}^n a^2k^2 = \frac{a^2n(n+1)(2n+1)}{6}\,.$$

Si noti che per ogni $\varepsilon > 0$ esiste un valore N_ε (che non dipende da a) tale che per $n > N_\varepsilon$ si ha $\varepsilon v_n > an$, in quanto questa disuguaglianza può essere espressa come

$$\frac{(n+1)(2n+1)}{6n} > \frac{1}{\varepsilon^2}\,.$$

Quindi, si ha

$$\lim_n \sum_{k=1}^n \mathrm{E}[X_k^2\,\mathbf{1}_{]\varepsilon v_n,\infty[}(|X_k|)] < \infty\,.$$

Inoltre, risulta $\lim_n v_n^2 = \infty$ per cui

$$\lim_n \frac{1}{v_n^2} \sum_{k=1}^n \mathrm{E}[X_k^2\,\mathbf{1}_{]\varepsilon v_n,\infty[}(|X_k|)] = 0\,.$$

Dunque, la condizione del Teorema Centrale del Limite di Lindeberg-Lévy-Feller è verificata per ogni a. Pertanto, se si considera la successione di v.a. $(Z_n)_{n\geq 1}$ con $Z_n = \frac{1}{v_n} S_n$, allora $Z_n \xrightarrow{\mathcal{L}} Z$ per $n \to \infty$, dove la v.a. Z si distribuisce con legge Normale $\mathcal{N}(0,1)$. Si osservi che la condizione del Teorema Centrale del Limite di Lyapunov è in questo caso più facile da verificare. In effetti, scelto $\delta = 3$, si ha $\mathrm{E}[|X_n|^3] = a^3 n^3$, per cui risulta

$$\sum_{k=1}^{n} \mathrm{E}[|X_k|^3] = \sum_{k=1}^{n} a^3 k^3 = \frac{a^3 n^2 (n+1)^2}{4} \,.$$

Dunque, per ogni $a \in \mathbb{R}^+$ si ha

$$\lim_n \frac{1}{\sqrt{v_n^3}} \sum_{k=1}^{n} \mathrm{E}[|X_k|^3] = 0 \,. \qquad \square$$

• **Esercizio 8.5.2.** (Formula di Stirling) Si dimostri la formula di Stirling

$$n! \sim \sqrt{2\pi}\, n^{n+\frac{1}{2}}\, e^{-n}$$

mediante l'uso della Teorema Centrale del Limite di Lindeberg-Lévy.

Soluzione. Si consideri una successione $(X_n)_{n\geq 1}$ di v.a. indipendenti e ugualmente distribuite con legge Esponenziale ridotta. Evidentemente, si ha $\mathrm{E}[X_n] = \mathrm{Var}[X_n] = 1$. Posto $S_n = \sum_{k=1}^{n} X_k$, si consideri inoltre la successione di v.a. $(Z_n)_{n\geq 1}$ dove

$$Z_n = \frac{S_n - n}{\sqrt{n}} \,.$$

Quindi, dal Teorema Centrale del Limite di Lindeberg-Lévy risulta $Z_n \xrightarrow{\mathcal{L}} Z$, dove la v.a. Z si distribuisce con legge Normale $\mathcal{N}(0,1)$. Dal momento che $\mathrm{E}[Z_n^2] = 1$ per ogni n, allora risulta $\lim_n \mathrm{E}[Z_n^2] = \mathrm{E}[Z^2] = 1$ e questo implica che anche i momenti assoluti di ordine inferiore convergano, ovvero si ha

$$\lim_n \mathrm{E}[|Z_n|] = \mathrm{E}[|Z|] = 2\int_0^\infty z\phi(z)\, dz = \sqrt{\frac{2}{\pi}} \,.$$

Inoltre, tenendo presente che la v.a. S_n è distribuita con legge Gamma $\mathcal{G}(0,1,n)$, segue che

$$\begin{aligned} \mathrm{E}[|Z_n|] &= \int_0^\infty \left| \frac{z-n}{\sqrt{n}} \right| \frac{1}{\Gamma(n)}\, z^{n-1}\, e^{-z}\, dz \\ &= \frac{n^{n+\frac{1}{2}}}{(n-1)!} \int_0^\infty |u-1| u^{n-1} e^{-nu}\, du = \frac{2n^{n+\frac{1}{2}}\, e^{-n}}{n!} \,. \end{aligned}$$

Dunque, si deve concludere che

$$\lim_n \frac{\mathrm{E}[|Z_n|]}{\mathrm{E}[|Z|]} = \lim_n \frac{\sqrt{2\pi}\, n^{n+\frac{1}{2}}\, e^{-n}}{n!} = 1 \,,$$

che fornisce una verifica probabilistica della formula di Stirling. La formula prende il nome dal matematico scozzese James Stirling (1692-1770). $\square$

• **Esercizio 8.5.3.** Si consideri la successione $(X_n)_{n\geq 1}$ di v.a. indipendenti e con la medesima legge, tale che $\mu = \mathrm{E}[X_n]$ e $\sigma^2 = \mathrm{Var}[X_n] < \infty$. Si verifichi che la condizione del Teorema Centrale del Limite di Lindeberg-Lévy-Feller è soddisfatta per la successione considerata.

Soluzione. Senza perdita di generalità si può assumere $\mu = 0$. Si ha

$$v_n^2 = \sum_{k=1}^{n} \sigma^2 = n\sigma^2 \,,$$

mentre

$$\sum_{k=1}^{n} \mathrm{E}[X_k^2 \, \mathbf{1}_{]\varepsilon v_n, \infty[}(|X_k|)] = n\mathrm{E}[X_1^2 \, \mathbf{1}_{]\varepsilon\sqrt{n}\sigma, \infty[}(|X_1|)] \;.$$

Dunque, si ottiene infine

$$\lim_n \frac{1}{v_n^2} \sum_{k=1}^{n} \mathrm{E}[X_k^2 \mathbf{1}_{]\varepsilon v_n, \infty[}(|X_k|)] = \lim_n \frac{1}{\sigma^2} \, \mathrm{E}[X_1^2 \, \mathbf{1}_{]\varepsilon\sqrt{n}\sigma, \infty[}(|X_1|)] = 0 \,,$$

ovvero la condizione del Teorema Centrale del Limite di Lindeberg-Lévy-Feller è soddisfatta. Quindi, si noti che quando le condizioni del Teorema Centrale del Limite di Lindeberg-Lévy sono soddisfatte, anche quella del Teorema Centrale del Limite di Lindeberg-Lévy-Feller è soddisfatta. □

• **Esercizio 8.5.4.** Si consideri la successione $(X_n)_{n\geq 1}$ di v.a. indipendenti e con la medesima legge, tale che la d.p. della v.a. X_n è data da

$$f_{X_n}(x) = \frac{1}{c|x|^3 \log^2(|x|)} \, \mathbf{1}_{]2,\infty[}(|x|) \,,$$

dove

$$c = 2 \int_2^{\infty} \frac{1}{x^3 \log^2(x)} \, dx = 4 \, \mathrm{Ei}(-\log(4)) + \frac{1}{\log(4)} \,,$$

mentre

$$\mathrm{Ei}(z) = \int_{-\infty}^{z} \frac{e^u}{u} \, du$$

rappresenta la funzione integrale esponenziale. Si verifichi che la condizione del Teorema Centrale del Limite di Lyapunov non è soddisfatta per la successione considerata.

Soluzione. Si ha $\mu = \mathrm{E}[X_n] = 0$ e

$$\sigma^2 = \mathrm{Var}[X_n] = \frac{2}{c} \int_2^{\infty} \frac{1}{x \log^2(x)} \, dx = \frac{2}{c \log(2)} < \infty \,.$$

Inoltre, per $\delta > 2$ risulta

$$\mathrm{E}[|X_n|^{\delta}] = \frac{2}{c} \int_2^{\infty} \frac{x^{\delta-3}}{\log^2(x)} \, dx = \infty$$

e quindi la condizione del Teorema Centrale del Limite di Lyapunov non può essere soddisfatta. Tuttavia, si noti che $(X_n)_{n\geq 1}$ è una successione di v.a. indipendenti con la medesima legge e tale che $\mathrm{Var}[X_n] < \infty$. Dunque, sulla base dell'Esercizio 8.5.3, la condizione del Teorema Centrale del Limite di Lindeberg-Lévy-Feller è invece soddisfatta. □

Sezione 8.6

• **Esercizio 8.6.1.** Si consideri la successione $(X_n)_{n\geq 1}$ di v.a. indipendenti e con la medesima legge della v.a. X tale che $\mathrm{E}[X^4] < \infty$. Si consideri inoltre la successione di v.v.a. $(S_n)_{n\geq 1}$ tale che $S_n = (S_{n,1}, S_{n,2})^{\mathrm{T}} = (\sum_{k=1}^n X_k, \sum_{k=1}^n X_k^2)^{\mathrm{T}}$ e la successione di v.v.a. $(Z_n)_{n\geq 1}$ tale che

$$Z_n = \frac{1}{\sqrt{n}} \left(S_n - n\mathrm{E}[S_n]\right) .$$

Si verifichi che $Z_n \xrightarrow{\mathcal{L}} Z$ per $n \to \infty$ e si determini la legge relativa al v.v.a. Z.

Soluzione. Si assuma che $\mathrm{E}[X] = \mu$ e $\mathrm{E}[X^r] = \mu_r$ per $r = 2, 3, 4$, mentre $\mathrm{Var}[X] = \sigma^2$. Inoltre, si consideri la successione $(Y_n)_{n\geq 1}$ tale che $Y_n = (Y_{n,1}, Y_{n,2})^{\mathrm{T}} = (X_n, X_n^2)^{\mathrm{T}}$. Dunque, risulta immediato verificare che $\nu = \mathrm{E}[Y_n] = (\mu, \mu_2)^{\mathrm{T}}$. Inoltre, si ha $\mathrm{Var}[Y_{n,1}] = \sigma^2$ e $\mathrm{Var}[Y_{n,2}] = \mu_4 - \mu_2^2$, mentre

$$\mathrm{Cov}[Y_{n,1}, Y_{n,2}] = \mu_3 - \mu\mu_2 .$$

Quindi, si ha

$$\Sigma = \mathrm{Var}[Y_n] = \begin{pmatrix} \sigma^2 & \mu_3 - \mu\mu_2 \\ \mu_3 - \mu\mu_2 & \mu_4 - \mu_2^2 \end{pmatrix} .$$

Dal momento che $S_n = \sum_{k=1}^n Y_k$, allora risulta $\mathrm{E}[S_n] = n\nu$ e $\mathrm{Var}[S_n] = n\Sigma$. Per il Teorema Centrale Multivariato del Limite si ha dunque che il v.v.a. $Z = \frac{1}{\sqrt{n}}(S_n - n\nu)$ è distribuito con legge Normale Multivariata $\mathcal{N}_2(0, \Sigma)$. In un contesto di Statistica inferenziale, si osservi che le componenti del v.v.a. $U_n = (U_{n,1}, U_{n,2})^{\mathrm{T}} = n^{-1}S_n$ sono date rispettivamente dalla media campionaria e dal secondo momento campionario. Dunque, il risultato ottenuto implica che $\sqrt{n}(U_n - \nu) \xrightarrow{\mathcal{L}} Z$, ovvero si è ottenuta la distribuzione per "grandi campioni" dei primi due momenti campionari $U_{n,1}$ e $U_{n,2}$. □

• **Esercizio 8.6.2.** Si consideri la successione $(X_n)_{n\geq 1}$ di v.a. indipendenti e con la medesima legge della v.a. X con $\mathrm{E}[X^4] < \infty$. Inoltre, sia $(V_n)_{n\geq 1}$ una ulteriore successione di v.a. tale che

$$V_n = \frac{1}{n}\sum_{k=1}^n X_k^2 - \left(\frac{1}{n}\sum_{k=1}^n X_k\right)^2 .$$

Considerata la successione di v.a. $(Z_n)_{n\geq 1}$ con $Z_n = \sqrt{n}(V_n - \upsilon)$, si determini la legge relativa alla v.a. Z tale che $Z_n \xrightarrow{\mathcal{L}} Z$ per $n \to \infty$ e il valore di υ.

Soluzione. Si adotti le notazioni dell'Esercizio 8.6.1. Si osservi che

$$V_n = \frac{1}{n}\sum_{k=1}^n (X_k - \mu)^2 - \left(\frac{1}{n}\sum_{k=1}^n (X_k - \mu)\right)^2$$

e quindi per semplicità si può assumere $\mu = 0$, da cui $\mu_2 = \sigma^2$. Inoltre, risulta $V_n = g(U_n)$, dove $g(u_1, u_2) = u_2 - u_1^2$, per cui

$$A = \frac{\partial}{\partial x} g(x)\Big|_{x=\nu} = \begin{pmatrix} -2\mu \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix} .$$

Quindi, si ha $A^{\mathrm{T}}\Sigma A = \mu_4 - \sigma^2$ e $\upsilon = g(\nu) = \sigma^2$. Sulla base del metodo Delta Multivariato, si deve dunque concludere che $Z_n = \sqrt{n}(V_n - \sigma^2) \xrightarrow{\mathcal{L}} Z$, dove la v.a. Z è distribuita con legge Normale $\mathcal{N}(0, \mu_4 - \sigma^2)$. In un contesto di Statistica inferenziale, si osservi che la v.a. V_n è in effetti la varianza campionaria, di cui si è ottenuta la distribuzione per "grandi campioni". □

Pagina intenzionalmente vuota

Capitolo 9

Processi aleatori

9.1. Concetti preliminari

Molti esperimenti o fenomeni aleatori danno luogo ad un insieme di realizzazioni all'evolversi di un "parametro", come ad esempio il tempo. In questo caso, l'interesse si concentra sull'analisi di collezioni di v.a. indicizzate su un determinato insieme, che nella Teoria della Probabilità sono dette processi aleatori. Formalmente, si ha la seguente definizione.

Definizione 9.1.1. Se $\mathbb{T} \subseteq [0, \infty[$ è un insieme di indici, una collezione di v.a. $X = (X_t)_{t\in\mathbb{T}}$ definita sullo spazio probabilizzato $(\Omega, \mathcal{F}, P)$ è detta *processo aleatorio* (p.a.) se l'applicazione $X : \mathbb{T} \times \Omega \to \mathbb{R}$, dove $X(t, \omega) = X_t(\omega)$, è misurabile rispetto alla σ-algebra $\mathcal{B}(\mathbb{T}) \otimes \mathcal{F}$. Per un dato $t \in \mathbb{T}$, la v.a. X_t è detta *stato* del p.a. a tempo t. Al contrario, per un dato $\omega \in \Omega$, l'applicazione $t \mapsto X_t(\omega)$ è detta *traiettoria* del p.a. associata con ω. □

Anche se in generale si possono considerare spazi molto generali per $\mathbb{T}$, si noti che nella precedente Definizione l'insieme $\mathbb{T}$ è stato assimilato per semplicità ad un insieme di tempi. Nel seguito si assumerà principalmente che $\mathbb{T} = \mathbb{N}$ o che $\mathbb{T} = [0, \infty[$ (o eventualmente un loro sottoinsieme). Più esattamente, il p.a. sarà detto *a tempo discreto* se card$(\mathbb{T})$ è finita o numerabile o *a tempo continuo* se card$(\mathbb{T})$ non è numerabile. Inoltre, quando $\mathbb{T} = \mathbb{N}$, il p.a. si riduce ad una successione di v.a. e si enfatizzerà questo fatto adottando l'indice n piuttosto che l'indice t.

• **Esempio 9.1.1.** Si consideri una successione di v.a. $(Z_n)_{n\geq 1}$ a componenti indipendenti distribuiti con legge di Bernoulli di parametro p e si definisca il p.a. a tempo discreto $X = (X_n)_{n\geq 0}$ dove $X_n = \sum_{k=1}^{n} Z_k$ e $X_0 = 0$ $q.c.$ In questo caso, si ha ovviamente $\mathbb{T} = \mathbb{N}$. Inoltre, per quanto visto nella Sezione 6.1, per un determinato $n \geq 1$ lo stato X_n è una v.a. con legge Binomiale $\mathcal{B}(n, p)$, mentre una traiettoria è una successione non decrescente a valori su $\mathbb{N}$ tale che gli incrementi fra due termini successivi sono nulli o pari all'unità. In un ambito di gioco d'azzardo, piuttosto che la successione originale $(Z_n)_{n\geq 1}$, si considera invece la successione $(Y_n)_{n\geq 1}$, con $Y_n = 2Z_n - 1$, ovvero si ottiene il p.a. a tempo discreto $X = (X_n)_{n\geq 0}$ dove $X_n = \sum_{k=1}^{n} Y_k = 2\sum_{k=1}^{n} Z_k - n$. In questo caso, il p.a. X viene assimilato ad una successione di giocate in cui si vince un'unità con probabilità p e si perde un'unità con probabilità $(1 - p)$ ad ogni giocata. Lo stato X_n risulta dunque la v.a. che rappresenta la somma vinta (o perduta) a tempo n, mentre una traiettoria è la successione delle somme vinte (o perdute). Con un termine pittoresco, questo p.a. è detto passeggiata aleatoria e sarà analizzato in dettaglio nella Sezione 9.3. □

• **Esempio 9.1.2.** Sia $\mathbb{T} = [0, T]$ e sia $0 = t_0 < t_1 < \ldots < t_n = T$ una partizione di $[0, T]$. Inoltre, si consideri il v.v.a. $Z = (Z_0, \ldots, Z_{n-1})^{\mathrm{T}}$ e si definisca il p.a. a tempo continuo $X = (X_t)_{t\in[0,T]}$ dove

$$X_t = \sum_{k=0}^{n-1} Z_k \, \mathbf{1}_{[t_k, t_{k+1}[}(t) \, .$$

Per un determinato $t \in [t_k, t_{k+1}[$, lo stato X_t è una v.a. con la stessa legge della v.a. Z_k, mentre una traiettoria è una funzione semplice. Di conseguenza, questo p.a. è detto semplice. Questo tipo di p.a. sarà fondamentale nel calcolo stocastico introdotto nel Capitolo 10. □

Definizione 9.1.2. Si consideri uno spazio probabilizzabile $(\Omega, \mathcal{F})$. Se $\mathbb{T} \subseteq [0, \infty[$, una *filtrazione* è una collezione di σ-algebre $(\mathcal{F}_t)_{t\in\mathbb{T}}$ per cui si ha $\mathcal{F}_s \subset \mathcal{F}_t \subset \mathcal{F}$ per ogni $s, t \in \mathbb{T}$ tali che $s \leq t$. Inoltre, dato il p.a. $X = (X_t)_{t\in\mathbb{T}}$ definito sullo spazio probabilizzato $(\Omega, \mathcal{F}, P)$, sia $\mathcal{F}_t = \sigma(\{X_s : s \leq t\})$ la più piccola σ-algebra che contiene gli eventi del tipo $\{X_s \in B\}$ con $B \in \mathcal{B}(\mathbb{R})$ e $s \leq t$. In questo caso, la corrispondente filtrazione viene detta *filtrazione naturale* per il p.a. X. □

Per comodità, si assume usualmente che $\inf(\{t : t \in \mathbb{T}\}) = 0$ e $\mathcal{F}_0 = \{\emptyset, \Omega\}$. Inoltre, uno spazio di probabilità $(\Omega, \mathcal{F}, P)$ a cui è associata la filtrazione $(\mathcal{F}_t)_{t\in\mathbb{T}}$ è detto *spazio di probabilità con filtrazione*. Una filtrazione può essere interpretata come l'incremento di informazione nel p.a. che si ottiene al passare del tempo. In effetti, $\mathcal{F}_t$ può essere vista come l'informazione disponibile ad un osservatore del p.a. fino a tempo t.

• **Esempio 9.1.3.** Si consideri di nuovo il p.a. X dell'Esempio 9.1.1. Si osservi che ogni traiettoria del p.a. può essere posta in corrispondenza biunivoca con un punto dell'intervallo $[0, 1]$, dal momento che ogni determinazione della successione $(Z_n)_{n\geq 1}$ può essere vista come l'insieme delle cifre nella rappresentazione binaria di un punto in $[0, 1]$. In questo caso, si può considerare la filtrazione $(\mathcal{F}_n)_{n\geq 0}$ con $\mathcal{F}_0 = \{\emptyset, \Omega\}$ e dove $\mathcal{F}_n$ è la σ-algebra generata dagli intervalli del tipo $[2^{-n}(k-1), 2^{-n}k[$ con $k = 1, \ldots, 2^n$. Evidentemente, la partizione binaria di $[0, 1]$ che genera $\mathcal{F}_n$ diventa sempre più "fine" all'aumentare di n e la relativa σ-algebra diventa contemporaneamente più "ricca". □

Definizione 9.1.3. Sia $X = (X_t)_{t\in\mathbb{T}}$ un p.a. definito sullo spazio di probabilità $(\Omega, \mathcal{F}, P)$ con filtrazione $(\mathcal{F}_t)_{t\in\mathbb{T}}$. Il p.a. X è detto *adattato* se X_t è misurabile rispetto a $\mathcal{F}_t$ per ogni $t \in \mathbb{T}$. Inoltre, il p.a. X è detto *progressivamente misurabile* se l'applicazione $(s, \omega) \mapsto X_s(\omega)$ è misurabile rispetto alla σ-algebra $\mathcal{B}([0, t]) \otimes \mathcal{F}_t$ per ogni $t \in \mathbb{T}$. □

Dalla Definizione 9.1.3 è evidente che un p.a. X è adattato se la filtrazione è naturale.

Definizione 9.1.4. Sia $X = (X_t)_{t\in\mathbb{T}}$ un p.a. definito sullo spazio di probabilità $(\Omega, \mathcal{F}, P)$. Il p.a. X è detto *continuo* se le corrispondenti traiettorie sono continue $q.c.$ Inoltre, il p.a. è detto *continuo a destra* se le corrispondenti traiettorie sono continue a destra $q.c.$ □

• **Esempio 9.1.4.** Si consideri di nuovo il p.a. semplice X introdotto nell'Esempio 9.1.2. È immediato verificare che le traiettorie del p.a. X non sono continue $q.c.$ In effetti, risulta $\{X_{t_k} = Z_k\}$ $q.c.$, mentre per $\varepsilon > 0$ si ha $\{\lim_{\varepsilon\to 0^+} X_{t_k-\varepsilon} = Z_{k-1}\}$ $q.c.$, dove $k = 1, \ldots, n$. Dunque, il p.a. X non è continuo $q.c.$ Il p.a. X è continuo a destra $q.c.$, dal momento che $\{\lim_{\varepsilon\to 0^+} X_{t_k+\varepsilon} = Z_k\}$ $q.c.$ □

Definizione 9.1.5. Si consideri i p.a. $X = (X_t)_{t\in\mathbb{T}}$ e $Y = (Y_t)_{t\in\mathbb{T}}$ definiti sullo spazio di probabilità $(\Omega, \mathcal{F}, P)$. Si dice che il p.a. Y è una *modificazione* del p.a. X se per ogni $t \in \mathbb{T}$ si ha

$$P(\{\omega \in \Omega : X_t(\omega) = Y_t(\omega)\}) = 1 .$$

Inoltre, si dice che i p.a. X e Y sono *indistinguibili* se si ha

$$P(\{\omega \in \Omega : X_t(\omega) = Y_t(\omega), \forall t \in \mathbb{T}\}) = 1 . \qquad \square$$

Risulta evidente dalla precedente Definizione che se i p.a. X e Y sono indistinguibili, allora il p.a. X è anche una modificazione del p.a. Y, mentre non è vero il contrario. Sulla base di queste classi di

equivalenza fra p.a., si può selezionare il p.a. più opportuno ai fini teorici all'interno di una classe, come ad esempio un p.a. continuo.

• **Esempio 9.1.5.** Si consideri i p.a. $X = (X_t)_{t\in[0,\infty[}$ e $Y = (Y_t)_{t\in[0,\infty[}$ definiti sullo spazio di probabilità $(\Omega, \mathcal{F}, P)$, dove $\Omega = [0, \infty[$ e $\mathcal{F} = \mathcal{B}([0, \infty[)$. Inoltre, i p.a. X e Y sono tali che $X_t = 0$, mentre $Y_t = \mathbf{1}_{\{\omega\}}(t)$. Dunque, i p.a. X e Y non sono indistinguibili, in quanto $X_t(\omega) \neq Y_t(\omega)$ per $t = \omega$. Tuttavia, il p.a. Y è una modificazione del p.a. X, in quanto

$$P(\{\omega \in \Omega : X_t(\omega) \neq Y_t(\omega)\}) = 0 \, .$$

e dunque

$$P(\{\omega \in \Omega : X_t(\omega) = Y_t(\omega)\}) = 1 - P(\{\omega \in \Omega : X_t(\omega) \neq Y_t(\omega)\}) = 1 \, . \qquad \square$$

9.2. Martingale

Le cosiddette "martingale" costituiscono una vasta classe di p.a. con proprietà notevoli. Il termine "martingala" ha origine nel gioco d'azzardo, anche se la sua etimologia non è del tutto chiara. La prima apparizione del termine in ambito probabilistico è dovuta al matematico Jean Ville (1910-1989) e lo studio di questi p.a. è stato profondamente influenzato dal probabilista Joseph Leo Doob (1910-2004), che ne ha costruito le basi teoriche.

Figura 9.2.1. Joseph Leo Doob (1910-2004).

Definizione 9.2.1. Se $\mathbb{T} \subseteq [0, \infty[$, sia $M = (M_t)_{t\in\mathbb{T}}$ un p.a. adattato definito sullo spazio di probabilità $(\Omega, \mathcal{F}, P)$ con filtrazione $(\mathcal{F}_t)_{t\in\mathbb{T}}$ e tale che $\mathrm{E}[|M_t|] < \infty$ per ogni $t \in \mathbb{T}$. Si dice che M è una *martingala* se per ogni $0 \leq s < t$ si ha

$$\{\mathrm{E}[M_t \mid \mathcal{F}_s] = M_s\} \; q.c. \, ,$$

mentre si dice che M è una *super-martingala* se

$$\{\mathrm{E}[M_t \mid \mathcal{F}_s] \leq M_s\} \; q.c. \, ,$$

e infine si dice che M è una *sub-martingala* se

$$\{\mathrm{E}[M_t \mid \mathcal{F}_s] \geq M_s\} \; q.c. \, . \qquad \square$$

È evidente che la proprietà di martingala dipende dalla filtrazione e dalla misura di probabilità P che sono state considerate. Inoltre, dalla Definizione 9.2.1, per ogni t si ha

$$\{\mathrm{E}[M_t \mid \mathcal{F}_0] = M_0\} \; q.c. \, ,$$

ovvero, una martingala può essere considerato in effetti un p.a. che rimane "costante in media". Analogamente, una super-martingala è un p.a. che "decresce in media" e una sub-martingala è un p.a. che "cresce in media".

Risulta importante sottolineare che ogni martingala ammette sempre una modificazione continua a destra che è unica, a meno di p.a. indistinguibili (si veda Pascucci, 2011, p.115). Dal momento che questa condizione non è restrittiva, nel seguito verrà considerata tacitamente la versione continua a destra di ogni martingala.

• **Esempio 9.2.1.** Si consideri una successione di v.a. $(X_n)_{n\geq 1}$ a componenti indipendenti tale che $\mathrm{E}[X_n] = c < \infty$ per ogni n. Si definisca inoltre il p.a. a tempo discreto $M = (M_n)_{n\geq 0}$ tale che

$$M_n = \sum_{k=1}^{n} X_k$$

con $\{M_0 = 0\}$ $q.c.$ e sia $(\mathcal{F}_n)_{n\geq 0}$ la filtrazione naturale, ovvero si ha $\mathcal{F}_n = \sigma(\{X_1, \ldots, X_n\})$. Tenendo presente le assunzioni fatte sul p.a. M, per ogni n si ha

$$\mathrm{E}[|M_n|] = \mathrm{E}[|\sum_{k=1}^{n} X_k|] \leq \mathrm{E}[\sum_{k=1}^{n} |X_k|] = \sum_{k=1}^{n} \mathrm{E}[|X_k|] < \infty \, .$$

Inoltre, dal momento che $\{\mathrm{E}[M_n \mid \mathcal{F}_n] = M_n\}$ $q.c.$ sulla base della Proposizione 5.2.1 e che la σ-algebra generata da X_n è indipendente dalla σ-algebra $\mathcal{F}_{n-1}$, si ha $q.c.$

$$\begin{aligned} \mathrm{E}[M_n \mid \mathcal{F}_{n-1}] &= \mathrm{E}[M_{n-1} + X_n \mid \mathcal{F}_{n-1}] = \mathrm{E}[M_{n-1} \mid \mathcal{F}_{n-1}] + \mathrm{E}[X_n \mid \mathcal{F}_{n-1}] \\ &= M_{n-1} + \mathrm{E}[X_n] = M_{n-1} + c \, . \end{aligned}$$

Procedendo in modo iterativo, per $m \in \mathbb{N}$ con $m < n$, dalla precedente espressione si ha $q.c.$

$$\mathrm{E}[M_n \mid \mathcal{F}_m] = M_m + (n-m)c \, .$$

Quindi, il p.a. M è una martingala se $c = 0$, una super-martingala se $c < 0$ e una sub-martingala se $c > 0$. □

• **Esempio 9.2.2.** (Strategia di gioco) Si consideri una successione $(Y_n)_{n\geq 1}$ di v.a. indipendenti con la medesima legge, tali che $P(Y_n = 1) = p$ e $P(Y_n = -1) = q$, dove $q = 1 - p$, e una successione $(U_n)_{n\geq 1}$ di v.a. tale che $U_n = g_n(Y_1, \ldots, Y_{n-1})$ con $P(U_n \geq 0) = 1$ e $\mathrm{E}[U_n] < \infty$ per ogni $n \in \mathbb{N}$. Si definisca il p.a. a tempo discreto $M = (M_n)_{n\geq 0}$ tale che

$$M_n = \sum_{k=1}^{n} U_k Y_k$$

con $\{M_0 = 0\}$ $q.c.$ e sia $(\mathcal{F}_n)_{n\geq 0}$ la filtrazione naturale, ovvero si assume che $\mathcal{F}_n = \sigma(\{Y_1, \ldots, Y_n\})$. Dalle assunzioni fatte si noti che la v.a. U_n è misurabile rispetto a $\mathcal{F}_{n-1}$. Questo p.a. può essere assimilato ad un insieme di giocate ad un gioco d'azzardo con esito dicotomico, in cui si vince con probabilità p e si perde con probabilità q, e dove si scommette una quantità di denaro pari a U_n all'n-esima giocata. In questo caso, M_n rappresenta la somma vinta (o perduta) dal giocatore all'n-esima giocata. Nella terminologia del gioco d'azzardo, la successione di v.a. $(U_n)_{n\geq 1}$ è detta *strategia di gioco*. Risulta evidente il motivo per cui la v.a. U_n dipende solo dal v.v.a. $(Y_1, \ldots, Y_{n-1})^{\mathrm{T}}$. In effetti, il giocatore può prendere la decisione sulla somma da scommettere all'n-esima giocata solo basandosi

sugli esiti delle $(n-1)$ giocate precedenti. Tenendo presente che le v.a. U_n e Y_n sono indipendenti e che $\mathrm{E}[|Y_n|] = 1$, per ogni n si ha

$$\mathrm{E}[|M_n|] = \mathrm{E}[|\sum_{k=0}^{n} U_k Y_k|] \leq \mathrm{E}[\sum_{k=0}^{n} |U_k Y_k|] = \sum_{k=0}^{n} \mathrm{E}[|U_k|]\mathrm{E}[|Y_k|] = \sum_{k=0}^{n} \mathrm{E}[|U_k|] < \infty \, .$$

Inoltre, sulla base della Proposizione 5.2.1, dal momento che la σ-algebra generata da Y_n è indipendente dalla σ-algebra $\mathcal{F}_{n-1}$ e che U_n è misurabile rispetto a $\mathcal{F}_{n-1}$, si ha $q.c.$

$$\begin{aligned} \mathrm{E}[M_n \mid \mathcal{F}_{n-1}] &= \mathrm{E}[M_{n-1} + U_n Y_n \mid \mathcal{F}_{n-1}] = \mathrm{E}[M_{n-1} \mid \mathcal{F}_{n-1}] + \mathrm{E}[U_n Y_n \mid \mathcal{F}_{n-1}] \\ &= M_{n-1} + U_n \mathrm{E}[Y_n \mid \mathcal{F}_{n-1}] = M_{n-1} + U_n \mathrm{E}[Y_n] = M_{n-1} + U_n(2p-1) \, . \end{aligned}$$

Quindi, a prescindere dalla strategia di gioco adottata dal giocatore, il p.a. M è una martingala se $p = \frac{1}{2}$, una super-martingala se $p < \frac{1}{2}$ e una sub-martingala se $p > \frac{1}{2}$. Di conseguenza, per qualsiasi strategia di gioco considerata, un gioco equo rimane equo, un gioco sfavorevole rimane sfavorevole e un gioco favorevole rimane favorevole. Questo esempio mostra che è impossibile determinare una strategia di gioco per cui un gioco d'azzardo sfavorevole può essere trasformato in un gioco favorevole. □

Di seguito vengono dati due esempi di importanti martingale. Ulteriori esempi di martingale, super-martingale e sub-martingale verranno considerati nell'analisi delle passeggiate aleatorie nella Sezione 9.3 e del moto Browniano nella Sezione 9.4.

• **Esempio 9.2.3.** (Martingala di Doob) Si consideri una v.a. X con $\mathrm{E}[|X|] < \infty$ e si consideri il p.a. $M = (M_t)_{t \in \mathbb{T}}$ tale che

$$M_t = \mathrm{E}[X \mid \mathcal{F}_t] \, ,$$

dove $(\mathcal{F}_t)_{t \in \mathbb{T}}$ è una filtrazione. Tenendo presente la disuguaglianza di Jensen condizionata (Teorema 5.2.3), per ogni $s < t$ si ha

$$\mathrm{E}[|M_t|] = \mathrm{E}[|\mathrm{E}[X \mid \mathcal{F}_t]|] \leq \mathrm{E}[\mathrm{E}[|X| \mid \mathcal{F}_t]] = \mathrm{E}[|X|] < \infty \, .$$

Inoltre, sulla base del Teorema 5.2.6 si ha $\{\mathrm{E}[\mathrm{E}[X \mid \mathcal{F}_t] \mid \mathcal{F}_s] = \mathrm{E}[X \mid \mathcal{F}_s]\}$ $q.c.$ se $\mathcal{F}_s \subset \mathcal{F}_t$, da cui si ha $q.c.$

$$\mathrm{E}[M_t \mid \mathcal{F}_s] = \mathrm{E}[\mathrm{E}[X \mid \mathcal{F}_t] \mid \mathcal{F}_s] = \mathrm{E}[X \mid \mathcal{F}_s] = M_s \, .$$

Quindi, il p.a. M è una martingala, anche detta *martingala di Doob*. Questo esempio evidenzia una procedura pratica per costruire una martingala. □

• **Esempio 9.2.4.** (Urna di Pólya) Sia data un'urna che contiene r palline rosse e b palline blu. Si consideri inoltre un esperimento aleatorio nel quale si estrae una pallina dall'urna e, dopo averne verificato il colore, si reimmette la pallina nell'urna con un'ulteriore pallina dello stesso colore. La procedura viene successivamente ripetuta con le stesse modalità. Questo esperimento aleatorio è detto *schema dell'urna di Pólya* dal momento che è stato proposto dal matematico George Pólya (1887-1985). Se X_n è la v.a. che rappresenta il numero di palline rosse nell'urna alla n-esima ripetizione dell'esperimento, si consideri la successione di v.a. $(X_n)_{n \geq 1}$. Si osservi che si può scrivere $X_n = r + \sum_{k=1}^{n} Z_k$, dove ogni v.a. Z_k è tale che $Z_k = 1$ se la pallina estratta alla k-esima estrazione è rossa e $Z_k = 0$ altrimenti. Dalla costruzione dell'esperimento aleatorio risulta immediato verificare che

$$\mathrm{E}[Z_k \mid X_{k-1} = x] = P(Z_k \mid X_{k-1} = x) = \frac{x}{k-1+r+b} \, ,$$

dove $x = r, r+1, \ldots, r+k-1$. Si consideri dunque il p.a. $M = (M_n)_{n\geq 0}$ tale che $M_n = \frac{X_n}{n+r+b}$ e $M_0 = \frac{r}{r+b}$ con $(\mathcal{F}_n)_{n\geq 0}$ filtrazione naturale. Per ogni n si ha

$$\mathrm{E}[|M_n|] = \mathrm{E}[M_n] = \frac{1}{n+r+b}\,\mathrm{E}[X_n] = \frac{1}{n+r+b}\left(r + \sum_{k=1}^{n}\mathrm{E}[Z_k]\right) \leq \frac{n+r}{n+r+b} < 1\,.$$

Inoltre, tenendo presente la Proposizione 5.2.1 e che $\mathrm{E}[Z_n \mid \mathcal{F}_{n-1}] = \mathrm{E}[Z_n \mid X_{n-1}]$, si ha $q.c.$

$$\begin{aligned}\mathrm{E}[M_n \mid \mathcal{F}_{n-1}] &= \frac{1}{n+r+b}\left(\mathrm{E}[X_{n-1} \mid \mathcal{F}_{n-1}] + \mathrm{E}[Z_n \mid \mathcal{F}_{n-1}]\right)\\ &= \frac{1}{n+r+b}\left(X_{n-1} + \frac{X_{n-1}}{n-1+r+b}\right) = \frac{1}{n-1+r+b}\,X_{n-1} = M_{n-1}\,.\end{aligned}$$

Procedendo in modo iterativo, per $m \in \mathbb{N}$ con $m < n$, si ha dunque $q.c.$

$$\mathrm{E}[X_n \mid \mathcal{F}_m] = \frac{1}{m+r+b}\,X_m = M_m\,,$$

ovvero dalla Definizione 9.2.1 il p.a. M è una martingala. □

Di seguito viene introdotta la definizione di tempo di arresto. Anche se il tempo di arresto può essere considerato in generale per un qualsiasi p.a., questo concetto risulta particolarmente importante nell'ambito delle martingale.

Definizione 9.2.2. Si consideri lo spazio di probabilità $(\Omega, \mathcal{F}, P)$. Se $(\mathcal{F}_n)_{n\geq 0}$ è una filtrazione, un *tempo di arresto discreto* è un'applicazione $\tau : \Omega \to \mathbb{N} \cup \{\infty\}$, tale che l'evento $\{\tau \leq n\} \in \mathcal{F}_n$ per ogni $n \in \mathbb{N}$. Inoltre, se $(\mathcal{F}_t)_{t\in[0,\infty[}$ è una filtrazione, un *tempo di arresto continuo* è un'applicazione $\tau : \Omega \to [0,\infty[\cup \{\infty\}$, tale che l'evento $\{\tau \leq t\} \in \mathcal{F}_t$ per ogni $t \in [0,\infty[$. □

Intuitivamente, si può pensare ad un tempo d'arresto come al momento in cui si prende una decisione relativa ad un fenomeno aleatorio nel tempo (per esempio, la decisione di raddoppiare la scommessa ad un certo istante in una serie di giocate ad un gioco d'azzardo). I vincoli sugli eventi, ovvero $\{\tau \leq n\} \in \mathcal{F}_n$ e $\{\tau \leq t\} \in \mathcal{F}_t$, si basano sul fatto che tale decisione deve dipendere solo dalle informazioni disponibili fino a quel momento. Si osservi infine che un tempo di arresto non è in effetti una v.a. in senso proprio in quanto potrebbe sussistere che $P(\tau < \infty) < 1$, ovvero il tempo di arresto potrebbe non essere finito.

• **Esempio 9.2.5.** Si consideri di nuovo il p.a. $M = (M_n)_{n\geq 0}$ introdotto nell'Esempio 9.2.2 e sia dato il tempo di arresto

$$\tau = \min(\{n \in \mathbb{N} : Y_n = 1\})\,.$$

Nell'interpretazione in termini di gioco d'azzardo, il tempo di arresto τ può essere visto come l'istante in cui il giocatore ottiene la prima vincita. Si noti che τ è in effetti un tempo di arresto, dal momento che $\mathcal{F}_n$ è la σ-algebra generata dal v.v.a. $(Y_1, \ldots, Y_n)^{\mathrm{T}}$ e

$$\{\tau = n\} = \{Y_1 = -1, \ldots, Y_{n-1} = -1, Y_n = 1\} \in \mathcal{F}_n\,,$$

per cui

$$\{\tau \leq n\} = \bigcup_{k=1}^{n}\{\tau = k\} \in \mathcal{F}_n\,.$$

Per quanto visto nella Sezione 6.3, la legge della v.a. $(\tau - 1)$ è Geometrica di parametro p, ovvero la relativa f.p. è data da

$$p_\tau(x) = pq^{x-1}\,\mathbf{1}_{\{1,2,\ldots\}}(x)\ .$$

Dunque, in questo caso si ha $P(\tau < \infty) = 1$. □

Definizione 9.2.3. Si consideri lo spazio di probabilità $(\Omega, \mathcal{F}, P)$ e sia $(\mathcal{F}_t)_{t\in\mathbb{T}}$ una filtrazione. In questo caso

$$\mathcal{F}_\tau = \{E \in \mathcal{F} : E \cap \{\tau < t\}, t \in \mathbb{T}\}$$

è detta *σ-algebra associata al tempo di arresto τ*. Inoltre, se $M = (M_t)_{t\in\mathbb{T}}$ è una martingala e τ è un tempo di arresto, allora posto $M_\tau = M(\tau(\omega), \omega)$, il p.a. $M^\tau = (M_{\tau\wedge t})_{t\in\mathbb{T}}$ è detta *martingala arrestata al tempo di arresto τ*.

Il prossimo Teorema riveste un'importanza notevole, dal momento che evidenzia che una martingala arrestata conserva ancora le caratteristiche di martingala.

Teorema 9.2.4. (Teorema del campionamento opzionale di Doob) *Sia $M = (M_t)_{t\in\mathbb{T}}$ una martingala e τ_1 e τ_2 tempi di arresto tali che $\{\tau_1 \le \tau_2 < T\}$ q.c. dove $T > 0$. Si ha q.c.*

$$\mathrm{E}[M_{\tau_2} \mid \mathcal{F}_{\tau_1}] = M_{\tau_1}\ .$$

In particolare, il p.a. $M^\tau = (M_{\tau\wedge t})_{t\in\mathbb{T}}$ è una martingala. Inoltre, se M è una super-martingala, si ha q.c.

$$\mathrm{E}[M_{\tau_2} \mid \mathcal{F}_{\tau_1}] \le M_{\tau_1}$$

e il p.a. M^τ è una super-martingala, mentre se M è una sub-martingala, si ha q.c.

$$\mathrm{E}[M_{\tau_2} \mid \mathcal{F}_{\tau_1}] \ge M_{\tau_1}$$

e il p.a. M^τ è una sub-martingala.

Dimostrazione. Si veda Revuz e Yor (2005, p.123). □

• **Esempio 9.2.6.** (Strategia di gioco con martingala) Si consideri il p.a. $M = (M_n)_{n\ge0}$ introdotto nell'Esempio 9.2.2 e si assuma inoltre una successione di v.a. $(U_n)_{n\ge1}$ tale che

$$U_n = 2^{n-1}\prod_{k=1}^{n-1}\mathbf{1}_{\{-1\}}(Y_k)\ .$$

Si consideri inoltre il tempo d'arresto τ introdotto nell'Esempio 9.2.5 e la martingala arrestata $M^\tau = (M_{\tau\wedge n})_{n\ge0}$. Dal Teorema 9.2.4, si ha dunque che il p.a. M^τ è una martingala se $p = \frac{1}{2}$, una super-martingala se $p < \frac{1}{2}$ o una sub-martingala se $p > \frac{1}{2}$. Nell'interpretazione in un ambito di gioco d'azzardo, la strategia di gioco basata sulla successione di v.a. $(U_n)_{n\ge1}$ è comunemente detta *martingala* (con una terminologia ovviamente fuorviante). In pratica, seguendo questa strategia, il giocatore scommette un'unità alla prima giocata. Nel caso in cui perda, il giocatore raddoppia la scommessa e continua in questo modo fino a quando non ottiene la prima vincita e conclude il gioco. Si noti che

$$\begin{aligned} M_{\tau\wedge n} &= \mathbf{1}_{\{1,\ldots,n\}}(\tau) - \sum_{k=1}^{n} U_k\,\mathbf{1}_{\{n+1,n+2,\ldots\}}(\tau) = \mathbf{1}_{\{1,\ldots,n\}}(\tau) - \sum_{k=1}^{n} 2^{k-1}\,\mathbf{1}_{\{n+1,n+2,\ldots\}}(\tau) \\ &= \mathbf{1}_{\{1,\ldots,n\}}(\tau) - (2^n - 1)\,\mathbf{1}_{\{n+1,n+2,\ldots\}}(\tau)\ . \end{aligned}$$

In pratica, $M_{\tau\wedge n} = M_\tau = 1$ se $\tau = n$ (ovvero il giocatore guadagna un'unità alla prima vincita concludendo il gioco), mentre $M_{\tau\wedge n} = M_n = -(2^n - 1)$ se $\tau > n$ (ovvero il giocatore perde $2^n - 1$ unità all'n-esima giocata precedente la prima vincita). Dunque, poichè $P(\tau < \infty) = 1$, si ha anche $P(M_\tau = 1) = 1$. Dal momento che questo risultato vale per ogni p, la martingala sembra apparentemente una strategia di gioco vincente (il giocatore termina il gioco con un'unità vinta). Tuttavia, si deve tenere presente che questa strategia presuppone una disponibilità illimitata di denaro e di tempo da parte del giocatore (ovviamente un'assunzione irrealistica). Inoltre, tenendo presente l'Esempio 9.2.5, si ha

$$\mathrm{E}[M_{\tau-1}] = \sum_{n=1}^{\infty} M_{n-1} P(\tau = n) = -\sum_{n=1}^{\infty} (2^{n-1} - 1) pq^{n-1} = -p \sum_{n=0}^{\infty} (2q)^n + 1 \,.$$

Dunque, risulta $\mathrm{E}[M_{\tau-1}] = -\infty$ per $q \geq \frac{1}{2}$, ovvero per $p \leq \frac{1}{2}$. Quindi, in questo caso la perdita attesa nella giocata prima della vincita finale è infinita. Sorprendentemente, questo risultato si ha anche per un gioco equo, ovvero anche nel caso in cui $p = \frac{1}{2}$. □

Teorema 9.2.5. (Disuguaglianze di Doob) *Sia $M = (M_n)_{t\in\mathbb{T}}$ una martingala. Per ogni $T > 0$ e $\lambda > 0$ vale la disuguaglianza*

$$P\left(\sup_{t\in[0,T]} |M_t| \geq \lambda \right) \leq \frac{1}{\lambda} \mathrm{E}[M_T \vee 0] \leq \frac{1}{\lambda} \mathrm{E}[|M_T|] \,.$$

Inoltre, per $p > 1$ e $q = \frac{p}{p-1}$ vale la disuguaglianza

$$\mathrm{E}[\sup_{t\in[0,T]} |M_t|^p] \geq q^p \mathrm{E}[|M_T|^p] \,.$$

Dimostrazione. Si veda Pascucci (2011, p.113). □

9.3. Passeggiate aleatorie

La passeggiata aleatoria è uno dei p.a. più semplici (anche se fondamentali) nella teoria della probabilità. Più esattamente, la passeggiata aleatoria è un processo a tempo discreto che viene formalmente definito come segue.

Definizione 9.3.1. Sia $(\Omega, \mathcal{F}, P)$ con $(\mathcal{F}_n)_{n\geq 0}$ uno spazio di probabilità con filtrazione. Sia inoltre $(Y_n)_{n\geq 1}$ una successione di v.a. indipendenti con la medesima legge, tali che $P(Y_n = 1) = p$ e $P(Y_n = -1) = q$ con $q = 1 - p$. Si dice *passeggiata aleatoria* il p.a. $X = (X_n)_{n\geq 0}$ dove $X_n = \sum_{k=1}^n Y_k$ e $\{X_0 = 0\}$ *q.c.* □

Evidentemente, vista l'indipendenza delle componenti della successione $(Y_n)_{n\geq 1}$, si può scegliere la filtrazione naturale $(\mathcal{F}_n)_{n\geq 0}$ tale che $\mathcal{F}_n = \sigma(\{Y_1, \ldots, Y_n\})$. I tempi $n \in \mathbb{N}$ sono anche detti *passi* della passeggiata aleatoria. Inoltre, nel caso particolare in cui $p = \frac{1}{2}$, la passeggiata aleatoria è detta *simmetrica*. La seguente Proposizione fornisce la legge del generico stato della passeggiata aleatoria.

Proposizione 9.3.2. *Sia $X = (X_n)_{n\geq 0}$ una passeggiata aleatoria. Per ogni $n \in \mathbb{N}$ la f.p. dello stato X_n è data da*

$$p_{X_n}(x) = \binom{n}{\frac{1}{2}(n+x)} p^{\frac{1}{2}(n+x)} q^{\frac{1}{2}(n-x)} \mathbf{1}_{\{-n,2-n,\ldots,n-2,n\}}(x) \,.$$

Inoltre, $\mathrm{E}[X_n] = n(p-q)$ *e* $\mathrm{Var}[X_n] = 4npq$.

Dimostrazione. Si osservi che sussiste $X_n = 2V_n - n$, dove $V_n = \sum_{k=1}^{n} Z_k$ e $(Z_n)_{n\geq 1}$ è una successione di v.a. indipendenti con legge di Bernoulli di parametro p. Per quanto visto nella Sezione 6.1, la legge della v.a. V_n è Binomiale $\mathcal{B}(n,p)$. Dunque, dal momento che la v.a. X_n è una trasformata lineare della v.a. V_n è immediato ottenerne la relativa f.p. Inoltre, per le proprietà del valore atteso si ha

$$\mathrm{E}[X_n] = 2\mathrm{E}[V_n] - n = 2np - n = n(p-q)$$

e

$$\mathrm{Var}[X_n] = 4\mathrm{Var}[V_n] = 4npq. \qquad \square$$

• **Esempio 9.3.1.** Si consideri il gioco d'azzardo della roulette introdotto nell'Esempio 2.6.3. Si supponga inoltre un giocatore ostinato (con disponibilità illimitate) che scommette un'unità sul rosso ad ogni giocata. La successione delle somme vinte (o perdute) dal giocatore è una passeggiata aleatoria con $p = \frac{18}{37}$ se la roulette è di tipo Monte Carlo (evitando di considerare per semplicità le regole basate su *la partage* o *en prison* quando si presenta lo zero). Sulla base della Proposizione 9.3.2, la probabilità di essere in attivo dopo $n = 20$ giocate è data da $\sum_{k=1}^{10} p_{X_{20}}(2k) = 0.365$, quella di essere in pareggio è data da $p_{X_{20}}(0) = 0.175$ e quella di essere in perdita è data da $\sum_{k=-10}^{-1} p_{X_{20}}(2k) = 0.460$. Inoltre, le medesime quantità dopo $n = 100$ giocate sono rispettivamente $\sum_{k=1}^{50} p_{X_{100}}(2k) = 0.355$, $p_{X_{100}}(0) = 0.077$ e $\sum_{k=-50}^{-1} p_{X_{100}}(2k) = 0.568$. $\square$

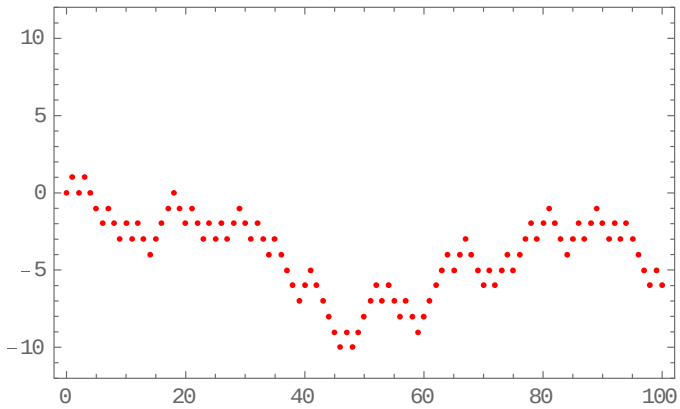


Figura 9.3.1. Una traiettoria della passeggiata aleatoria con $p = \frac{18}{37}$ e $n = 100$ passi.

La seguente Proposizione evidenzia che la passeggiata aleatoria è una martingala, una sub-martingala o una super-martingala a secondo dei valori assunti da p.

Proposizione 9.3.3. *La passeggiata aleatoria $X = (X_n)_{n\geq 0}$ è una martingala se $p = \frac{1}{2}$, una super-martingala se $p < \frac{1}{2}$ e una sub-martingala se $p > \frac{1}{2}$.*

Dimostrazione. Sulla base della Definizione 9.3.1, dal momento che $\mathrm{E}[|Y_k|] = 1$ si ha

$$\mathrm{E}[|X_n|] = \mathrm{E}[|\sum_{k=1}^{n} Y_k|] \leq \sum_{k=1}^{n} \mathrm{E}[|Y_k|] = n < \infty\,.$$

Inoltre, tenendo presente che $\{\mathrm{E}[X_n \mid \mathcal{F}_n] = X_n\}$ $q.c.$ sulla base della Proposizione 5.2.1 e che la σ-algebra generata da Y_n è indipendente dalla σ-algebra $\mathcal{F}_{n-1}$, si ha $q.c.$

$$\mathrm{E}[X_n \mid \mathcal{F}_{n-1}] = \mathrm{E}[X_{n-1} \mid \mathcal{F}_{n-1}] + \mathrm{E}[Y_n \mid \mathcal{F}_{n-1}] = X_{n-1} + \mathrm{E}[Y_n] = X_{n-1} + 2p - 1\,.$$

In modo analogo alla precedente relazione, si ha $q.c.$

$$\begin{aligned}\mathrm{E}[X_n \mid \mathcal{F}_{n-2}] &= \mathrm{E}[X_{n-2} \mid \mathcal{F}_{n-2}] + \mathrm{E}[Y_{n-1} \mid \mathcal{F}_{n-2}] + \mathrm{E}[Y_n \mid \mathcal{F}_{n-2}] \\ &= X_{n-2} + \mathrm{E}[Y_{n-1}] + \mathrm{E}[Y_n] = X_{n-2} + 2(2p-1)\,.\end{aligned}$$

Procedendo in modo iterativo, per $m \in \mathbb{N}$ con $m < n$, si ha dunque $q.c.$

$$\mathrm{E}[X_n \mid \mathcal{F}_m] = X_m + (n-m)(2p-1)\,.$$

Dalla precedente espressione e dalla Definizione 9.2.1 segue il risultato. □

• **Esempio 9.3.2.** Si consideri il gioco d'azzardo della testa e croce introdotto nell'Esempio 1.1.1 e si supponga di nuovo un giocatore ostinato (con disponibilità illimitate) che scommette un'unità sulla croce ad ogni giocata. Se la moneta è bilanciata, ovvero $p = \frac{1}{2}$, allora la successione delle somme vinte (o perdute) dal giocatore è una passeggiata aleatoria che è anche una martingala. La testa e croce è dunque un gioco equo. Nel caso della roulette considerato nell'Esempio 9.3.1, la passeggiata aleatoria è invece una super-martingala, dal momento che $p = \frac{18}{37} < \frac{1}{2}$. La roulette non è dunque un gioco d'azzardo vantaggioso. □

Risulta interessante ottenere alcune proprietà della traiettoria di una passeggiata aleatoria. In questo ambito, è conveniente introdurre le seguente definizione.

Definizione 9.3.4. Sia $X = (X_n)_{n\geq 0}$ una passeggiata aleatoria. Posto $n = 2k$ con $k \in \mathbb{N}$, la quantità $u_n = P(X_n = 0)$ dove

$$u_{2k} = P(X_{2k} = 0) = \binom{2k}{k}(pq)^k\,,$$

mentre $u_{2k+1} = P(X_{2k-1} = 0) = 0$ è detta *probabilità di ritorno all'origine* in n passi. □

Dunque, u_n non è nulla solo se n è pari. La seguente Proposizione fornisce alcune caratteristiche della successione delle probabilità di ritorno all'origine.

Proposizione 9.3.5. *La successione $u = (u_n)_{n\geq 0}$ ammette funzione generatrice data da*

$$G_u(t) = \sum_{n=0}^{\infty} t^n u_n = \frac{1}{\sqrt{1-4pqt^2}}\,.$$

Inoltre, posto $n = 2k$ con $k \in \mathbb{N}$, per $k \to \infty$ si ha

$$u_{2k} \sim \frac{1}{\sqrt{\pi k}}\,(4pq)^k\,.$$

Dimostrazione. Dal momento che sussiste la serie

$$\sum_{k=0}^{\infty}\binom{2k}{k}x^k = \frac{1}{\sqrt{1-4x}}\,,$$

la funzione generatrice della successione u è data da

$$G_u(t) = \sum_{n=0}^{\infty} t^n u_n = \sum_{k=0}^{\infty} t^{2k} u_{2k} = \sum_{k=0}^{\infty}\binom{2k}{k}(pqt^2)^k = \frac{1}{\sqrt{1-4pqt^2}}\,.$$

Inoltre, per $k \to \infty$ l'approssimazione di Stirling applicata al coefficiente binomiale nell'espressione di u_{2k} fornisce

$$\binom{2k}{k} \sim \frac{4^k}{\sqrt{\pi k}} ,$$

da cui si ottiene la seconda parte. □

Dalla Proposizione 9.3.5, per $p = \frac{1}{2}$ si ottiene in particolare che

$$u_{2k} \sim \frac{1}{\sqrt{\pi k}} .$$

Dunque, in modo controintuitivo, la probabilità di ritorno all'origine per una passeggiata aleatoria simmetrica tende ad annullarsi all'aumentare dei passi.

La seguente definizione introduce un particolare tempo di arresto nell'ambito delle passeggiate aleatorie.

Definizione 9.3.6. Sia $X = (X_n)_{n\geq 0}$ una passeggiata aleatoria. Il tempo di arresto

$$\tau_k = \min(\{n > 0 : X_n = k\}) ,$$

dove $k \in \mathbb{Z}$, è detto *tempo del primo passaggio per* k. In particolare,

$$\tau_0 = \min(\{n > 0 : X_n = 0\})$$

è detto *tempo del primo ritorno all'origine*. □

Evidentemente, il tempo del primo ritorno all'origine τ_0 può prendere valori solo sui numeri pari. La seguente Proposizione fornisce la distribuzione di τ_0 e le relative caratteristiche.

Proposizione 9.3.7. *Sia $X = (X_n)_{n\geq 0}$ una passeggiata aleatoria. La funzione generatrice della successione $v = (v_n)_{n\geq 1}$ con $v_n = P(\tau_0 = n)$ è data da*

$$G_v(t) = \sum_{n=1}^{\infty} t^n v_n = 1 - \sqrt{1 - 4pqt^2} .$$

Posto $n = 2k$ con $k \in \{1, 2, \dots\}$, si ha

$$v_{2k} = P(\tau_0 = 2k) = \binom{2k}{k} \frac{(pq)^k}{2k - 1} ,$$

mentre $v_{2k-1} = P(\tau_0 = 2k - 1) = 0$.

Dimostrazione. Gli eventi $\{\tau_0 = k\}$, $k = 1, 2, \dots, n$, sono incompatibili e dunque risulta

$$P(X_n = 0) = \sum_{k=1}^{n} P(X_n = 0 \mid \tau_0 = k) P(\tau_0 = k) .$$

Inoltre, si ha $P(X_n = 0 \mid \tau_0 = k) = P(X_{n-k} = 0)$ dal momento che, se si verifica $\{\tau_0 = k\}$, allora $\sum_{j=k}^{n} Y_j$ può essere considerato come il passo di ordine $(n - k)$ di una nuova passeggiata aleatoria. Dunque, dalla precedente espressione segue

$$u_n = \sum_{k=1}^{n} u_{n-k} v_k .$$

Quindi, si ha

$$G_u(t) = \sum_{n=0}^{\infty} t^n u_n = \sum_{n=0}^{\infty} \sum_{k=1}^{n} t^n u_{n-k} v_k \, .$$

Tenendo presente la definizione di convoluzione di funzioni generatrici e poichè $u_0 = 1$, dalla precedente espressione si ottiene

$$G_u(t) = 1 + G_u(t) G_v(t) \, ,$$

ovvero

$$G_v(t) = 1 - \frac{1}{G_u(t)}$$

e la prima parte segue dalla Proposizione 9.3.5. Inoltre, considerando l'espressione della serie binomiale, la funzione generatrice della successione v può essere scritta come

$$G_v(t) = 1 - \sum_{k=0}^{\infty} \binom{\frac{1}{2}}{k} (-4pqt^2)^k = \sum_{k=1}^{\infty} (-1)^{k-1} \binom{\frac{1}{2}}{k} (4pq)^k t^{2k} \, ,$$

da cui segue la seconda parte, tenendo presente l'identità

$$(-1)^{k-1} \binom{\frac{1}{2}}{k} 2^{2k} = \binom{2k}{k} \frac{1}{2k-1}$$

applicata all'espressione di v_{2k}. □

Si osservi che dalle Proposizioni 9.3.5 e 9.3.7 risulta in modo notevole

$$v_{2k} = \frac{1}{2k-1} \, u_{2k} \, ,$$

mentre per $k \to \infty$ si ha

$$v_{2k} \sim \frac{1}{\sqrt{4\pi k^3}} \, (4pq)^k \, .$$

Inoltre, si deve evidenziare che G_v non è in generale una funzione generatrice delle probabilità. In effetti, la successione v non costituisce in generale una legge essenziale, dal momento che dalla Proposizione 9.3.7 si ha

$$P(\tau_0 < \infty) = \sum_{n=1}^{\infty} v_n = G_v(1) = 1 - \sqrt{1-4pq} = 1 - \sqrt{(p-q)^2} = 1 - |p-q| \leq 1 \, .$$

• **Esempio 9.3.3.** Si consideri di nuovo il gioco d'azzardo della roulette. In questo caso, sulla base della Proposizione 9.3.7, la probabilità del tempo di ritorno all'origine dopo $n = 20$ giocate è data da $v_{20} = 0.009$. Inoltre, la probabilità che il tempo di ritorno all'origine sia finito risulta $P(\tau_0 < \infty) = \frac{36}{37}$. Dunque, in modo sorprendente per il giocatore ingenuo (anche se con disponibilità illimitate), la probabilità che il tempo di ritorno all'origine non sia finito è dato da $1 - P(\tau_0 < \infty) = \frac{1}{37}$. □

Dalla Proposizione 9.3.7 è evidente che $P(\tau_0 < \infty) = 1$ esclusivamente per una passeggiata aleatoria simmetrica. Dunque, τ_0 è una v.a. propria solamente per $p = \frac{1}{2}$. In questo caso, si ha il seguente risultato.

Proposizione 9.3.8. *Sia $X = (X_n)_{n\geq 0}$ una passeggiata aleatoria simmetrica. La f.p. della v.a. τ_0 è data da*

$$p_{\tau_0}(x) = \binom{x}{\frac{1}{2}x} \frac{2^{-x}}{x-1} \mathbf{1}_{\{2,4,\ldots\}}(x)$$

e $\mathrm{E}[\tau_0] = \infty$. Inoltre, posto $n = 2k$ con $k \in \mathbb{N}$ e data la successione $z = (z_n)_{n\geq 0}$ dove

$$z_{2k} = P(\tau_0 > 2k)$$

e $z_{2k+1} = 0$, si ha $z = u$.

Dimostrazione. Per $p = \frac{1}{2}$ la v.a. τ_0 possiede f.g. data da

$$G_{\tau_0}(t) = 1 - \sqrt{1-t^2}$$

e la prima parte segue immediatamente dalla Proposizione 9.3.7. Inoltre, si ha

$$G^{(1)}_{\tau_0}(t) = \frac{t}{\sqrt{1-t^2}} ,$$

e dunque

$$\mathrm{E}[\tau_0] = \lim_{t\to 1^-} G^{(1)}_{\tau_0}(t) = \infty .$$

Per quanto riguarda l'ultima parte, poichè $z_{2k} = \mathrm{E}[\mathbf{1}_{\{1,2,\ldots\}}(\tau_0 - 2k)]$, la funzione generatrice della successione z è data da

$$G_z(t) = \sum_{n=0}^{\infty} t^n z_n = \sum_{k=0}^{\infty} t^{2k}\, \mathrm{E}[\mathbf{1}_{\{1,2,\ldots\}}(\tau_0 - 2k)] = \mathrm{E}[\sum_{k=0}^{\infty} t^{2k}\, \mathbf{1}_{\{1,2,\ldots\}}(\tau_0 - 2k)]$$
$$= \mathrm{E}[\sum_{k=0}^{\frac{1}{2}\tau_0 - 1} t^{2k}] = \frac{1}{1-t^2}\, \mathrm{E}[1 - t^{\tau_0}] = \frac{1}{1-t^2}\, (1 - G_{\tau_0}(t)) .$$

Tenendo presente l'espressione di $G_{\tau_0}(t)$ e dalla Proposizione 9.3.5, si ha

$$G_z(t) = \frac{1}{\sqrt{1-t^2}} = G_u(t) ,$$

da cui segue la tesi. □

Dalla Proposizione 9.3.8 risulta quindi che, pur essendo $P(\tau_0 < \infty) = 1$ per $p = \frac{1}{2}$, si ha comunque $\mathrm{E}[\tau_0] = \infty$. Inoltre, dalla medesima Proposizione segue anche una relazione notevole fra le successioni u e z per una passeggiata aleatoria simmetrica. In effetti, da questa relazione risulta che la probabilità di non ritornare all'origine in $2k$ passi è pari alla probabilità di ritornare all'origine in $2k$ passi quando $p = \frac{1}{2}$, ovvero $P(\tau_0 > 2k) = P(X_{2k} = 0)$.

• **Esempio 9.3.4.** Si consideri il gioco d'azzardo della testa e croce analizzato nell'Esempio 9.3.2. In questo caso, la probabilità del tempo di ritorno all'origine dopo $n = 20$ giocate è data da $P(X_{20} = 0) = 0.176$, per cui risulta anche $P(\tau_0 > 20) = 0.176$. □

Il prossimo Teorema fornisce un risultato classico per la passeggiata aleatoria simmetrica, ovvero la cosiddetta *legge dell'arcoseno*. Se si considera i primi $2n$ passi di una passeggiata aleatoria, questa legge fornisce la distribuzione dell'ultimo tempo di ritorno all'origine.

Teorema 9.3.9. (Legge dell'arcoseno) *Sia $X = (X_n)_{n\geq 0}$ una passeggiata aleatoria simmetrica. Si consideri la v.a. $\varrho_n = \max(\{l \in \{1,\dots,n\} : X_l = 0\})$, detta ultimo tempo di ritorno all'origine in n passi. Posto $n = 2k$ con $k \in \mathbb{N}$, se $\alpha_{2l,2k} = P(\varrho_{2k} = 2l)$, si ha*

$$\alpha_{2l,2k} = u_{2l}u_{2k-2l} = \binom{2l}{l}\binom{2k-2l}{k-l}2^{-2k} ,$$

per $l = 0,\dots,k$, ovvero la f.p. della v.a. ϱ_n è data da

$$p_{\varrho_n}(x) = \binom{x}{\frac{1}{2}x}\binom{n-x}{\frac{1}{2}(n-x)}2^{-n}\,\mathbf{1}_{\{0,2,\dots,n\}}(x) .$$

Inoltre, risulta

$$\lim_n P(n^{-1}\varrho_n \leq x) = \frac{2}{\pi}\arcsin(\sqrt{x})\,\mathbf{1}_{[0,1]}(x) + \mathbf{1}_{]1,\infty[}(x) ,$$

ovvero $n^{-1}\varrho_n \xrightarrow{\mathcal{L}} Z$ per $n \to \infty$, dove Z è una v.a. con legge Beta di tipo $\mathcal{BE}(0,1,\frac{1}{2},\frac{1}{2})$ che ammette d.p. data da

$$f_Z(x) = \frac{1}{\pi\sqrt{x(1-x)}}\,\mathbf{1}_{[0,1]}(x) .$$

Dimostrazione. Per quanto riguarda la prima parte, tenendo presente che se si verifica $\{X_{2l} = 0\}$, allora i successivi passi possono essere considerati come una nuova passeggiata aleatoria, dalla Proposizione 9.3.8 si ha

$$\begin{aligned}\alpha_{2l,2k} &= P(X_{2l} = 0, X_{2l+2} \neq 0,\dots,X_{2k} \neq 0) = P(X_{2l} = 0)P(X_{2l+2} \neq 0,\dots,X_{2k} \neq 0 \mid X_{2l} = 0)\\ &= P(X_{2l} = 0)P(X_{2l+2} \neq 0,\dots,X_{2k} \neq 0) = P(X_{2l} = 0)P(\tau_0 > 2k-2l) = u_{2l}u_{2k-2l} .\end{aligned}$$

Per quanto riguarda la seconda parte, dalla Proposizione 9.3.5 si ha

$$\alpha_{2l,2k} \sim \frac{1}{\pi\sqrt{l(k-l)}} .$$

Dunque, posto $x = \frac{l}{k}$ per $l = 0,\dots,k$, si ha

$$P(n^{-1}\varrho_n \leq x) \sim \sum_{j=0}^{kx}\frac{1}{\pi\sqrt{j(k-j)}} ,$$

da cui, tenendo presente la definizione dell'integrale di Riemann, si ottiene

$$\lim_n P(n^{-1}\varrho_n \leq x) = \lim_k \sum_{j=0}^{kx}\frac{1}{\pi\sqrt{j(k-j)}} = \int_0^x \frac{1}{\pi\sqrt{u(1-u)}}\,du .$$

La seconda parte segue dunque dalla definizione di convergenza in legge e dai risultati della Sezione 6.9. □

Si osservi che $\alpha_{2l,2k} = \alpha_{2k-2l,2k}$. Inoltre, si deve sottolineare che ϱ_n non è un tempo di arresto. La legge dell'arcoseno fornisce un risultato sulla legge della v.a. ϱ_n che può risultare controintuitiva. In effetti, la f.p. della v.a. ϱ_n mostra la cosiddetta "forma ad U" (si veda la Figura 9.3.2), così come la d.p. della v.a. limite della successione di v.a. $(n^{-1}\varrho_n)_{n\geq 1}$ mostra la medesima morfologia. In effetti, anche se $\mathrm{E}[\varrho_n] = \frac{1}{2}n$, la probabilità del tempo dell'ultimo ritorno all'origine in n passi è molto più elevata per valori prossimi a 0 e a n piuttosto che a $\frac{1}{2}n$. Dal momento che la v.a. limite può essere

interpretata come la proporzione di tempo dall'ultimo ritorno all'origine, la legge dell'arcoseno è ancora più sorprendente per $n \to \infty$.

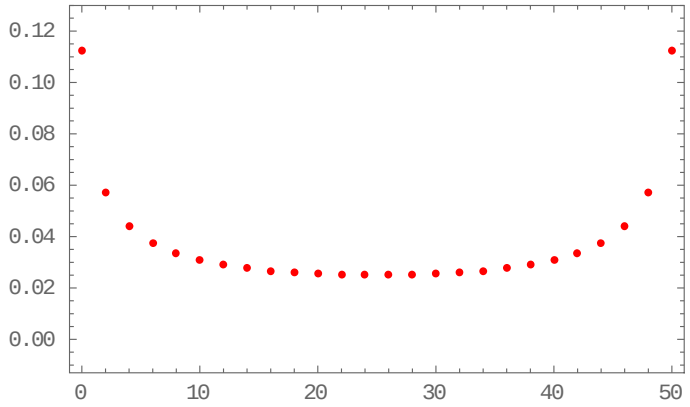


Figura 9.3.2. Funzione di probabilità per la legge dell'arcoseno con $n = 50$.

Per concludere, viene considerato un risultato classico nella Teoria della Probabilità, ovvero la cosiddetta *rovina del giocatore d'azzardo*. In questo problema, un giocatore ostinato con capitale iniziale pari ad a unità fronteggia in giocate ripetute un secondo giocatore con capitale iniziale pari a b unità. Il gioco continua fino a quando il primo giocatore esaurisce il proprio capitale o riesce a vincere il capitale del secondo giocatore. In questo caso, l'obiettivo è quello di determinare la probabilità di rovina, che viene ottenuta nel seguente risultato.

Teorema 9.3.10. (La rovina del giocatore d'azzardo) *Sia $X = (X_n)_{n\geq 0}$ una passeggiata aleatoria. Se $a, b \in \mathbb{N}$, si consideri il tempo di arresto $\tau = \min(\tau_{-a}, \tau_b)$. Si ha $P(\tau < \infty) = 1$. Inoltre, per $p \neq \frac{1}{2}$*

$$P(X_\tau = -a) = 1 - P(X_\tau = b) = \frac{1-\rho^b}{\rho^{-a}-\rho^b} ,$$

dove $\rho = \frac{q}{p}$, mentre se $p = \frac{1}{2}$

$$P(X_\tau = -a) = 1 - P(X_\tau = b) = \frac{b}{a+b} .$$

Dimostrazione. Per quanto riguarda la prima parte, si ponga $c = a + b$ e si definisca il tempo di arresto

$$Z = \min(\{n \in \{1, 2, \dots\} : Y_{(n-1)c+1} = 1, \dots, Y_{(n-1)c+c} = 1\}) .$$

Il tempo di arresto Z è quindi dato dal primo passo della passeggiata dopo il quale si verificano c vincite del primo giocatore. Dal momento che per n fissato si ha

$$P(Y_{(n-1)c+1} = 1, \dots, Y_{(n-1)c+c} = 1) = p^c ,$$

la legge della v.a. Z è Geometrica di parametro p^c e dunque $\mathrm{E}[Z] < \infty$. Inoltre, $\{\tau < cZ\}$ *q.c.* dal momento che τ si deve necessariamente realizzare nell'insieme $\{(n-1)c+1, \dots, (n-1)c+c\}$ se $Z = n$. Dunque, risulta $\mathrm{E}[\tau] < c\mathrm{E}[Z] < \infty$, da cui si ha anche $P(\tau < \infty) = 1$. Per quanto riguarda la seconda parte per $p \neq \frac{1}{2}$, si noti che $M = (M_n)_{n\geq 0}$ con $M_n = \rho^{X_n}$ è una martingala. In effetti, dal momento che $\mathrm{E}[\rho^{Y_k}] = 1$ per $k = 1, \dots, n$, si ha

$$\mathrm{E}[|M_n|] = \mathrm{E}[\prod_{k=1}^{n} \rho^{Y_k}] = \prod_{k=1}^{n} \mathrm{E}[\rho^{Y_k}] = 1 < \infty ,$$

mentre risulta $q.c.$

$$\mathrm{E}[M_n \mid \mathcal{F}_{n-1}] = \mathrm{E}[\rho^{Y_n} M_{n-1} \mid \mathcal{F}_{n-1}] = \mathrm{E}[\rho^{Y_n}]\mathrm{E}[M_{n-1} \mid \mathcal{F}_{n-1}] = M_{n-1}$$

e dunque, in modo iterativo, per $m \in \mathbb{N}$ con $m < n$, si ha $q.c.$

$$\mathrm{E}[X_n \mid \mathcal{F}_m] = X_m \, .$$

Quindi, dal Teorema 9.2.4 si ha che

$$\mathrm{E}[M_\tau] = \mathrm{E}[M_0] = 1 < \infty \, .$$

Dal momento che i due eventi

$$E = \{X_\tau = -a\} = \{M_\tau = \rho^{-a}\} = \{\tau_{-a} < \tau_b\}$$

e

$$F = \{X_\tau = b\} = \{M_\tau = \rho^{b}\} = \{\tau_{-a} > \tau_b\} \, ,$$

sono incompatibili ed esaustivi, si ha

$$\mathrm{E}[M_\tau] = \mathrm{E}[M_\tau \mid E]P(E) + \mathrm{E}[M_\tau \mid F]P(F) = \rho^{-a}P(X_\tau = -a) + \rho^{b}P(X_\tau = b) = 1$$

e la tesi segue tenendo presente che $P(X_\tau = -a) + P(X_\tau = b) = 1$. Nel caso che $p = \frac{1}{2}$, dalla Proposizione 9.3.3 segue che X è una martingala. Dunque, dal Teorema 9.2.4 si ha che

$$\mathrm{E}[X_\tau] = \mathrm{E}[X_0] = 0 < \infty \, .$$

In modo simile a quanto visto in precedenza, si ha

$$\mathrm{E}[X_\tau] = \mathrm{E}[X_\tau \mid E]P(E) + \mathrm{E}[X_\tau \mid F]P(F) = (-a)P(X_\tau = -a) + bP(X_\tau = b) = 0 \, ,$$

da cui segue la seconda parte. □

Dal precedente Teorema, si evidenzia che il gioco termina $q.c.$ dal momento che $P(\tau < \infty) = 1$ a prescindere dai valori di a e b e di p. Inoltre, nel caso che il secondo giocatore sia infinitamente ricco, ovvero se $b \to \infty$, allora si ha

$$P(X_\tau = -a) = 1 \, ,$$

se $\rho \geq 1$, mentre

$$P(X_\tau = -a) = \rho^{a} \, ,$$

se $\rho < 1$. Dunque, sorprendentemente, la rovina del primo giocatore contro un giocatore infinitamente ricco avviene $q.c.$ anche se il gioco è equo.

• **Esempio 9.3.5.** Si consideri di nuovo il gioco d'azzardo della roulette. Come è stato visto nell'Esempio 9.3.2, la roulette non è un gioco equo essendo $p = \frac{18}{37}$ e inoltre la casa da gioco dispone solitamente di capitali molto più elevati dei giocatori, ovvero b è molto più grande di a. Quindi, $P(X_\tau = -a)$ è prossimo ad uno nei casi usuali. Tuttavia, supponendo che $b = a$, ovvero che la casa da gioco disponga dello stesso capitale del giocatore ostinato, per $a = 50$ si ha $P(X_\tau = -50) = 0.937$ e per $a = 100$ si ha $P(X_\tau = -100) = 0.996$. Inoltre, supponendo che $b = \frac{a}{2}$, ovvero che la casa da gioco disponga della metà del capitale rispetto al giocatore ostinato, per $a = 50$ si ha $P(X_\tau = -50) = 0.754$ e per $a = 100$ si ha $P(X_\tau = -100) = 0.933$. Dunque, anche se il gioco della roulette è solo leggermente sfavorevole, questo fatto conferisce comunque un vantaggio determinante alla casa da gioco. □

9.4. Moto Browniano

Il moto Browniano è uno dei p.a. fondamentali nella Teoria della Probabilità. Questo p.a. prende nome dal botanico Robert Brown (1773-1858), che per primo aveva notato il moto erratico delle particelle nei fluidi. Il probabilista Norbert Wiener (1894-1964) ha determinato numerose proprietà del relativo modello matematico e per questo motivo il p.a. è anche detto processo di Wiener.

Figura 9.4.1. Norbert Wiener (1894-1964).

Definizione 9.4.1. Sia $(\Omega, \mathcal{F}, P)$ con $(\mathcal{F}_t)_{t\in[0,\infty[}$ uno spazio di probabilità con filtrazione. Il *moto Browniano* o *processo di Wiener* è un p.a. $B = (B_t)_{t\in[0,\infty[}$ tale che
i) $P(B_0 = 0) = 1$;
ii) per ogni $0 \leq s < t$, la legge della v.a. $B_t - B_s$ è Normale $\mathcal{N}(0, t-s)$;
iii) il p.a. B ha incrementi indipendenti, ovvero per ogni $0 \leq t_1 < \ldots < t_n$ le v.a. $B_{t_1}, B_{t_2} - B_{t_1}, \ldots, B_{t_n} - B_{t_{n-1}}$ sono indipendenti;
iv) il p.a. B ha traiettorie continue $q.c.$ □

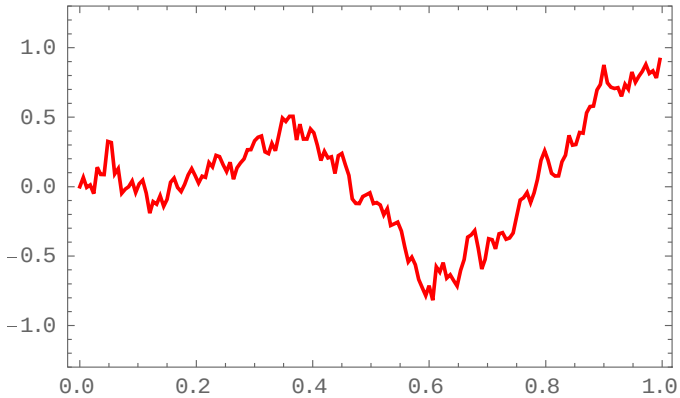


Figura 9.4.2. Una traiettoria del moto Browniano.

Per quanto riguarda la costruzione del moto Browniano, si può consultare Kuo (2006, p.23), dove a tale fine sono analizzati tre principali metodi. Il moto Browniano può essere anche ottenuto come limite in legge di una versione generale di passeggiata aleatoria "scalata" i cui passi tendono a zero. Più esattamente, si ha il seguente Teorema che prende il nome dal matematico Monroe David Donsker (1924-1991).

Teorema 9.4.2. (Teorema di Donsker) *Sia* $(\Omega, \mathcal{F}, P)$ *con* $(\mathcal{F}_t)_{t\in[0,\infty[}$ *uno spazio di probabilità con filtrazione. Sia inoltre* $(Y_n)_{n\geq 1}$ *una successione di v.a. indipendenti con la medesima legge, tali che* $\mathrm{E}[Y_n] = 0$ *e* $\mathrm{Var}[Y_n] = 1$ *e sia* $X_n = \sum_{k=1}^{n} Y_k$ *e* $X_0 = 0$ *q.c. Se* $Z_n = (Z_{n,t})_{t\in[0,1]}$ *è una passeggiata aleatoria scalata tale che*

$$Z_{n,t} = \frac{1}{\sqrt{n}} \left(X_{\lfloor nt \rfloor} + (nt - \lfloor nt \rfloor) Y_{\lfloor nt \rfloor + 1} \right) ,$$

dove $\lfloor x \rfloor$ *rappresenta il più grande intero minore o uguale a* x*, allora* $Z_n \xrightarrow{\mathcal{L}} B$ *per* $n \to \infty$*, dove* $B = (B_t)_{t\in[0,1]}$.

Dimostrazione. Si veda Capasso (2015, p.430). □

In particolare, il Teorema di Donsker vale per la passeggiata aleatoria simmetrica, per cui si ha $P(Y_n = 1) = P(Y_n = -1) = \frac{1}{2}$. Il Teorema di Donsker può essere considerato come la versione funzionale del Teorema Centrale del Limite (Teorema 8.4.1). Molte proprietà del moto Browniano possono essere quindi ottenute nel "limite" da quelle della passeggiata aleatoria simmetrica (si veda Bhattacharya e Waymire, 2016).

Di seguito vengono considerate una serie di Proposizioni che forniscono alcune delle principali proprietà del moto Browniano.

Proposizione 9.4.3. *Sia dato il moto Browniano* $B = (B_t)_{t\in[0,\infty[}$*. Per ogni* $t > 0$*, la v.a.* B_t *si distribuisce con legge Normale* $\mathcal{N}(0, t)$*. Inoltre, per ogni* $s, t \geq 0$ *si ha* $\mathrm{Cov}[B_s, B_t] = \min(s, t)$.

Dimostrazione. Dall'assunzione *i*) della Definizione 9.4.1 si ottiene che $B_t \overset{\mathcal{L}}{=} B_t - B_0$ e dunque la prima parte segue dalle assunzioni *i*) e *ii*) della Definizione 9.4.1. Per quanto riguarda la seconda parte, banalmente si ha che $B_t \overset{\mathcal{L}}{=} (B_t - B_s) + B_s$ per $s < t$. Tenendo presente l'assunzione *iii*) della Definizione 9.4.1 si ha inoltre $\mathrm{E}[(B_t - B_s)B_s] = 0$ e dunque

$$\mathrm{Cov}[B_s, B_t] = \mathrm{E}[B_s B_t] = \mathrm{E}[(B_t - B_s)B_s + B_s^2] = \mathrm{E}[B_s^2] = s \, .$$

La seconda parte segue immediatamente scambiando gli indici s e t. □

La seguente Proposizione mostra che il moto Browniano è un p.a. Gaussiano, ovvero un p.a. per cui la legge di ogni scelta finita di stati è Normale Multivariata.

Proposizione 9.4.4. *Sia dato il moto Browniano* $B = (B_t)_{t\in[0,\infty[}$*. Se* $t_1 < \ldots < t_n$*, allora la legge del v.v.a.* $(B_{t_1}, \ldots, B_{t_n})^{\mathrm{T}}$ *è Normale Multivariata* $\mathcal{N}_n(0, \Sigma)$ con $\Sigma = (\min(t_k, t_j))$.

Dimostrazione. Innanzitutto, si osservi che per un vettore di costanti $v = (v_1, \ldots, v_n)^{\mathrm{T}}$ risulta

$$v_1 B_{t_1} + \ldots + v_n B_{t_n} = (v_1 + \ldots + v_n) B_{t_1} + (v_2 + \ldots + v_n)(B_{t_2} - B_{t_1}) + \ldots + v_n (B_{t_n} - B_{t_{n-1}})$$

e che le v.a. $B_{t_1}, B_{t_2} - B_{t_1}, \ldots, B_{t_n} - B_{t_{n-1}}$ sono indipendenti sulla base della assunzione *iii*) della Definizione 9.4.1. Dunque, la f.c.m. del v.v.a. $(B_{t_1}, \ldots, B_{t_n})$ è data da

$$\varphi(v) = \mathrm{E}[\prod_{k=1}^{n} e^{\mathrm{i} v_k B_{t_k}}] = \mathrm{E}[e^{\mathrm{i}(v_1 + \ldots + v_n) B_{t_1}}] \mathrm{E}[e^{\mathrm{i}(v_2 + \ldots + v_n)(B_{t_2} - B_{t_1})}] \times \cdots \times \mathrm{E}[e^{\mathrm{i} v_n (B_{t_n} - B_{t_{n-1}})}] \, .$$

Inoltre, per la Proposizione 9.4.3 la v.a. B_{t_1} è distribuita con legge Normale $\mathcal{N}(0, t_1)$, mentre dalla assunzione *ii*) della Definizione 9.4.1 la v.a. $B_{t_k} - B_{t_{k-1}}$ è distribuita con legge Normale $\mathcal{N}(0, t_k - t_{k-1})$ per $k = 2, \ldots, n$. Dunque, tenendo presente l'Esempio 7.1.6, risulta

$$\varphi(v) = e^{-\frac{1}{2}t_1(v_1+\ldots+v_n)^2} e^{-\frac{1}{2}(t_2-t_1)(v_2+\ldots+v_n)^2} \times \cdots \times e^{-\frac{1}{2}(t_n-t_{n-1})v_n^2}$$
$$= e^{-\frac{1}{2}(\sum_{k=1}^n t_k v_k^2 + 2\sum_{k=1}^n \sum_{j=k+1}^n t_k v_k v_j)} = e^{-\frac{1}{2}v^{\mathrm{T}}\Sigma v}$$

e quindi, sulla base dell'Esempio 7.3.1, segue la tesi. □

La seguente Proposizione mostra che il moto Browniano è in effetti una martingala.

Proposizione 9.4.5. *Sia dato il moto Browniano $B = (B_t)_{t\in[0,\infty[}$. Il p.a. B è una martingala.*

Dimostrazione. Sulla base della Proposizione 9.4.3 e dalla disuguaglianza di Lyapunov (Teorema 4.3.2), per ogni t si ha

$$\mathrm{E}[|B_t|] \leq \sqrt{\mathrm{E}[B_t^2]} = \sqrt{t} < \infty \,.$$

Inoltre, tenendo presente la dimostrazione della Proposizione 9.4.3 e le assunzioni *ii*) e *iii*) della Definizione 9.4.1, risulta

$$\mathrm{E}[B_t \mid \mathcal{F}_s] = \mathrm{E}[B_t - B_s + B_s \mid \mathcal{F}_s] = \mathrm{E}[B_t - B_s \mid \mathcal{F}_s] + \mathrm{E}[B_s \mid \mathcal{F}_s] = \mathrm{E}[B_t - B_s] + B_s = B_s \,,$$

da cui segue la tesi. □

Le seguenti Proposizioni evidenziano che se il moto Browniano viene riflesso, o traslato, o scalato, o invertito rispetto al tempo, si ottiene di nuovo un moto Browniano.

Proposizione 9.4.6. *Sia dato il moto Browniano $B = (B_t)_{t\in[0,\infty[}$. Posto $X_t = -B_t$ il p.a. $X = (X_t)_{t\in[0,\infty[}$ è un moto Browniano.*

Dimostrazione. Innanzitutto, si ha $P(X_0 = 0) = P(-B_0 = 0) = 1$. Inoltre, per ogni $0 \leq s < t$, sussiste

$$X_t - X_s = -(B_t - B_s) \stackrel{\mathcal{L}}{=} B_t - B_s$$

dal momento che la v.a. $(B_t - B_s)$ è simmetrica rispetto all'origine. Dunque, dall'assunzione *ii*) della Definizione 9.4.1, la legge della v.a. $(X_t - X_s)$ è Normale $\mathcal{N}(0, t-s)$. Dalla Proposizione 3.7.3 si ha che trasformate di v.a. indipendenti sono indipendenti, per cui le v.a. $X_t - X_s = -(B_t - B_s)$ e $X_s = -B_s$ risultano indipendenti sulla base dell'assunzione *ii*) della Definizione 9.4.1. In modo analogo, per ogni $0 \leq t_1 < \ldots < t_n$ le v.a. $X_{t_1}, X_{t_2} - X_{t_1}, \ldots, X_{t_n} - X_{t_{n-1}}$sono indipendenti. Infine, è evidente che il p.a. X ha traiettorie continue $q.c.$ Dunque, dalla Definizione 9.4.1 il p.a. X è un moto Browniano. □

Proposizione 9.4.7. *Sia dato il moto Browniano $B = (B_t)_{t\in[0,\infty[}$. Se $u \geq 0$ e posto $X_t = B_{t+u} - B_u$ il p.a. $X = (X_t)_{t\in[0,\infty[}$ è un moto Browniano.*

Dimostrazione. Risulta evidente che $P(X_0 = 0) = 1$. Inoltre, per ogni $0 \leq s < t$, la v.a. $X_t - X_s = B_{t+u} - B_{s+u}$ si distribuisce con legge Normale $\mathcal{N}(0, t-s)$ sulla base dell'assunzione *ii*) della Definizione 9.4.1. Tenendo presente la medesima assunzione, è immediato verificare che le v.a. $X_{t_1}, X_{t_2} - X_{t_1}, \ldots, X_{t_n} - X_{t_{n-1}}$ sono indipendenti. Infine, è evidente che il p.a. X ha traiettorie continue $q.c.$ Dunque, dalla Definizione 9.4.1 il p.a. X è un moto Browniano. □

Proposizione 9.4.8. *Sia dato il moto Browniano $B = (B_t)_{t\in[0,\infty[}$. Se $c \neq 0$ e posto $X_t = cB_{t/c^2}$ il p.a. $X = (X_t)_{t\in[0,\infty[}$ è un moto Browniano.*

Dimostrazione. Si ha $P(X_0 = 0) = P(cB_0 = 0) = 1$. Inoltre, per ogni $0 \leq s < t$, risulta $X_t - X_s = c(B_{t/c^2} - B_{s/c^2})$. Dunque, dall'assunzione *ii*) della Definizione 9.4.1 e dalla Proposizione 9.4.3, si ottiene

$$\mathrm{E}[X_t - X_s] = c(\mathrm{E}[B_{t/c^2}] - \mathrm{E}[B_{s/c^2}]) = 0$$

e

$$\mathrm{Var}[X_t - X_s] = c^2(\mathrm{Var}[B_{t/c^2}] + \mathrm{Var}[B_{s/c^2}] - 2\,\mathrm{Cov}[B_{t/c^2}, B_{s/c^2}]) = t - s\,.$$

Tenendo presente l'Esempio 7.3.2, la v.a. $(X_t - X_s)$ si distribuisce con legge Normale $\mathcal{N}(0, t-s)$. Dal momento che trasformate di v.a. indipendenti sono indipendenti (si veda la Proposizione 3.7.3), le v.a. $X_t - X_s = c(B_{t/c^2} - B_{s/c^2})$ e $X_s = cB_{s/c^2}$ risultano indipendenti sulla base dell'assunzione *ii*) della Definizione 9.4.1. In modo analogo, le v.a. $X_{t_1}, X_{t_2} - X_{t_1}, \ldots, X_{t_n} - X_{t_{n-1}}$ risultano indipendenti per ogni $0 \leq t_1 < \ldots < t_n$. Infine, è evidente che il p.a. X ha traiettorie continue $q.c.$ Dunque, dalla Definizione 9.4.1, il p.a. X è un moto Browniano. □

Proposizione 9.4.9. *Sia dato il moto Browniano $B = (B_t)_{t\in[0,\infty[}$. Posto $X_t = tB_{1/t}$ se $t > 0$ e $X_0 = 0$ q.c. il p.a. $X = (X_t)_{t\in[0,\infty[}$ è un moto Browniano.*

Dimostrazione. Evidentemente si ha $P(X_0 = 0) = 1$ dalle assunzioni fatte. Dalla Proposizione 9.4.3, la v.a. $B_{1/t}$ è distribuita con legge Normale $\mathcal{N}(0, \frac{1}{t})$ e dunque la v.a. X_t è distribuita con legge Normale $\mathcal{N}(0, t)$, mentre per ogni $0 \leq s < t$ si ha

$$\mathrm{E}[X_t - X_s] = t\,\mathrm{E}[B_{1/t}] - s\,\mathrm{E}[B_{1/s}] = 0$$

e

$$\mathrm{Var}[X_t - X_s] = t^2\,\mathrm{Var}[B_{1/t}] + s^2\,\mathrm{Var}[B_{1/s}] - 2ts\,\mathrm{Cov}[B_{1/t}, B_{1/s}] = t - s\,.$$

Dal momento il v.v.a. $(B_{1/t}, B_{1/s})$ è distribuito con legge Normale Multivariata sulla base della Proposizione 9.4.4 e poichè $X_t - X_s = tB_{1/t} - sB_{1/s}$ è una combinazione lineare delle componenti di questo v.v.a., si ha dunque che la v.a. $(X_t - X_s)$ è distribuita con legge Normale $\mathcal{N}(0, t-s)$. Inoltre, tenendo presente che $\mathrm{E}[B_{1/t}B_{1/s}] = \min(\frac{1}{t}, \frac{1}{s}) = \frac{1}{t}$, risulta

$$\mathrm{Cov}[X_t - X_s, X_s] = \mathrm{E}[(X_t - X_s)X_s] = \mathrm{E}[X_tX_s] - \mathrm{E}[X_s^2] = st\,\mathrm{E}[B_{1/t}B_{1/s}] - s^2\,\mathrm{E}[B_{1/s}^2] = 0\,.$$

Il v.v.a. $(X_s, X_t - X_s)$ si distribuisce con legge Normale Multivariata essendo una trasformata affine del v.v.a. $(B_{1/t}, B_{1/s})$ ed inoltre le sue componenti sono indipendenti tenendo presente la precedente relazione. In modo analogo, le v.a. $X_{t_1}, X_{t_2} - X_{t_1}, \ldots, X_{t_n} - X_{t_{n-1}}$ risultano indipendenti per ogni $0 \leq t_1 < \ldots < t_n$. Dal momento che il moto Browniano è continuo $q.c.$, risulta

$$P(\lim_{t\to 0^+} X_t = 0) = P(\lim_{t\to 0^+} tB_{1/t} = 0) = 1$$

e quindi il p.a. X ha traiettorie continue $q.c.$ Dunque, dalla Definizione 9.4.1, il p.a. X è un moto Browniano. □

I prossimi Teoremi evidenziano che le traiettorie del moto Browniano sono molto "irregolari". In effetti, anche se le traiettorie di questo p.a. sono continue $q.c.$ per definizione, queste sono non differenziabili $q.o.$ Inoltre, la variazione di un moto Browniano è infinita $q.c.$, mentre la rispettiva variazione quadratica è finita (per una discussione approfondita di queste quantità si veda Klebaner, 2005)

Teorema 9.4.10. *Sia dato il moto Browniano $B = (B_t)_{t\in[0,\infty[}$. Allora, $q.c.$ le traiettorie del moto Browniano non sono differenziabili $q.o.$*

Dimostrazione. Si veda Capasso (2015, p.138). □

Teorema 9.4.11. *Sia dato un moto Browniano $B = (B_t)_{t\in[0,T]}$. Si consideri inoltre la partizione $0 = t_{0,n} < t_{1,n} < \ldots < t_{n,n} = T$ e le successioni di v.a. $(V_n)_{n\geq 1}$ e $(V_{n,2})_{n\geq 1}$, dove*

$$V_n = \sum_{k=0}^{n-1} |B_{t_{k+1,n}} - B_{t_{k,n}}|$$

e

$$V_{n,2} = \sum_{k=0}^{n-1} (B_{t_{k+1,n}} - B_{t_{k,n}})^2 .$$

Allora, se $\lim_n \max(t_{k+1,n} - t_{k,n}) = 0$*, si ha*

$$P(\lim_n V_n = \infty) = 1$$

e

$$\lim_n \mathrm{E}[(V_{n,2} - T)^2] = 0 .$$

Dimostrazione. Si veda Klebaner (2005, p.63). □

La seguente definizione introduce il cosiddetto tempo di primo passaggio del moto Browniano.

Definizione 9.4.12. Sia $B = (B_t)_{t\in[0,\infty[}$ il moto Browniano. Il tempo di arresto

$$\tau_a = \inf\{t \in]0,\infty[: B_t = a\} ,$$

dove $a \in \mathbb{R}$, è detto *tempo del primo passaggio per* a. □

Teorema 9.4.13. *Sia $B = (B_t)_{t\in[0,\infty[}$ il moto Browniano. La v.a. τ_a ammette d.p. data da*

$$f_{\tau_a}(x) = \frac{|a|}{\sqrt{2\pi x^3}}\, e^{-\frac{a^2}{2x}}\, \mathbf{1}_{]0,\infty[}(x) .$$

Inoltre, si ha $\mathrm{E}[\tau_a] = \infty$.

Dimostrazione. Si veda Klebaner (2005, p.75). □

Dal precedente Teorema è immediato verificare che il tempo del primo passaggio del moto Browniano possiede in effetti la legge di Lévy introdotta nella sua versione ridotta nell'Esempio 4.1.3. Il prossimo Teorema contiene un risultato classico, ovvero la legge dell'arcoseno per il moto Browniano che fornisce la distribuzione dell'ultimo tempo di ritorno all'origine.

Teorema 9.4.14. (Legge dell'arcoseno per il moto Browniano) *Sia $B = (B_t)_{t\in[0,1]}$ il moto Browniano su $[0,1]$ e si consideri la v.a. $\tau = \sup\{t \in [0,1] : B_t = 0\}$. La v.a. τ ammette d.p. data da*

$$f_\tau(x) = \frac{1}{\pi\sqrt{x(1-x)}}\, \mathbf{1}_{[0,1]}(x) .$$

ovvero τ è una v.a. con legge Beta di tipo $\mathcal{BE}(0,1,\frac{1}{2},\frac{1}{2})$.

Dimostrazione. Si veda Klebaner (2005, p.75). □

Il prossimo Teorema fornisce un risultato che permette di ottenere notevoli semplificazioni nel calcolo delle leggi connesse al massimo del moto Browniano e al tempo di primo passaggio.

Teorema 9.4.15. (Principio di riflessione) *Sia $B = (B_t)_{t\in[0,1]}$ il moto Browniano su $[0,1]$ e si consideri la v.a. $B_t^* = \sup_{0\leq s\leq t} B_s$. Fra le v.a. B_t^*, τ_a e B_t sussiste la seguente relazione*

$$P(B_t^* \geq a) = P(\tau_a \leq t) = 2P(B_t \geq a)\,,$$

dove $a > 0$.

Dimostrazione. Si veda Baldi (2017, p.70). □

9.5. Riferimenti bibliografici

I testi di Bass (2011), Bhattacharya e Waymire (2016), Koralov e Sinai (2007), Knill (2009) e Mishura e Shevchenko (2017) sono espressamente focalizzati sulla teoria dei processi stocastici. Ampie parti dei testi avanzati di Ash e Doléans-Dade (2000), Billingsley (1995), Çinlar (2010), Dudley (2004), Kallenberg (2002) e Stroock (2013) sono dedicati ai processi stocastici. Per un approccio introduttivo a questo argomento si possono consultare i testi di Brzeźniak e Zastawniak (2002), Durrett (2012), Resnick (2005) e Tijms (2003). I testi di Capasso e Bakstein (2015) e Revuz e Yor (2005) considerano in modo esaustivo i processi stocastici a tempo continuo. I processi a tempo discreto, con particolare enfasi sulle martingale, sono considerati in Brzeźniak e Zastawniak (2002), Pascucci (2011) e Williams (1992). In particolar modo le applicazioni delle martingale al gioco d'azzardo sono analizzate in Ethier (2010). Per una introduzione alle passeggiate aleatorie si può consultare il testo di Klafter e Sokolov (2011), mentre è fondamentale il capitolo dedicato a questo argomento in Feller (1968). Il moto Browniano è considerato in dettaglio nei testi di Mörters e Peres (2010) e Schilling e Partzsch (2012).

9.6. Esercizi svolti

• **Nota.** Al fine di evitare notazioni ridondanti e ripetizioni, nei seguenti esercizi si assume l'esistenza di uno spazio probabilizzato $(\Omega, \mathcal{F}, P)$ e di una filtrazione $(\mathcal{F}_t)_{t\in\mathbb{T}}$, a meno che non venga specificato diversamente.

Sezione 9.2

• **Esercizio 9.2.1.** Si consideri una successione di v.a. $(X_n)_{n\geq 1}$ a componenti indipendenti tale che $\mathrm{E}[X_n] = 0$ e $\mathrm{E}[X_n^2] = c^2 < \infty$ per ogni n. Inoltre, si definisca il p.a. a tempo discreto $M = (M_n)_{n\geq 0}$ tale che

$$M_n = S_n^2 - nc^2\,,$$

dove $S_n = \sum_{k=1}^n X_k$, mentre $\{M_0 = 0\}$ $q.c.$ e $(\mathcal{F}_n)_{n\geq 0}$ è la filtrazione naturale, ovvero $\mathcal{F}_n = \sigma(\{X_1, \ldots, X_n\})$. Si verifichi che M è una martingala.

Soluzione. Tenendo presente le assunzioni fatte sul p.a. M, per ogni n si ha

$$\mathrm{E}[|M_n|] \leq \mathrm{E}[S_n^2] + nc^2 = 2nc^2 < \infty\,.$$

Inoltre, dal momento che $\{\mathrm{E}[M_n \mid \mathcal{F}_n] = M_n\}$ $q.c.$ e che la σ-algebra generata da X_n è indipendente dalla σ-algebra $\mathcal{F}_{n-1}$, si ha $q.c.$

$$\begin{aligned}\mathrm{E}[M_n \mid \mathcal{F}_{n-1}] &= \mathrm{E}[S_n^2 - nc^2 \mid \mathcal{F}_{n-1}] = \mathrm{E}[M_{n-1} + X_n^2 - c^2 + 2X_n S_{n-1} \mid \mathcal{F}_{n-1}] \\ &= \mathrm{E}[M_{n-1} \mid \mathcal{F}_{n-1}] + \mathrm{E}[X_n^2 - c^2] + 2S_{n-1}\mathrm{E}[X_n] = M_{n-1} \, .\end{aligned}$$

Inoltre, per la Proprietà della torre (Proposizione 5.2.6), si ha $q.c.$

$$\mathrm{E}[\mathrm{E}[M_n \mid \mathcal{F}_{n-1}] \mid \mathcal{F}_{n-2}] = \mathrm{E}[M_n \mid \mathcal{F}_{n-2}]$$

e quindi dalle precedenti relazioni sussiste $q.c.$

$$\mathrm{E}[M_n \mid \mathcal{F}_{n-2}] = \mathrm{E}[M_{n-1} \mid \mathcal{F}_{n-2}] = M_{n-2} \, .$$

Procedendo in modo iterativo, si ha infine $\{\mathrm{E}[M_n \mid \mathcal{F}_m] = M_m\}$ $q.c.$ per $m \in \mathbb{N}$ con $m < n$. Dunque, il p.a. M è una martingala. □

• **Esercizio 9.2.2.** Si consideri una successione di v.a. $(X_n)_{n\geq 1}$ a componenti indipendenti tale che $\mathrm{E}[X_n] = 1$ per ogni n e il p.a. $M = (M_n)_{n\geq 0}$ tale che

$$M_n = \prod_{k=1}^{n} X_k \, ,$$

dove $\{M_0 = 1\}$ $q.c.$ e $(\mathcal{F}_n)_{n\geq 0}$ è la filtrazione naturale. Si verifichi che M è una martingala.

Soluzione. Tenendo presente le assunzioni fatte sul p.a. M, per ogni n si ha $\mathrm{E}[|M_n|] < \infty$. Inoltre, si ha $q.c.$

$$\mathrm{E}[M_n \mid \mathcal{F}_{n-1}] = \mathrm{E}[X_n M_{n-1} \mid \mathcal{F}_{n-1}] = M_{n-1}\mathrm{E}[X_n] = M_{n-1} \, .$$

In modo simile all'Esercizio 9.2.1, si ha dunque $\{\mathrm{E}[M_n \mid \mathcal{F}_m] = M_m\}$ $q.c.$ per $m \in \mathbb{N}$ con $m < n$, ovvero il p.a. M è una martingala. □

• **Esercizio 9.2.3.** (Martingala di Wald) Si consideri una successione di v.a. $(X_n)_{n\geq 1}$ a componenti indipendenti e con la medesima legge, tale che $\psi(t) = \mathrm{E}[e^{tX_n}] < \infty$ per ogni $t \in \mathbb{R}$ e per ogni n. Inoltre, si definisca il p.a. a tempo discreto $M = (M_n)_{n\geq 0}$ tale che

$$M_n = \frac{e^{tS_n}}{\psi(t)^n} \, ,$$

dove $S_n = \sum_{k=1}^{n} X_k$, mentre $\{M_0 = 1\}$ $q.c.$ e $(\mathcal{F}_n)_{n\geq 0}$ è la filtrazione naturale. Si verifichi che M è una martingala.

Soluzione. Si osservi che risulta

$$M_n = \prod_{k=1}^{n} \frac{e^{tX_k}}{\psi(t)}$$

e $\mathrm{E}[\frac{e^{tX_k}}{\psi(t)}] = 1$. Dunque, sulla base dell'Esercizio 9.2.2, il p.a. M è una martingala. □

• **Esercizio 9.2.4.** Si consideri una successione di v.a. $(X_n)_{n\geq 1}$ con $\mathrm{E}[|X_n|] < \infty$ per ogni n. Inoltre, si definisca il p.a. a tempo discreto $M = (M_n)_{n\geq 0}$ tale che

$$M_n = \sum_{k=1}^{n} (X_k - \mathrm{E}[X_k \mid \mathcal{F}_{k-1}]) \, ,$$

dove $\{M_0 = 0\}$ $q.c.$ e $(\mathcal{F}_n)_{n\geq 0}$ è la filtrazione naturale. Si verifichi che M è una martingala.

Soluzione. Tenendo presente le assunzioni fatte sul p.a. M, per ogni n si ha

$$\mathrm{E}[|M_n|] \le \sum_{k=1}^{n} \mathrm{E}[|X_k| + |\mathrm{E}[X_k \mid \mathcal{F}_{k-1}]|] \le \sum_{k=1}^{n} \mathrm{E}[|X_k| + \mathrm{E}[|X_k| \mid \mathcal{F}_{k-1}]] = 2n\,\mathrm{E}[|X_1|] < \infty\,.$$

Inoltre, tenendo presente che $\mathrm{E}[X_k \mid \mathcal{F}_{n-1}] = X_k$ per $k \le n-1$ in quanto $\mathcal{F}_k \subset \mathcal{F}_{n-1}$, si ha $q.c.$

$$\begin{aligned}\mathrm{E}[M_n \mid \mathcal{F}_{n-1}] &= \mathrm{E}[M_{n-1} + X_n - \mathrm{E}[X_n \mid \mathcal{F}_{n-1}] \mid \mathcal{F}_{n-1}] \\ &= \mathrm{E}[M_{n-1} \mid \mathcal{F}_{n-1}] + \mathrm{E}[X_n - \mathrm{E}[X_n \mid \mathcal{F}_{n-1}] \mid \mathcal{F}_{n-1}] \\ &= M_{n-1} + \mathrm{E}[X_n \mid \mathcal{F}_{n-1}] - \mathrm{E}[X_n \mid \mathcal{F}_{n-1}] = M_{n-1}\,.\end{aligned}$$

In modo simile all'Esercizio 9.2.1, si ha dunque $\{\mathrm{E}[M_n \mid \mathcal{F}_m] = M_m\}$ $q.c.$ per $m \in \mathbb{N}$ con $m < n$, ovvero il p.a. M è una martingala. Quindi, sulla base di questo risultato, si può costruire una martingala partendo da una successione qualsiasi di v.a. con media finita. □

Sezione 9.3

• **Esercizio 9.3.1.** Data una traiettoria $(x_n)_{n\ge0}$ di una passeggiata aleatoria $X = (X_n)_{n\ge0}$, dove $x_n \in \{-n, 2-n, \ldots, n-2, n\}$ per $n \in \mathbb{N}$, si consideri la cardinalità dei percorsi $(x_m, \ldots, x_{m+n})$ della passeggiata dal passo m al passo $(m+n)$, dove $m \in \mathbb{N}$. Posto $x = x_{m+n} - x_m$, si verifichi che questa cardinalità è pari a

$$N_n(x) = \binom{n}{\frac{1}{2}(n-x)}\,.$$

Soluzione. Data una realizzazione $(y_n)_{n\ge1}$ della successione di v.a. $(Y_n)_{n\ge1}$, si ponga

$$r = \sum_{k=m}^{m+n} \mathbf{1}_{\{1\}}(y_k)$$

e

$$s = \sum_{k=m}^{m+n} \mathbf{1}_{\{-1\}}(y_k)\,.$$

Evidentemente, r e s rappresentano rispettivamente le cardinalità dei valori 1 e -1 nel vettore $(y_m, \ldots, y_{m+n})$. Dal momento che $n = r + s$ e $x = r - s$, allora si ha $r = \frac{1}{2}(n+x)$ e $s = \frac{1}{2}(n-x)$. Inoltre, vi sono

$$\binom{r+s}{s} = \binom{n}{\frac{1}{2}(n-x)}$$

possibili combinazioni dei valori 1 e -1 nel vettore $(y_m, \ldots, y_{m+n})$, da cui il risultato richiesto. Si osservi che la cardinalità dei percorsi $(x_m, \ldots, x_{m+n})$ dal passo m al passo $(m+n)$ coincide con la cardinalità dei percorsi $(x_0, \ldots, x_n)$ dal passo 0 al passo n. □

• **Esercizio 9.3.2.** (Principio di riflessione) Assumendo le notazioni dell'Esercizio 9.3.1 e posto $x_m, x_{m+n} > 0$, si verifichi che la cardinalità dei percorsi $(x_m, \ldots, x_{m+n})$ tali che $x_k = 0$ per qualche $k = m+1, \ldots, m+n-1$ è pari alla cardinalità dei percorsi del tipo $(z_m, \ldots, z_{m+n})$ con $z_m = -x_m$ e $z_{m+n} = x_{m+n}$.

Soluzione. Considerato un dato percorso $(x_m, \ldots, x_{m+n})$, sia k il minimo indice per cui la passeggiata ritorna all'origine, ovvero $k = \min\{j : x_j = 0, j = m, \ldots, m+n\}$. Dunque, a questo percorso può essere associato il percorso $(-x_m, \ldots, -x_{k-1}, x_k, \ldots, x_{m+n})$. Evidentemente, quest'ultimo percorso costituisce la "riflessione" del percorso originale per valori negativi della

passeggiata. Sulla base di questa osservazione, esiste dunque una corrispondenza biunivoca fra l'insieme dei percorsi $(x_m, \ldots, x_{m+n})$ e l'insieme dei percorsi $(z_m, \ldots, z_{m+n})$, che hanno perciò la stessa cardinalità. Tenendo presente l'Esercizio 9.3.1, questa cardinalità è data da $N_n(x)$ dove in questo caso $x = x_m + x_{m+n}$. □

• **Esercizio 9.3.3.** (Problema del ballottaggio) Considerata un'elezione con due candidati che ricevono rispettivamente a e b voti con $a > b$, si determini la probabilità p che il candidato con più voti risulti strettamente in vantaggio durante lo spoglio.

Soluzione. Si assuma che $n = a + b$ e $x = a - b$. Tenendo presente le notazioni dell'Esercizio 9.3.1, il problema può essere assimilato ad un percorso $(x_0, \ldots, x_n)$ di una passeggiata aleatoria, dove $x_n = x$. In questo caso, x_k rappresenta la differenza di voti fra i due candidati alla k-esima iterazione dello spoglio, dove $k = 0, 1, \ldots, n$. Si osservi che, se il candidato con più voti è sempre in vantaggio nello spoglio, allora deve risultare $x_1 = 1$. Sulla base degli Esercizi 9.3.1 e 9.3.2, la cardinalità dell'insieme dei percorsi $(x_1, \ldots, x_n)$ è data da $N_{n-1}(x-1)$, mentre la cardinalità di questi percorsi tali che $x_k = 0$ per qualche k è data da $N_{n-1}(x+1)$. Dal momento che la cardinalità dell'insieme dei percorsi $(x_0, \ldots, x_n)$ è data da $N_n(x)$, la probabilità richiesta risulta

$$p = \frac{N_{n-1}(x-1) - N_{n-1}(x+1)}{N_n(x)} = \frac{\binom{n-1}{\frac{1}{2}(n+x-2)} - \binom{n-1}{\frac{1}{2}(n+x)}}{\binom{n}{\frac{1}{2}(n+x)}} = \frac{x}{n} = \frac{a-b}{a+b} ,$$

un risultato intuibile, sebbene abbastanza complicato da verificare con metodi diversi da quello proposto. □

Sezione 9.4

• **Esercizio 9.4.1.** Dato il moto Browniano $B = (B_t)_{t \in [0,\infty[}$ e assumendo che $s \leq t$, si calcoli il valore atteso $\mathrm{E}[B_s^2 B_t^2]$.

Soluzione. Si osservi che

$$\mathrm{E}[B_s^2 B_t^2] = \mathrm{E}[B_s^2 (B_t - B_s + B_s)^2] = \mathrm{E}[B_s^2 (B_t - B_s)^2 + 2B_s^3 (B_t - B_s) + B_s^4] .$$

Inoltre, se la v.a. Z è distribuita con legge Normale $\mathcal{N}(0,1)$, si ha $B_s \stackrel{\mathcal{L}}{=} \sqrt{s} Z$ e quindi

$$\mathrm{E}[B_s^4] = s^2 \, \mathrm{E}[Z^4] = 3s^2 .$$

Dunque, per le proprietà del moto Browniano, dalle precedenti espressioni risulta

$$\mathrm{E}[B_s^2 B_t^2] = \mathrm{E}[B_s^2] \mathrm{E}[(B_t - B_s)^2] + 2 \, \mathrm{E}[B_s^3] \mathrm{E}[B_t - B_s] + \mathrm{E}[B_s^4] = st + 2s^2 ,$$

dal momento che $\mathrm{E}[B_s^2] = s$, $\mathrm{E}[(B_t - B_s)^2] = t - s$ e $\mathrm{E}[B_t - B_s] = 0$. □

• **Esercizio 9.4.2.** Dato il moto Browniano $B = (B_t)_{t \in [0,\infty[}$ e assumendo che $s \leq t$, si verifichi che il v.v.a. $(B_1, B_2)^{\mathrm{T}}$ ammette d.p.c. data da

$$f_{(B_1,B_2)}(x_1, x_2) = \frac{1}{\sqrt{2\pi s}} \, e^{-\frac{1}{2s} x_1^2} \, \frac{1}{\sqrt{2\pi (t-s)}} \, e^{-\frac{1}{2(t-s)} (x_2 - x_1)^2} .$$

Soluzione. Tenendo presente che la matrice di varianza-covarianza del v.v.a. $(B_1, B_2)^{\mathrm{T}}$ è data da

$$\Sigma = \begin{pmatrix} s & s \\ s & t \end{pmatrix} ,$$

si ha

$$\Sigma^{-1} = \frac{1}{s(t-s)} \begin{pmatrix} t & -s \\ -s & s \end{pmatrix} .$$

Dunque, posto $x = (x_1, x_2)^{\mathrm{T}}$ si ha

$$x^{\mathrm{T}} \Sigma^{-1} x = \frac{1}{s(t-s)} \left(t x_1^2 - 2 s x_1 x_2 + s x_2^2 \right) = \frac{1}{s(t-s)} \left((t-s) x_1^2 + s (x_2 - x_1)^2 \right) .$$

Il v.v.a. $(B_1, B_2)^{\mathrm{T}}$ è distribuito con legge Normale bivariata $\mathcal{N}_2(0, \Sigma)$ e quindi si ottiene infine

$$f_{(B_1,B_2)}(x_1, x_2) = \det(2\pi\Sigma)^{-\frac{1}{2}} \, e^{-\frac{1}{2} x^{\mathrm{T}} \Sigma^{-1} x} = \frac{1}{\sqrt{2\pi s}} \, e^{-\frac{1}{2s} x_1^2} \, \frac{1}{\sqrt{2\pi(t-s)}} \, e^{-\frac{1}{2(t-s)} (x_2 - x_1)^2} .$$

Si osservi che considerando la trasformata $(Y_1, Y_2)^{\mathrm{T}} = g(B_1, B_2)$ dove $g(x_1, x_2) = (x_1, x_2 - x_1)^{\mathrm{T}}$ e tale che $|J(g^{-1}(y))| = 1$, si ha

$$f_{(Y_1,Y_2)}(y_1, y_2) = \frac{1}{\sqrt{2\pi s}} \, e^{-\frac{1}{2s} y_1^2} \, \frac{1}{\sqrt{2\pi(t-s)}} \, e^{-\frac{1}{2(t-s)} y_2^2} ,$$

ovvero B_1 e $(B_2 - B_1)$ sono v.a. indipendenti con rispettive leggi $\mathcal{N}(0, s)$ e $\mathcal{N}(0, t-s)$, come evidentemente deve risultare per le proprietà del moto Browniano. □

• **Esercizio 9.4.3.** Dato il moto Browniano $B = (B_t)_{t \in [0,\infty[}$, si calcoli i valori attesi $\mathrm{E}[\mathbf{1}_{]-\infty,a]}(B_t)]$ e $\mathrm{E}[B_t \mathbf{1}_{]-\infty,a]}(B_t)]$, dove $a \in \mathbb{R}$, e se ne determini il limite per $t \to \infty$.

Soluzione. Tenendo presente che $B_t \stackrel{\mathcal{L}}{=} \sqrt{t} Z$ dove Z è una v.a. con legge Normale $\mathcal{N}(0, 1)$, si ha

$$\mathrm{E}[\mathbf{1}_{]-\infty,a]}(B_t)] = P(B_t \le a) = P\left(Z \le \frac{a}{\sqrt{t}} \right) = \Phi\left(\frac{a}{\sqrt{t}} \right)$$

e quindi, in modo controintuitivo, per ogni a si ha

$$\lim_{t \to \infty} \mathrm{E}[\mathbf{1}_{]-\infty,a]}(B_t)] = \Phi(0) = \frac{1}{2} .$$

Inoltre, risulta

$$\mathrm{E}[B_t \, \mathbf{1}_{]-\infty,a]}(B_t)] = \mathrm{E}[\sqrt{t} Z \, \mathbf{1}_{]-\infty,a]}(\sqrt{t} Z)] = \sqrt{t} \int_{-\infty}^{\frac{a}{\sqrt{t}}} x \, \frac{1}{\sqrt{2\pi}} \, e^{-\frac{1}{2} x^2} \, dx = \, - \sqrt{\frac{t}{2\pi}} \, e^{-\frac{1}{2t} a^2}$$

e quindi, di nuovo in modo controintuitivo, per ogni a si ha

$$\lim_{t \to \infty} \mathrm{E}[B_t \mathbf{1}_{]-\infty,a]}(B_t)] = \, - \infty . \qquad \square$$

• **Esercizio 9.4.4.** Dato il moto Browniano $B = (B_t)_{t \in [0,\infty[}$, si consideri la v.a. $B_t^* = \sup_{0 \le s \le t} B_s$ e si determini la d.p. $f_{B_t^*}$. Si ottenga inoltre il valore atteso $\mathrm{E}[B_t^*]$.

Soluzione. Dal momento che per le proprietà del moto Browniano si ha $\{B_0 = 0\}$ $q.c.$, allora risulta $\{B_t^* \ge 0\}$ $q.c.$ Inoltre, assumendo che $a > 0$, per il Principio di riflessione si ha

$$P(B_t^* \ge a) = 2 P(B_t \ge a) = P(|B_t| \ge a) .$$

Quindi, risulta $B_t^* \stackrel{\mathcal{L}}{=} |B_t|$, da cui

$$f_{B_t^*}(x) = 2 f_{B_t}(x)\, \mathbf{1}_{]0,\infty[}(x) = \sqrt{\frac{2}{\pi t}}\, e^{-\frac{1}{2t}x^2}\, \mathbf{1}_{]0,\infty[}(x)\,.$$

Infine, dal precedente risultato si ottiene

$$\mathrm{E}[B_t^*] = \int_0^\infty x \sqrt{\frac{2}{\pi t}}\, e^{-\frac{1}{2t}x^2}\, dx = \sqrt{\frac{2t}{\pi}}\,. \qquad \square$$

• **Esercizio 9.4.5.** (Tempo di occupazione) Dato il moto Browniano $B = (B_t)_{t\in[0,\infty[}$, si consideri l'insieme misurabile $A \subset \mathbb{R}$ con misura di Lebesgue $\lambda(A) > 0$ e l'insieme aleatorio

$$S_A = \{\omega \in \Omega : B_t(\omega) \in A,\, t \in [0,\infty[\}\,.$$

Si determini il valore atteso $\mathrm{E}[\lambda(S_A)]$.

Soluzione. Si osservi che S_A è in effetti l'insieme dei valori di t per cui il moto Browniano visita l'insieme A e per questo motivo S_A viene anche definito come tempo di occupazione dell'insieme A. Ovviamente, per ogni traiettoria $B_t(\omega)$ del moto Browniano, S_A si realizza in un insieme dato di valori. Dunque, si ha che

$$\lambda(S_A) = \int_0^\infty \mathbf{1}_A(B_t)\, dt$$

è una v.a. che rappresenta la misura di Lebesgue dell'insieme aleatorio S_A. Sulla base di questa espressione e del Teorema di Fubini, risulta

$$\mathrm{E}[\lambda(S_A)] = \mathrm{E}[\textstyle\int_0^\infty \mathbf{1}_A(B_t)\, dt] = \displaystyle\int_0^\infty \mathrm{E}[\mathbf{1}_A(B_t)]\, dt = \int_0^\infty P(B_t \in A)\, dt\,.$$

Tenendo presente che la v.a. B_t si distribuisce con legge Normale $\mathcal{N}(0,t)$ e applicando di nuovo il Teorema di Fubini, si ha

$$\mathrm{E}[\lambda(S_A)] = \int_0^\infty \left(\int_A \frac{1}{\sqrt{2\pi t}}\, e^{-\frac{1}{2t}x^2}\, dx \right) dt = \int_A \int_0^\infty \frac{1}{\sqrt{2\pi t}}\, e^{-\frac{1}{2t}x^2}\, dt\, dx\,.$$

Tuttavia, risulta

$$\int_0^\infty \frac{1}{\sqrt{2\pi t}}\, e^{-\frac{1}{2t}x^2}\, dt = \frac{|x|}{2\sqrt{\pi}} \int_0^\infty \frac{e^{-y}}{\sqrt{y^3}}\, dy = \infty$$

e dunque $\mathrm{E}[\lambda(S_A)] = \infty$, ovvero il moto Browniano visita un qualsiasi insieme misurabile A infinitamente spesso. Risulta notevole osservare che, con una dimostrazione simile alla precedente, per il moto Browniano multivariato in $\mathbb{R}^d$, dove $d > 2$, e per un insieme misurabile $A \subset \mathbb{R}^d$, si ha

$$\mathrm{E}[\lambda^d(S_A)] = \frac{\Gamma(\frac{1}{2}d - 1)}{2\pi^{\frac{1}{2}d}} \int_A \|x\|^{2-d}\, dx\,.$$

In questo caso, l'insieme aleatorio S_A viene definito opportunamente per il moto Browniano multivariato. Dunque, per $d > 2$ si può avere $\mathrm{E}[\lambda^d(S_A)] < \infty$ per un insieme A, ovvero il moto Browniano può non visitare infinitamente spesso l'insieme A. $\square$

Pagina intenzionalmente vuota

Capitolo 10

Calcolo stocastico

10.1. Introduzione all'integrazione stocastica

Il problema dell'integrazione stocastica può essere fatto risalire a Louis Bachelier (1870-1946). Se X_t rappresenta il prezzo di un bene al variare del tempo, nella sua tesi di laurea *Théorie de la spéculation* (1900), Louis Bachelier propose di modellare gli incrementi infinitesimi del prezzo in modo proporzionale agli incrementi infinitesimi del moto Browniano, ovvero (con una simbologia moderna) introdusse la relazione euristica $dX_t = \sigma dB_t$ dove $\sigma \in]0, \infty[$. In questo caso, la precedente relazione porta intuitivamente a concludere che $X_t = X_0 + \sigma B_t$. Tuttavia, anche se questo modello può essere discretamente valido per descrivere la dinamica del prezzo nel breve periodo, in generale non è realistico. In effetti, tenendo presenti le proprietà del moto Browniano, il prezzo X_t può assumere valori negativi con probabilità crescente all'aumentare del tempo.

Figura 9.2.1. Kiyosi Itô (1915-2008).

L'inconveniente del modello potrebbe essere aggirato supponendo che l'incremento infinitesimo del prezzo, rapportato al prezzo stesso, sia proporzionale agli incrementi infinitesimi del moto Browniano considerando la relazione euristica $dX_t/X_t = \sigma dB_t$, ovvero $dX_t = \sigma X_t dB_t$. Anche se questa ultima espressione potrebbe ricordare una equazione differenziale ordinaria, evidentemente presenta una difficoltà teorica dal momento che il moto Browniano non è differenziabile *q.o.* e quindi non ha significato matematico. Il problema è stato risolto da Kiyosi Itô (1915-2008), che ha dato un'interpretazione rigorosa alla precedente relazione con una formulazione in termini di integrale del tipo $X_t = X_0 + \sigma \int_0^t X_s dB_s$. Nel secondo termine di quest'espressione compare appunto il cosiddetto *integrale stocastico di Itô*.

Di seguito vengono introdotte le basi per la costruzione dell'integrale stocastico di Itô partendo da una classe elementare di processi, al fine di ottenere la definizione generale nella prossima sezione.

Definizione 10.1.1. Si consideri uno spazio di probabilità $(\Omega, \mathcal{F}, P)$ con filtrazione $(\mathcal{F}_t)_{t\in[0,T]}$ dove $T > 0$. Sia inoltre $0 = t_0 < t_1 < \ldots < t_n = T$ una partizione di $[0, T]$ e sia $Z = (Z_0, \ldots, Z_{n-1})^{\mathrm{T}}$ un v.v.a. tale che ogni v.a. Z_k è misurabile rispetto a $\mathcal{F}_{t_k}$ con $\mathrm{E}[Z_k^2] < \infty$ per $k = 0, \ldots, n-1$. Il p.a. $X = (X_t)_{t\in[0,T]}$ è detto *semplice* se

$$X_t = \sum_{k=0}^{n-1} Z_k \, \mathbf{1}_{[t_k, t_{k+1}[}(t) \, .$$

La classe dei p.a. semplici è denotata con $\mathcal{S}_T^2$. □

Si osservi che la misurabilità delle v.a. Z_k rispetto alle σ-algebre $\mathcal{F}_{t_k}$ implica che il p.a. semplice X sia adattato alla filtrazione $(\mathcal{F}_t)_{t\in[0,\infty[}$. Inoltre, l'assunzione $\mathrm{E}[Z_k^2] < \infty$ implica che $\mathrm{E}[X_t^2] < \infty$ per ogni $t \in [0, T]$. Infine, $\mathcal{S}_T^2$ è uno spazio vettoriale, dal momento che è immediato verificare che $aX + bY \in \mathcal{S}_T^2$ per ogni $X, Y \in \mathcal{S}_T^2$ e $a, b \in \mathbb{R}$.

Definizione 10.1.2. Si consideri uno spazio di probabilità $(\Omega, \mathcal{F}, P)$ con filtrazione $(\mathcal{F}_t)_{t\in[0,\infty[}$ e il moto Browniano $B = (B_t)_{t\in[0,\infty[}$ adattato alla filtrazione. Dato il p.a. $X \in \mathcal{S}_T^2$, l'*integrale stocastico di Itô* di X è definito come

$$I(X) = \sum_{k=0}^{n-1} Z_k (B_{t_{k+1}} - B_{t_k})$$

e si scrive

$$I(X) = \int_0^T X_t \, dB_t \, .$$ □

Dalla Definizione 10.1.2 è evidente che l'integrale di Itô di un p.a. semplice è una v.a. Inoltre, risulta opportuno sottolineare la differenza teorica fra l'integrale di Itô e l'integrale di Lebesgue di un p.a. semplice. Più esattamente, l'integrale del p.a. semplice X è definito come

$$\int_0^T X_t \, dt = \sum_{k=0}^{n-1} Z_k (t_{k+1} - t_k) \, .$$

Dunque, anche l'integrale del p.a. è ovviamente una v.a. Tuttavia, l'integrale del p.a. è basato su incrementi deterministici, al contrario dell'integrale di Itô, dove gli incrementi sono proporzionali agli incrementi di un moto Browniano. In generale, si noti che l'integrale di un p.a. (le cui traiettorie sono integrabili *q.c.*) è una v.a., dal momento che ad ogni realizzazione del processo viene associato un integrale di Lebesgue.

Si osservi inoltre che per semplicità di notazione si pone anche

$$\int_a^b X_t \, dB_t = \int_0^T \mathbf{1}_{[a,b[}(t) X_t \, dB_t \, ,$$

dove $a, b \in [0, T]$ e $a < b$.

Il seguente Teorema evidenzia che l'integrale stocastico di Itô è una v.a. con media nulla e secondo momento finito quando il p.a. è semplice.

Teorema 10.1.3. *Se $X \in \mathcal{S}_T^2$, l'integrale stocastico di Itô è una v.a. tale che* $\mathrm{E}[I(X)] = 0$ *e*

$$\mathrm{Var}[I(X)] = \mathrm{E}[I(X)^2] = \int_0^T \mathrm{E}[X_t^2]\, dt = \mathrm{E}[\textstyle\int_0^T X_t^2\, dt] < \infty\,.$$

Dimostrazione. Per semplicità di notazione si assuma che $\Delta_k = B_{t_{k+1}} - B_{t_k}$. Si noti che la v.a. Z_k è misurabile rispetto a $\mathcal{F}_{t_k}$ e che l'incremento Δ_k è indipendente dalla σ-algebra $\mathcal{F}_{t_k}$ per l'assunzione *iii*) della Definizione 9.4.1. Dunque, tenendo presente la Proposizione 5.2.1 e il Teorema 5.2.4 si ha

$$\mathrm{E}[Z_k\Delta_k] = \mathrm{E}[\mathrm{E}[Z_k\Delta_k \mid \mathcal{F}_{t_k}]] = \mathrm{E}[Z_k\mathrm{E}[\Delta_k \mid \mathcal{F}_{t_k}]] = \mathrm{E}[Z_k\mathrm{E}[\Delta_k]] = 0\,,$$

in quanto sussiste $\mathrm{E}[\Delta_k] = 0$ sulla base dell'assunzione *ii*) della Definizione 9.4.1. Quindi, risulta

$$\mathrm{E}[I(X)] = \sum_{k=0}^{n-1} \mathrm{E}[Z_k\Delta_k] = 0\,.$$

Inoltre, si ha

$$I(X)^2 = \sum_{k=0}^{n-1}\sum_{j=0}^{n-1} Z_k Z_j \Delta_k \Delta_j\,.$$

Tenendo presente i precedenti commenti, assumendo $k < j$, si ha

$$\mathrm{E}[Z_kZ_j\Delta_k\Delta_j] = \mathrm{E}[\mathrm{E}[Z_kZ_j\Delta_k\Delta_j \mid \mathcal{F}_{t_j}]] = \mathrm{E}[Z_kZ_j\Delta_k\mathrm{E}[\Delta_j \mid \mathcal{F}_{t_j}]] = \mathrm{E}[Z_kZ_j\Delta_k\mathrm{E}[\Delta_j]] = 0\,,$$

mentre per $k = j$, si ha

$$\mathrm{E}[Z_k^2\Delta_k^2] = \mathrm{E}[\mathrm{E}[Z_k^2\Delta_k^2 \mid \mathcal{F}_{t_k}]] = \mathrm{E}[Z_k^2\mathrm{E}[\Delta_k^2 \mid \mathcal{F}_{t_k}]] = \mathrm{E}[Z_k^2\mathrm{E}[\Delta_k^2]] = \mathrm{E}[Z_k^2](t_{k+1} - t_k)\,,$$

dal momento che $\mathrm{E}[\Delta_k^2] = t_{k+1} - t_k$ dall'assunzione *ii*) della Definizione 9.4.1. Dunque, risulta

$$\mathrm{E}[I(X)^2] = \sum_{k=0}^{n-1} \mathrm{E}[Z_k^2](t_{k+1} - t_k)\,.$$

Inoltre, tenendo presente che $\mathbf{1}_{[t_k,t_{k+1}[}(t)\mathbf{1}_{[t_j,t_{j+1}[}(t) = 0$ per $k \neq j$, si ha

$$\begin{aligned}\int_0^T X_t^2\, dt &= \int_0^T \sum_{k=0}^{n-1}\sum_{j=0}^{n-1} Z_kZ_j\mathbf{1}_{[t_k,t_{k+1}[}(t)\mathbf{1}_{[t_j,t_{j+1}[}(t)\, dt \\ &= \int_0^T \sum_{k=0}^{n-1} Z_k^2\mathbf{1}_{[t_k,t_{k+1}[}(t)\, dt = \sum_{k=0}^{n-1} Z_k^2(t_{k+1} - t_k)\,,\end{aligned}$$

da cui risulta

$$\mathrm{E}[\textstyle\int_0^T X_t^2\, dt] = \sum_{k=0}^{n-1} \mathrm{E}[Z_k^2](t_{k+1} - t_k)\,.$$

Inoltre, per il Teorema di Fubini (Teorema A.11), si ha

$$\mathrm{E}[\textstyle\int_0^T X_t^2\, dt] = \int_0^T \mathrm{E}[X_t^2]\, dt\,.$$

La tesi segue dalle precedenti relazioni. □

• **Esempio 10.1.1.** Si assuma il p.a. semplice $X = (X_t)_{t\in[0,5]}$, dove

$$X_t = Z_0\,\mathbf{1}_{[0,1[}(t) + Z_1\,\mathbf{1}_{[1,2[}(t) + Z_2\,\mathbf{1}_{[2,5]}(t)\,,$$

e $Z = (Z_0, Z_1, Z_2)^{\mathrm{T}}$ è un v.v.a. degenere con $P(Z_0 = 1, Z_1 = 2, Z_2 = 1) = 1$. In questo caso, l'integrale stocastico di Itô è dato da

$$I(X) = Z_0(B_1 - B_0) + Z_1(B_2 - B_1) + Z_2(B_5 - B_2) \stackrel{\mathcal{L}}{=} B_1 + 2(B_2 - B_1) + (B_5 - B_2)\,.$$

Dunque, dall'assunzione *iii*) della Definizione 9.4.1, si ha che $I(X)$ è una v.a. con legge Normale $\mathcal{N}(0, 8)$. In effetti, dal Teorema 10.1.3 risulta $\mathrm{E}[I(X)] = 0$ e

$$\mathrm{Var}[I(X)] = \mathrm{E}[Z_0^2] + \mathrm{E}[Z_1^2] + 3\,\mathrm{E}[Z_2^2] = 8\,. \qquad \square$$

• **Esempio 10.1.2.** Si assuma il p.a. semplice $X = (X_t)_{t\in[0,1]}$, dove

$$X_t = Z_0\,\mathbf{1}_{[0,1]}(t)\,,$$

mentre Z_0 è una v.a. con legge di Bernoulli di parametro $\frac{1}{2}$. In questo caso, l'integrale stocastico di Itô è dato da

$$I(X) = Z_0(B_1 - B_0) \stackrel{\mathcal{L}}{=} Z_0 B_1\,.$$

Dunque, la f.r. della v.a. $I(X)$ è data da

$$\begin{aligned} F_{I(X)}(x) &= P(I(X) \le x) = P(\{I(X) \le x\} \cap \{Z_0 = 0\}) + P(\{I(X) \le x\} \cap \{Z_0 = 1\}) \\ &= P(I(X) \le x \mid Z_0 = 0)P(Z_0 = 0) + P(I(X) \le x \mid Z_0 = 1)P(Z_0 = 1) \\ &= \frac{1}{2}\,\mathbf{1}_{[0,\infty[}(x) + \frac{1}{2}\,\Phi(x)\,. \end{aligned}$$

Dunque, $I(X)$ è una v.a. mista, dal momento che la corrispondente f.r. è una combinazione convessa di una f.r. di una v.a. degenere nello zero e una v.a. con legge Normale ridotta. Inoltre, risulta $\mathrm{E}[I(X)] = 0$ e

$$\mathrm{Var}[I(X)] = \mathrm{E}[Z_0^2] = \frac{1}{2}\,. \qquad \square$$

I seguenti risultati forniscono rispettivamente la proprietà di linearità l'integrale stocastico di Itô e la covarianza fra due integrali stocastici di Itô nel caso di p.a. semplici.

Teorema 10.1.4. *Dati i p.a. $X, Y \in \mathcal{S}_T^2$, per ogni $a, b \in \mathbb{R}$ si ha*

$$I(aX + bY) = aI(X) + bI(Y)\,.$$

Dimostrazione. Si consideri la partizione $0 = t_0 < t_1 < \ldots < t_n = T$ di $[0, T]$, in modo tale che

$$X_t = \sum_{k=0}^{n-1} Z_k\,\mathbf{1}_{[t_k,t_{k+1}[}(t)$$

e

$$Y_t = \sum_{k=0}^{n-1} U_k\,\mathbf{1}_{[t_k,t_{k+1}[}(t)\,.$$

dove le v.a. Z_k e U_k sono misurabili rispetto a $\mathcal{F}_{t_k}$ con $\mathrm{E}[Z_k^2] < \infty$ e $\mathrm{E}[U_k^2] < \infty$ per $k = 0, \ldots, n-1$. Se le due partizioni per i p.a. X e Y sono differenti, è sempre possibile considerare una partizione comune considerando un raffinamento delle due partizioni originali. Dunque, si ha

$$aX_t + bY_t = \sum_{k=0}^{n-1} (aZ_k + bU_k)\, \mathbf{1}_{[t_k, t_{k+1}[}(t)\,.$$

Assumendo le notazioni della dimostrazione del Teorema 10.1.3, risulta

$$I(aX + bY) = \sum_{k=0}^{n-1} (aZ_k + bU_k)\Delta_k = a\sum_{k=0}^{n-1} Z_k\Delta_k + b\sum_{k=0}^{n-1} U_k\Delta_k = aI(X) + bI(Y)\,,$$

da cui segue la tesi. □

Proposizione 10.1.5. *Dati i p.a. $X, Y \in \mathcal{S}_T^2$, si ha*

$$\mathrm{Cov}[I(X), I(Y)] = \mathrm{E}[I(X)I(Y)] = \int_0^T \mathrm{E}[X_tY_t]\, dt = \mathrm{E}[\textstyle\int_0^T X_tY_t\, dt]\,.$$

Dimostrazione. Si tenga presente l'identità

$$ab = \frac{1}{2}\,(a+b)^2 - \frac{1}{2}\,a^2 - \frac{1}{2}\,b^2\,,$$

dove $a, b \in \mathbb{R}$. Dai Teoremi 10.1.3 e 10.1.4, si ha

$$\begin{aligned}\mathrm{E}[I(X)I(Y)] &= \frac{1}{2}\,\mathrm{E}[(I(X)+I(Y))^2] - \frac{1}{2}\,\mathrm{E}[I(X)^2] - \frac{1}{2}\,\mathrm{E}[I(Y)^2] \\ &= \frac{1}{2}\,\mathrm{E}[\textstyle\int_0^T (X_t+Y_t)^2\, dt] - \frac{1}{2}\,\mathrm{E}[\int_0^T X_t^2\, dt] - \frac{1}{2}\,\mathrm{E}[\int_0^T Y_t^2\, dt] = \mathrm{E}[\int_0^T X_tY_t\, dt]\,.\end{aligned}$$

L'identità

$$\mathrm{E}[\textstyle\int_0^T X_tY_t\, dt] = \displaystyle\int_0^T \mathrm{E}[X_tY_t]\, dt$$

segue dal Teorema di Fubini (Teorema A.11). □

10.2. Integrale di Itô

La seguente classe di p.a. è centrale nella definizione generale dell'integrale stocastico di Itô. In effetti, gli elementi di questa classe saranno le "funzioni integrande" opportune per cui l'integrale stocastico è definito.

Definizione 10.2.1. Si consideri uno spazio di probabilità $(\Omega, \mathcal{F}, P)$ con filtrazione $(\mathcal{F}_t)_{t\in[0,\infty[}$. La classe $\mathcal{M}_T^2$ è costituita dai p.a. $X = (X_t)_{t\in[0,T]}$ adattati alla filtrazione e tali che $\mathrm{E}[\int_0^T X_t^2 dt] < \infty$. □

• **Esempio 10.2.1.** Si consideri il p.a. $X = (B_t)_{t\in[0,T]}$. Tenendo presente la Definizione 9.4.1 si ha

$$\mathrm{E}[\textstyle\int_0^T B_t^2\, dt] = \displaystyle\int_0^T \mathrm{E}[B_t^2]\, dt = \int_0^T t\, dt = \frac{T^2}{2} < \infty\,,$$

ovvero $X \in \mathcal{M}_T^2$. Si consideri inoltre il p.a. $X = (B_t^2)_{t\in[0,T]}$. Tenendo presente che in base alla Proposizione 9.4.3 si ha $\mathrm{E}[B_t^4] = 3t^2$ (si veda anche l'Esempio 7.1.8), allora risulta

$$\mathrm{E}[\textstyle\int_0^T B_t^4\, dt] = \displaystyle\int_0^T \mathrm{E}[B_t^4]\, dt = \int_0^T 3t^2\, dt = T^3 < \infty\,,$$

ovvero si ha anche $X \in \mathcal{M}_T^2$. □

È immediato verificare che $\mathcal{S}_T^2 \subset \mathcal{M}_T^2$. Inoltre, la seguente Proposizione fornisce una semplice condizione per verificare se un p.a. appartiene alla classe $\mathcal{M}_T^2$, mentre il Teorema successivo evidenzia che la classe $\mathcal{S}_T^2$ è densa nella classe $\mathcal{M}_T^2$.

Proposizione 10.2.2. *Se il p.a. $X = (X_t)_{t\in[0,T]}$ è adattato alla filtrazione e limitato, ovvero se esiste $K > 0$ tale che $|X_t(\omega)| < K$ per ogni $(t,\omega) \in [0,T]\times\Omega$, allora $X \in \mathcal{M}_T^2$.*

Dimostrazione. Dalle assunzioni si ha

$$\mathrm{E}[\textstyle\int_0^T X_t^2\, dt] \le \mathrm{E}[\int_0^T K^2\, dt] = TK^2\,,$$

e sulla base della Definizione 10.2.1 segue la tesi. □

Teorema 10.2.3. *Per ogni p.a. $X = (X_t)_{t\in[0,T]}$ tale che $X \in \mathcal{M}_T^2$ esiste una successione di p.a. $(X_n)_{n\ge1}$, dove $X_n = (X_{t,n})_{t\in[0,T]}$ con $X_n \in \mathcal{S}_T^2$, che approssima il p.a. X, ovvero*

$$\lim_n \mathrm{E}[\textstyle\int_0^T (X_t - X_{t,n})^2\, dt] = 0\,.$$

Dimostrazione. Si veda Da Prato (2014). □

Sulla base dei precedenti risultati, si può dunque introdurre la versione generale dell'integrale stocastico di Itô di un p.a. definito su $\mathcal{M}_T^2$ come limite in media quadratica di una successione di integrali stocastici relativi ad una successione di p.a. semplici.

Definizione 10.2.4. Si consideri uno spazio di probabilità $(\Omega, \mathcal{F}, P)$ con filtrazione $(\mathcal{F}_t)_{t\in[0,\infty[}$ e un moto Browniano $B = (B_t)_{t\in[0,\infty[}$ adattato alla filtrazione. Dato il p.a. $X \in \mathcal{M}_T^2$, si dice che $I(X)$ è l'*integrale stocastico di Itô* del p.a. X se

$$\lim_n \mathrm{E}[(I(X) - I(X_n))^2] = 0$$

per ogni successione $(X_n)_{n\ge1}$ di p.a. tali che $X_n \in \mathcal{S}_T^2$ e si scrive

$$I(X) = \int_0^T X_t\, dB_t\,.$$ □

Evidentemente, la precedente Definizione implica che l'integrale stocastico di Itô può essere definito in termini di convergenza in media quadratica, ovvero risulta

$$I(X_n) \xrightarrow{L^2} I(X)$$

per $n \to \infty$. Inoltre, in modo equivalente a quanto fatto per i p.a. semplici, per comodità di notazione si pone anche

$$\int_a^b X_t\, dB_t = \int_0^T \mathbf{1}_{[a,b[}(t) X_t\, dB_t\,.$$

Nel caso particolare in cui $X = (X_t)_{t\in[0,T]}$ è un processo deterministico, ovvero se $X_t = g(t)$ *q.c.* con $g : \mathbb{R} \to \mathbb{R}$, l'integrale di Itô si riduce a

$$I(X) = \int_0^T g(t)\, dB_t \,,$$

che è anche detto *integrale di Wiener*.

Dal seguente Teorema si evince che l'integrale stocastico è ben definito, nel senso che non dipende dalla scelta della successione di p.a. semplici.

Teorema 10.2.5. *Per ogni p.a. $X \in \mathcal{M}_T^2$, l'integrale stocastico di Itô $I(X)$ non dipende dalla successione di p.a. semplici $(X_n)_{n\geq 1}$ che approssimano il p.a. X.*

Dimostrazione. Si consideri due successioni di p.a. semplici $(X_n)_{n\geq 1}$ e $(Y_n)_{n\geq 1}$ che approssimano X in $\mathcal{M}_T^2$, ovvero tali che $\lim_n \mathrm{E}[\int_0^T (X_t - X_{t,n})^2 dt] = 0$ e $\lim_n \mathrm{E}[\int_0^T (X_t - Y_{t,n})^2 dt] = 0$. Tenendo presente la disuguaglianza

$$(a+b)^2 \leq 2a^2 + 2b^2 \,,$$

dove $a, b \in \mathbb{R}$, si ha

$$\begin{aligned}\mathrm{E}[\textstyle\int_0^T (X_{t,n} - Y_{t,n})^2\, dt] &= \mathrm{E}[\textstyle\int_0^T ((X_{t,n} - X_t) + (X_t - Y_{t,n}))^2\, dt] \\ &\leq 2\,\mathrm{E}[\textstyle\int_0^T (X_t - X_{t,n})^2\, dt] + 2\,\mathrm{E}[\textstyle\int_0^T (X_t - Y_{t,n})^2\, dt]\end{aligned}$$

e dunque

$$\lim_n \mathrm{E}[\textstyle\int_0^T (X_{t,n} - Y_{t,n})^2\, dt] = 0 \,.$$

Inoltre, dai Teoremi 10.1.3 e 10.1.4 si ha

$$\mathrm{E}[(I(X_n) - I(Y_n))^2] = \mathrm{E}[I(X_n - Y_n)^2] = \mathrm{E}[\textstyle\int_0^T (X_{t,n} - Y_{t,n})^2\, dt] \,,$$

da cui, sulla base del precedente risultato, si ha

$$\lim_n \mathrm{E}[(I(X_n) - I(Y_n))^2] = 0 \,.$$

La tesi segue sulla base delle assunzioni fatte e della Definizione 10.2.4. □

In modo parallelo per quanto visto per i p.a. semplici si hanno i seguenti Teoremi.

Teorema 10.2.6. *Se $X \in \mathcal{M}_T^2$, l'integrale stocastico di Itô è una v.a. tale che* $\mathrm{E}[I(X)] = 0$ *e*

$$\mathrm{Var}[I(X)] = \mathrm{E}[I(X)^2] = \int_0^T \mathrm{E}[X_t^2]\, dt = \mathrm{E}[\textstyle\int_0^T X_t^2\, dt] < \infty \,.$$

Dimostrazione. Si consideri una successione di p.a. semplici $(X_n)_{n\geq 1}$ che approssima il p.a. X in $\mathcal{M}_T^2$. Per quanto riguarda la prima parte, tenendo presente che $\mathrm{E}[I(X_n)] = 0$ dal Teorema 10.1.3, sulla base della disuguaglianza di Jensen (Teorema 4.2.6) si ha

$$\mathrm{E}[I(X)]^2 = \mathrm{E}[I(X) - I(X_n)]^2 = \mathrm{E}[I(X - X_n)]^2 \leq \mathrm{E}[I(X - X_n)^2] = \mathrm{E}[(I(X) - I(X_n))^2] \,.$$

Quindi, per $n \to \infty$ segue che $\mathrm{E}[I(X)] = 0$ dalla Definizione 10.2.4. Per quanto riguarda la seconda parte, dal momento che $\mathcal{M}_T^2$ è uno spazio vettoriale, si ha

$$\left| \sqrt{\mathrm{E}[\textstyle\int_0^T X_t^2\, dt]} - \sqrt{\mathrm{E}[\textstyle\int_0^T X_{t,n}^2\, dt]} \right| \leq \sqrt{\mathrm{E}[\textstyle\int_0^T (X_t - X_{t,n})^2\, dt]} \,.$$

La precedente relazione implica che

$$\lim_n \mathrm{E}[\textstyle\int_0^T X_{t,n}^2\, dt] = \mathrm{E}[\int_0^T X_t^2\, dt]\,,$$

essendo $\lim_n \mathrm{E}[\int_0^T (X_t - X_{t,n})^2 dt] = 0$ sulla base del Teorema 10.2.3. In modo simile, la Definizione 10.2.4 implica che

$$\lim_n \mathrm{E}[(\textstyle\int_0^T X_{t,n}\, dB_t)^2] = \mathrm{E}[(\int_0^T X_t\, dB_t)^2]\,,$$

ovvero

$$\lim_n \mathrm{E}[I(X_n)^2] = \mathrm{E}[I(X)^2]\,.$$

Inoltre, dal momento che $X_n \in \mathcal{S}_T^2$, dal Teorema 10.1.3 si ha

$$\mathrm{E}[(\textstyle\int_0^T X_{t,n}\, dB_t)^2] = \mathrm{E}[\int_0^T X_{t,n}^2\, dt]$$

e quindi i due limiti coincidono. □

Teorema 10.2.7. *Dati i p.a. $X, Y \in \mathcal{M}_T^2$, per ogni $a, b \in \mathbb{R}$ si ha*

$$I(aX + bY) = aI(X) + bI(Y)\,.$$

Dimostrazione. Si consideri due successioni di p.a. semplici $(X_n)_{n\geq 1}$ e $(Y_n)_{n\geq 1}$ che approssimano i p.a. X e Y in $\mathcal{M}_T^2$, ovvero tali che $\lim_n \mathrm{E}[\int_0^T (X_t - X_{t,n})^2 dt] = 0$ e $\lim_n \mathrm{E}[\int_0^T (Y_t - Y_{t,n})^2 dt] = 0$. Tenendo presente la disuguaglianza evidenziata nella dimostrazione del Teorema 10.2.5, si ha

$$\begin{aligned}\mathrm{E}[\textstyle\int_0^T (aX_t + bY_t - aX_{t,n} - bY_{t,n})^2\, dt] &= \mathrm{E}[\int_0^T (a(X_t - X_{t,n}) + b(Y_t - Y_{t,n}))^2\, dt] \\ &\leq 2a^2\, \mathrm{E}[\textstyle\int_0^T (X_t - X_{t,n})^2\, dt] + 2b^2\, \mathrm{E}[\int_0^T (Y_t - Y_{t,n})^2\, dt]\,.\end{aligned}$$

Dunque, risulta

$$\lim_n \mathrm{E}[\textstyle\int_0^T (aX_t + bY_t - aX_{t,n} - bY_{t,n})^2\, dt] = 0\,,$$

ovvero la successione di p.a. semplici $(aX_n + bY_n)_{n\geq 1}$ approssima il p.a. $aX + bY$ in $\mathcal{M}_T^2$. Dunque, dalla Definizione 10.2.4 si ha

$$\lim_n \mathrm{E}[(I(aX + bY) - I(aX_n + bY_n))^2] = 0\,,$$

ovvero, tenendo presente il Teorema 10.1.4, risulta anche

$$\lim_n \mathrm{E}[(I(aX + bY) - aI(X_n) - bI(Y_n))^2] = 0\,.$$

Inoltre, si ha

$$\begin{aligned}\mathrm{E}[(aI(X) + bI(Y) - aI(X_n) - bI(Y_n))^2] &= \mathrm{E}[(a(I(X) - I(X_n)) + b(I(Y) - I(Y_n)))^2] \\ &\leq 2a^2\, \mathrm{E}[(I(X) - I(X_n))^2] + 2b^2\, \mathrm{E}[(I(X) - I(X_n))^2]\,.\end{aligned}$$

Poichè risulta $\lim_n \mathrm{E}[(I(X) - I(X_n))^2] = 0$ e $\lim_n \mathrm{E}[(I(Y) - I(Y_n))^2] = 0$ dalla Definizione 10.2.4, si ha infine che

$$\lim_n \mathrm{E}[(aI(X) + bI(Y) - aI(X_n) - bI(Y_n))^2] = 0\,.$$

Confrontando i precedenti limiti si ottiene la tesi. □

Proposizione 10.2.8. *Dati i p.a. $X, Y \in \mathcal{M}_T^2$, si ha*

$$\mathrm{Cov}[I(X), I(Y)] = \mathrm{E}[I(X)I(Y)] = \int_0^T \mathrm{E}[X_t Y_t]\, dt = \mathrm{E}[\textstyle\int_0^T X_t Y_t\, dt]\,.$$

Dimostrazione. È analoga a quella della Proposizione 10.1.5. □

• **Esempio 10.2.2.** Se $X = (X_t)_{[0,T]}$ è un p.a. per cui risulta che $X_t = \mathbf{1}_{[0,T]}(t)$, si vuole verificare che

$$I(X) = \int_0^T dB_t = B_T\,.$$

Evidentemente, X è un p.a. semplice e dunque $X \in \mathcal{M}_T^2$. Inoltre, si può banalmente scegliere la successione $(X_n)_{n\geq 1}$ con $X_n = X$ e quindi sussiste

$$I(X_n) = \sum_{k=0}^{n-1} (B_{t_{k+1,n}} - B_{t_{k,n}}) = B_T$$

dal momento che la somma è telescopica. Quindi, si ha

$$I(X_n) \overset{L^2}{\to} B_T$$

e dalla Definizione 10.2.4 risulta

$$I(X) = B_T\,.$$

Dunque, in questo caso l'integrale stocastico di Itô è una v.a. con legge Normale $\mathcal{N}(0,T)$. In modo coerente, dal Teorema 10.2.6 risulta $\mathrm{E}[I(X)] = 0$ e $\mathrm{Var}[I(X)] = T$. □

• **Esempio 10.2.3.** Se $X = (B_t)_{[0,T]}$, si vuole verificare che

$$I(X) = \int_0^T B_t\, dB_t = \frac{1}{2}\,(B_T^2 - T)\,.$$

Nell'Esempio 10.2.1 è stato verificato che $X \in \mathcal{M}_T^2$. Dunque, posto $0 = t_{0,n} < t_{1,n} < \ldots < t_{n,n} = T$, si consideri la successione di p.a. semplici $(X_n)_{n\geq 1}$, dove $X_n = (X_{t,n})_{t\in[0,T]}$ e

$$X_{t,n} = \sum_{k=0}^{n-1} B_{t_{k,n}}\, \mathbf{1}_{[t_{k,n}, t_{k+1,n}[}(t)\,.$$

Inoltre, dal momento che l'integrale stocastico non dipende dalla scelta della successione, si ponga $t_{k,n} = \frac{kT}{n}$. Tenendo presente l'identità

$$a(b-a) = \frac{1}{2}\,(b^2 - a^2) - \frac{1}{2}\,(b-a)^2\,,$$

dove $a, b \in \mathbb{R}$, si ha

$$\begin{aligned} I(X_n) &= \sum_{k=0}^{n-1} B_{t_{k,n}}(B_{t_{k+1,n}} - B_{t_{k,n}}) = \frac{1}{2}\sum_{k=0}^{n-1}(B_{t_{k+1,n}}^2 - B_{t_{k,n}}^2) - \frac{1}{2}\sum_{k=0}^{n-1}(B_{t_{k+1,n}} - B_{t_{k,n}})^2 \\ &= \frac{1}{2}\,B_T^2 - \frac{1}{2}\sum_{k=0}^{n-1}(B_{t_{k+1,n}} - B_{t_{k,n}})^2\,. \end{aligned}$$

Dunque, tenendo presente il Teorema 9.4.11 si ha

$$\frac{1}{2}\,B_T^2 - \frac{1}{2}\sum_{k=0}^{n-1}(B_{t_{k+1,n}} - B_{t_{k,n}})^2 \xrightarrow{L^2} \frac{1}{2}\,(B_T^2 - T)$$

per $n \to \infty$. Dalla Definizione 10.2.4 si ottiene infine il risultato richiesto. In questo caso, l'integrale di Itô è una trasformata lineare del quadrato di una v.a. con legge Normale $\mathcal{N}(0,T)$, ovvero $I(X)$ si distribuisce con legge Gamma $\mathcal{G}(-\frac{1}{2}T, T, \frac{1}{2})$. Inoltre, dal Teorema 10.2.6 si ha che $\mathrm{E}[I(X)] = 0$ e $\mathrm{Var}[I(X)] = \frac{T^2}{2}$. Si osservi infine che il risultato ottenuto nell'integrazione stocastica è simile a quello che si verifica in un ambito di integrazione con funzioni deterministiche, eccetto la presenza di un secondo termine. Questo termine è dovuto al fatto che la variazione di un moto Browniano non è nulla (vedi Teorema 9.4.11). □

Dai due precedenti esempi, è evidente che non è opportuno calcolare l'integrale di Itô mediante la Definizione 10.2.4. Tecniche pratiche di integrazione saranno presentate negli Esempi 10.3.5-10.3.8 sulla base della Formula di Itô introdotta nel prossimo paragrafo. Tuttavia, nel caso di un processo deterministico, si può ottenere un risultato generale. In particolare, il seguente Teorema fornisce la legge di un integrale di Wiener.

Teorema 10.2.9. *Sia $X = (X_t)_{t\in[0,T]}$ un p.a. tale che $X_t = g(t)$ q.c. con $g : \mathbb{R} \to \mathbb{R}$. Se si ha $\int_0^T g(t)^2 dt < \infty$, allora l'integrale di Wiener $I(X) = \int_0^T g(t) dB_t$ è una v.a. con legge Normale $\mathcal{N}(0, \int_0^T g(t)^2 dt)$.*

Dimostrazione. Tenendo presente il Teorema 10.2.5, si consideri la successione di p.a. semplici $(X_n)_{n\geq 1}$, dove $X_n = (X_{t,n})_{t\in[0,T]}$ con

$$X_{t,n} = \sum_{k=0}^{n-1} g(t_k)\,\mathbf{1}_{[t_k,t_{k+1}[}(t)\,,$$

che approssima il p.a. X. Dalla Definizione 10.1.2 si ha

$$I(X_n) = \sum_{k=0}^{n-1} g(t_k)(B_{t_{k+1}} - B_{t_k})\,.$$

Gli incrementi di un moto Browniano sono indipendenti per l'assunzione *iii*) della Definizione 9.4.1 e la legge di ogni incremento è Normale per l'assunzione *ii*) della Definizione 9.4.1. Dunque, tenendo presente l'Esempio 7.3.5, la v.a. $I(X_n)$ si distribuisce con legge Normale tale che $\mathrm{E}[I(X_n)] = 0$ e

$$\mathrm{Var}[I(X_n)] = \sum_{k=0}^{n-1} g(t_k)^2 (t_{k+1} - t_k)\,.$$

Inoltre, dalla definizione dell'integrale di Riemann, si ha

$$\lim_n \mathrm{Var}[I(X_n)] = \int_0^T g(t)^2 dt\,.$$

Dal momento che sulla base della Definizione 10.2.4 si ha $I(X_n) \xrightarrow{L^2} I(X)$ per $n \to \infty$, allora $I(X)$ si distribuisce necessariamente con legge Normale $\mathcal{N}(0, \int_0^T g(t)^2 dt)$ (si veda l'Esercizio 8.1.1). □

• **Esempio 10.2.4.** Se $X = (X_t)_{t\in[0,T]}$ è tale che $X_t = t$, allora si ha il seguente integrale di Wiener

$$I(X) = \int_0^T t\, dB_t \, .$$

Sulla base del Teorema 10.2.9, la v.a. $I(X)$ si distribuisce dunque con legge Normale $\mathcal{N}(0, \frac{T^3}{3})$. □

Seguendo un percorso analogo a quello che si considera nella teoria dell'integrazione ordinaria, di seguito viene dato il concetto di processo integrale.

Definizione 10.2.10. Si consideri uno spazio di probabilità $(\Omega, \mathcal{F}, P)$ con filtrazione $(\mathcal{F}_t)_{t\in[0,\infty[}$ e un moto Browniano $B = (B_t)_{t\in[0,\infty[}$ adattato alla filtrazione. Dato il p.a. $X \in \mathcal{M}_T^2$, si dice *processo integrale* il p.a. $M = (M_t)_{t\in[0,T]}$ tale che

$$M_t = \int_0^t X_s\, dB_s \, .$$ □

Il seguente Teorema considera le principali caratteristiche del processo integrale.

Teorema 10.2.11. *Se $M = (M_t)_{t\in[0,T]}$ è un processo integrale, allora M è unico q.c. e $M \in \mathcal{M}_T^2$. Inoltre, M è una martingala rispetto alla filtrazione $(\mathcal{F}_t)_{t\in[0,\infty[}$ con traiettorie continue q.c.*

Dimostrazione. Si veda Da Prato (2014). □

• **Esempio 10.2.5.** Dall'Esempio 10.2.3, si consideri il processo integrale $M = (M_t)_{t\in[0,T]}$ tale che

$$M_t = \int_0^t B_s\, dB_s = \frac{1}{2}\,(B_t^2 - t) \, .$$

Tenendo presente l'Esempio 10.2.1, si ha

$$\mathrm{E}[\textstyle\int_0^T M_t^2\, dt] \le \frac{1}{4}\,\mathrm{E}[\int_0^T (B_t^4 + t^2)\, dt] = \frac{1}{4}\int_0^T \mathrm{E}[B_t^4]\, dt + \frac{1}{4}\int_0^T t^2\, dt = \frac{T^3}{3} < \infty \, ,$$

e dunque risulta $M \in \mathcal{M}_T^2$ a conferma della prima parte del Teorema 10.2.11. Inoltre, si ha

$$\mathrm{E}[|M_t|] \le \frac{1}{2}\,\mathrm{E}[B_t^2] + \frac{t}{2} = t < \infty \, .$$

Tenendo presente che il moto Browniano è una martingala (Proposizione 9.4.5), per ogni $s < t$ risulta

$$\begin{aligned} \mathrm{E}[B_t^2 \mid \mathcal{F}_s] &= \mathrm{E}[(B_t - B_s + B_s)^2 \mid \mathcal{F}_s] \\ &= \mathrm{E}[(B_t - B_s)^2 \mid \mathcal{F}_s] + 2\,\mathrm{E}[(B_t - B_s)B_s \mid \mathcal{F}_s] + \mathrm{E}[B_s^2 \mid \mathcal{F}_s] \\ &= \mathrm{E}[(B_t - B_s)^2] + 2B_s\,\mathrm{E}[B_t - B_s]^2 + B_s^2 = t - s + B_s^2 \, , \end{aligned}$$

ovvero

$$\mathrm{E}[B_t^2 - t \mid \mathcal{F}_s] = B_s^2 - s \, ,$$

da cui infine si ha

$$\mathrm{E}[M_t \mid \mathcal{F}_s] = M_s \, .$$

Dunque, a conferma della seconda parte del Teorema 10.2.11, si ha che il p.a. M è una martingala. □

10.3. Formula di Itô

Di seguito viene considerata un'ampia classe di processi aleatori, detti processi di Itô, che generalizza la famiglia dei processi integrali presentata nella precedente sezione. Inoltre, viene introdotta la cosiddetta formula di Itô, che permette di determinare la struttura di una trasformata di un processo di Itô e che risulta fondamentale per il calcolo differenziale stocastico.

Definizione 10.3.1. Si consideri uno spazio di probabilità $(\Omega, \mathcal{F}, P)$ con filtrazione $(\mathcal{F}_t)_{t\in[0,\infty[}$. Un p.a. $X = (X_t)_{t\in[0,T]}$ è detto *processo di Itô* o *processo di diffusione* se si ha

$$X_t = X_0 + \int_0^t \alpha_s \, ds + \int_0^t \beta_s \, dB_s \,,$$

dove $\alpha = (\alpha_t)_{t\in[0,T]}$ e $\beta = (\beta_t)_{t\in[0,T]}$ sono p.a. tali che $\alpha, \beta \in \mathcal{M}_T^2$. □

Quando si considera il processo di Itô, è uso comune adottare la seguente notazione alternativa

$$dX_t = \alpha_t \, dt + \beta_t \, dB_t \,,$$

e denominare dX_t come *differenziale stocastico*. La precedente espressione viene detta anche *notazione differenziale di Itô*. Tuttavia, si deve evidenziare che il differenziale stocastico non ha un significato matematico ben definito e dovrebbe sempre essere inteso nel senso rigoroso della Definizione 10.3.1. La notazione differenziale di Itô va dunque vista solamente come un modo efficiente di scrivere il processo di Itô.

• **Esempio 10.3.1.** Considerato il p.a. $X = (B_t)_{[0,T]}$ e tenendo presente l'Esempio 10.2.2, si ha

$$B_t = \int_0^t dB_s \,.$$

Dunque, il moto Browniano è un processo di Itô, dal momento che è immediato verificare che $\alpha, \beta \in \mathcal{M}_T^2$, essendo $\alpha_t = 0$ e $\beta_t = 1$. □

• **Esempio 10.3.2.** Considerato il p.a. $X = (B_t^2)_{[0,T]}$ e tenendo presente l'Esempio 10.2.3 si ha

$$B_t^2 = \int_0^t ds + 2\int_0^t B_s \, dB_s$$

Dunque, X è un processo di Itô, dal momento che $\alpha, \beta \in \mathcal{M}_T^2$, essendo $\alpha_t = 1$ e $\beta_t = B_t$ (vedi anche Esempio 10.2.1). Inoltre, in notazione differenziale si può scrivere

$$dB_t^2 = dt + 2B_t \, dB_t \,.$$ □

La seguente Proposizione fornisce le caratteristiche principali del processo di Itô.

Proposizione 10.3.2. *Se il p.a. $X = (X_t)_{t\in[0,T]}$ è un processo di Itô, allora X ha traiettorie continue q.c. Inoltre, X è una martingala se il p.a. $\alpha = (\alpha_t)_{t\in[0,T]}$ è tale che $\alpha_t = 0$.*

Dimostrazione. Segue immediatamente dalle assunzioni e dal Teorema 10.2.10. □

Il seguente Teorema considera la celebrata formula di Itô, che fornisce il effetti la rappresentazione integrale di una trasformata del processo di Itô.

Teorema 10.3.3. (Formula di Itô per trasformate univariate) *Se il p.a. X è un processo di Itô e $g : \mathbb{R} \to \mathbb{R}$ è una funzione con derivata seconda continua, allora*

$$g(X_t) = g(X_0) + \int_0^t g'(X_s)\alpha_s\, ds + \int_0^t g'(X_s)\beta_s\, dB_s + \frac{1}{2}\int_0^t g''(X_s)\beta_s^2\, ds\,,$$

ovvero in notazione differenziale di Itô si ha

$$dg(X_t) = g'(X_t)\alpha_t\, dt + g'(X_t)\beta_t\, dB_t + \frac{1}{2}\, g''(X_t)\beta_t^2\, dt = g'(X_t)\, dX_t + \frac{1}{2}\, g''(X_t)\beta_t^2\, dt\,.$$

Dimostrazione. Si veda Da Prato (2014). □

È importante notare che nel precedente Teorema la funzione g non può dipendere da t. Di seguito vengono date alcune esemplificazioni di questa importante formula.

• **Esempio 10.3.3.** Si consideri il moto Browniano $X = (B_t)_{[0,T]}$ e la funzione $g(x) = x^m$ con $m \in \{1, 2, \dots\}$. Dal momento che è stato verificato che X è un processo di Itô nell'Esempio 10.3.1, dal Teorema 10.3.3 si ha

$$B_t^m = m\int_0^t B_s^{m-1}\, dB_s + \frac{m(m-1)}{2}\, \int_0^t B_s^{m-2}\, ds\,,$$

ovvero in notazione differenziale si ha

$$dB_t^m = mB_t^{m-1}\, dB_t + \frac{m(m-1)}{2}\, B_t^{m-2}\, dt\,.$$

In effetti, questa relazione generalizza quella ottenuta nell'Esempio 10.3.2 nel caso particolare $m = 2$. □

• **Esempio 10.3.4.** Dato il moto Browniano $(B_t)_{[0,T]}$, si consideri il p.a. $X = (X_t)_{[0,T]}$ tale che

$$dX_t = \frac{1}{2}\, X_t\, dt + X_t\, dB_t\,.$$

Se si suppone che $X \in \mathcal{M}_T^2$, il p.a. X è un processo di Itô con $\alpha = (\frac{1}{2}X_t)_{t\in[0,T]}$ e $\beta = (X_t)_{t\in[0,T]}$. Inoltre, dal Teorema 10.3.3 si ottiene

$$de^{B_t} = e^{B_t}\, dB_t + \frac{1}{2}\, e^{B_t}\, dt\,.$$

Si deve concludere dunque che $X_t = e^{B_t}$, ovvero si è determinata la struttura del p.a. X che soddisfa alla relazione differenziale stocastica richiesta. In effetti, risulta $X \in \mathcal{M}_T^2$ dal momento che si ha $\mathrm{E}[e^{B_t}] = e^{\frac{1}{2}t}$ tenendo presente la Proposizione 9.4.3, e quindi

$$\mathrm{E}[\textstyle\int_0^T e^{B_t}\, dt] = \displaystyle\int_0^T \mathrm{E}[e^{B_t}]\, dt = 2(e^{\frac{1}{2}T} - 1) < \infty\,.$$ □

Il seguente Teorema estende la formula di Itô nella situazione in cui si considera due processi di Itô e una loro trasformata. La formula può essere ancora generalizzata al caso di più processi di Itô (per maggiori dettagli, si veda Da Prato, 2014, p.121).

Teorema 10.3.4. (Formula di Itô per trasformate bivariate) *Se i p.a. X e Y sono processi di Itô tali che*

$$dX_t = \alpha_{X,t}\,dt + \beta_{X,t}\,dB_t$$

e

$$dY_t = \alpha_{Y,t}\,dt + \beta_{Y,t}\,dB_t\,,$$

mentre $g : \mathbb{R}^2 \to \mathbb{R}$ *è una funzione con derivate parziali seconde continue, allora*

$$\begin{aligned} g(X_t, Y_t) = g(X_0, Y_0) &+ \int_0^t \frac{\partial g(x,y)}{\partial x}\bigg|_{(X_s,Y_s)} \alpha_{X,s}\,ds + \int_0^t \frac{\partial g(x,y)}{\partial y}\bigg|_{(X_s,Y_s)} \alpha_{Y,s}\,ds \\ &+ \int_0^t \frac{\partial g(x,y)}{\partial x}\bigg|_{(X_s,Y_s)} \beta_{X,s}\,dB_s + \int_0^t \frac{\partial g(x,y)}{\partial y}\bigg|_{(X_s,Y_s)} \beta_{Y,s}\,dB_s \\ &+ \frac{1}{2}\int_0^t \frac{\partial^2 g(x,y)}{\partial x^2}\bigg|_{(X_s,Y_s)} \beta_{X,s}^2\,ds + \frac{1}{2}\int_0^t \frac{\partial g(x,y)}{\partial y}\bigg|_{(X_s,Y_s)} \beta_{Y,s}^2\,ds \\ &+ \int_0^t \frac{\partial^2 g(x,y)}{\partial x \partial y}\bigg|_{(X_s,Y_s)} \beta_{X,s}\beta_{Y,s}\,ds\,. \end{aligned}$$

Equivalentemente, in notazione differenziale stocastica si ha

$$\begin{aligned} dg(X_t, Y_t) = &\frac{\partial g(x,y)}{\partial x}\bigg|_{(X_t,Y_t)} \alpha_{X,t}\,dt + \frac{\partial g(x,y)}{\partial y}\bigg|_{(X_t,Y_t)} \alpha_{Y,t}\,dt \\ &+ \frac{\partial g(x,y)}{\partial x}\bigg|_{(X_t,Y_t)} \beta_{X,t}\,dB_t + \frac{\partial g(x,y)}{\partial y}\bigg|_{(X_t,Y_t)} \beta_{Y,t}\,dB_t \\ &+ \frac{1}{2}\,\frac{\partial^2 g(x,y)}{\partial x^2}\bigg|_{(X_t,Y_t)} \beta_{X,t}^2\,dt + \frac{1}{2}\,\frac{\partial^2 g(x,y)}{\partial y^2}\bigg|_{(X_t,Y_t)} \beta_{Y,t}^2\,dt \\ &+ \frac{\partial^2 g(x,y)}{\partial x \partial y}\bigg|_{(X_t,Y_t)} \beta_{X,t}\beta_{Y,t}\,dt\,, \end{aligned}$$

ovvero

$$\begin{aligned} dg(X_t, Y_t) = &\frac{\partial g(x,y)}{\partial x}\bigg|_{(X_t,Y_t)} dX_t + \frac{\partial g(x,y)}{\partial y}\bigg|_{(X_t,Y_t)} dY_t \\ &+ \frac{1}{2}\,\frac{\partial^2 g(x,y)}{\partial x^2}\bigg|_{(X_t,Y_t)} \beta_{X,t}^2\,dt + \frac{1}{2}\,\frac{\partial^2 g(x,y)}{\partial y^2}\bigg|_{(X_t,Y_t)} \beta_{Y,t}^2\,dt \\ &+ \frac{\partial^2 g(x,y)}{\partial x \partial y}\bigg|_{(X_t,Y_t)} \beta_{X,t}\beta_{Y,t}\,dt\,. \end{aligned}$$

Dimostrazione. Si veda Da Prato (2014, p.121). □

• **Esempio 10.3.5.** (Formula ridotta di Itô) Si consideri i p.a. di Itô X e Y, dove $dY_t = dt$. Evidentemente, Y è in effetti un processo deterministico. In questo caso, essendo $\alpha_{Y,t} = 1$ e $\beta_{Y,t} = 0$, la formula di Itô del Teorema 10.3.4 si riduce a

$$dg(X_t, t) = \frac{\partial g(x,y)}{\partial x}\bigg|_{(X_t,t)} dX_t + \frac{\partial g(x,y)}{\partial y}\bigg|_{(X_t,t)} dt + \frac{1}{2}\,\frac{\partial^2 g(x,y)}{\partial x^2}\bigg|_{(X_t,t)} \beta_{X,t}^2\,dt\,,$$

che è anche detta *formula ridotta di Itô*. □

• **Esempio 10.3.6.** (Moto Browniano geometrico) Al fine di esemplificare la formula ridotta di Itô, si supponga che X sia il moto Browniano, ovvero $dX_t = dB_t$, e si consideri inoltre la funzione $g(x,y) = e^{\sigma x + (\mu - \frac{1}{2}\sigma^2)y}$ con $\mu \in \mathbb{R}$ e $\sigma \in]0,\infty[$. Dunque, si ottiene

$$dg(B_t,t) = \sigma g(B_t,t)\, dB_t + \mu g(B_t,t)\, dt\,,$$

ovvero, se si pone $Z = (Z_t)_{[0,T]}$ con $Z_t = e^{(\mu - \frac{1}{2}\sigma^2)t + \sigma B_t}$, si è determinato quindi il p.a. che soddisfa alla relazione differenziale stocastica $dZ_t = \mu Z_t dy + \sigma Z_t dB_t$. Il p.a. Z è detto *moto Browniano geometrico* e riveste grande importanza nelle applicazioni finanziarie. Inoltre, seguendo una terminologia comunemente adottata, μ è detto *parametro di tendenza*, mentre σ è detto *parametro di volatilità*. Il cosiddetto moto Browniano geometrico ridotto si ha per $\mu = 0$ e $\sigma = 1$. Si osservi infine che per $\mu = 0$ il p.a. X si riduce a $dX_t = \sigma X_t\, dB_t$, ovvero si ha $X_t = X_0 + \sigma \int_0^t X_s dB_s$ in termini di integrale di Itô. Quindi, si è ottenuta la risposta formale al problema introdotto all'inizio della Sezione 10.1 e che ha condotto a considerare l'integrale di Itô. □

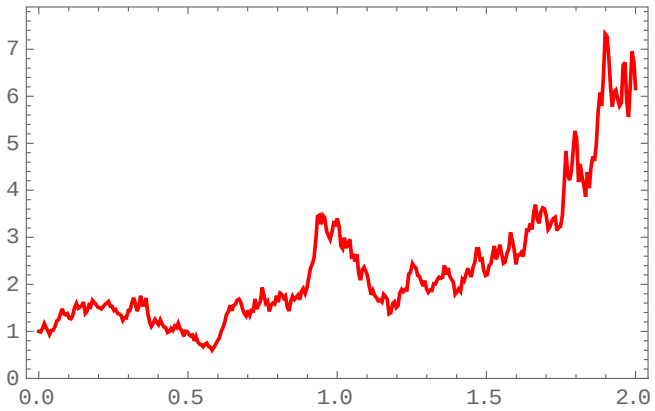


Figura 10.3.1. Una traiettoria del moto Browniano geometrico con $\mu = 1$ e $\sigma = 1$.

• **Esempio 10.3.7.** (Integrazione stocastica per sostituzione) Nella formula ridotta di Itô dell'Esempio 10.3.5 si consideri il p.a. X tale che $dX_t = dB_t$. In questo caso, la formula può essere espressa come

$$dg(B_t,t) = \left.\frac{\partial g(x,y)}{\partial x}\right|_{(B_t,t)} dB_t + \left.\left(\frac{\partial g(x,y)}{\partial y} + \frac{1}{2}\,\frac{\partial^2 g(x,y)}{\partial x^2}\right)\right|_{(B_t,t)} dt\,.$$

Si assuma che g sia una soluzione della cosiddetta equazione del calore, data dalla seguente equazione differenziale

$$\frac{\partial g(x,y)}{\partial y} + \frac{1}{2}\,\frac{\partial^2 g(x,y)}{\partial x^2} = 0\,.$$

In questo caso, la precedente formula ridotta di Itô si semplifica ulteriormente come

$$dg(B_t,t) = \left.\frac{\partial g(x,y)}{\partial x}\right|_{(B_t,t)} dB_t\,,$$

ovvero

$$\int_0^t \left.\frac{\partial g(x,y)}{\partial x}\right|_{(B_s,s)} dB_s = g(B_t,t)\,,$$

che fornisce un'utile espressione per l'*integrazione stocastica per sostituzione*. Nel caso particolare dell'integrale di Itô considerato nell'Esempio 10.2.3, una conveniente soluzione dell'equazione del calore è data da $g(x,y) = \frac{1}{2}(x^2 - y)$, da cui si ottiene immediatamente

$$\int_0^t B_s \, dB_s = \frac{1}{2}\,(B_t^2 - t)\,.$$

Come ulteriore caso particolare, se si vuole ottenere l'integrale di Itô del moto Browniano geometrico ridotto $X = (e^{-\frac{t}{2}+B_t})_{t\in[0,T]}$ considerato nell'Esempio 10.3.6, una conveniente soluzione dell'equazione del calore è data da $g(x,y) = e^{x-\frac{y}{2}}$, da cui si ottiene immediatamente

$$\int_0^t e^{B_s - \frac{s}{2}}\, dB_s = e^{-\frac{t}{2}+B_t}\,.$$

Questo risultato conferma effettivamente che il moto Browniano geometrico ridotto soddisfa alla relazione differenziale stocastica $dX_t = X_t dB_t$. Per ulteriori dettagli sulle soluzioni dell'equazione del calore si veda Calin (2015, p.157). □

• **Esempio 10.3.8.** (Integrazione stocastica del prodotto) Si consideri i processi di Itô X e Y e la funzione $g(x,y) = xy$. Calcolando le derivate parziali e sostituendo opportunamente, dal Teorema 10.3.4 si ha dunque

$$\begin{aligned} d(X_tY_t) &= Y_t\alpha_{X,t}\,dt + X_t\alpha_{Y,t}\,dt + Y_t\beta_{X,t}\,dB_t + X_t\beta_{Y,t}\,dB_t + \beta_{X,t}\beta_{Y,t}dt \\ &= Y_t\,dX_t + X_t\,dY_t + \beta_{X,t}\beta_{Y,t}dt\,. \end{aligned}$$

Si osservi infine che il risultato ottenuto nell'integrazione stocastica è simile a quello che si ha in un ambito di integrazione con funzioni deterministiche, eccetto la presenza di un terzo termine. Questo termine è dovuto al fatto in generale i processi di Itô X e Y non sono indipendenti, in quanto sono trasformate dello stesso moto Browniano. □

• **Esempio 10.3.9.** (Integrazione stocastica per parti) Nell'Esempio 10.3.8 si consideri il caso particolare per cui i p.a. X e Y sono tali che $dX_t = h(t)dt$ e $dY_t = g(B_t)dt$ dove $h : \mathbb{R} \to \mathbb{R}$ è una funzione derivabile e $g : \mathbb{R} \to \mathbb{R}$ è una funzione con derivata seconda continua. In questo caso, tenendo presente la Formula di Itô (Teorema 10.3.3) si ha

$$d(h(t)g(B_t)) = h'(t)g(B_t)dt + h(t)dg(B_t) = h'(t)g(B_t)dt + h(t)g'(B_t)dB_t + \frac{1}{2}\,h(t)g''(B_t)dt\,,$$

ovvero

$$\int_0^t h(s)g'(B_s)\,dB_s = h(s)g(B_s)|_0^t - \int_0^t h'(s)g(B_s)\,ds - \frac{1}{2}\int_0^t h(s)g''(B_s)\,ds\,.$$

Questa utile espressione è detta formula di *integrazione stocastica per parti*. Nel caso particolare dell'integrale di Itô considerato nell'Esempio 10.2.3, posto $h(t) = 1$ e $g(x) = \frac{1}{2}x^2$, si ha

$$\int_0^t B_s\,dB_s = \frac{1}{2}\,B_t^2 - \frac{1}{2}\int_0^t ds = \frac{1}{2}\,(B_t^2 - t)\,,$$

un risultato che è già stato verificato nel medesimo Esempio. Come ulteriore caso particolare, posto $h(t) = t$ e $g(x) = x$, si ottiene

$$\int_0^t s\,dB_s = tB_t - \int_0^t B_s\,ds\,,$$

che fornisce un interessante relazione fra l'integrale di Wiener e l'integrale del moto Browniano. Per ulteriori esemplificazioni del calcolo pratico dell'integrale di Itô si veda Calin (2015). □

10.4. Equazioni differenziali stocastiche

Il calcolo differenziale classico è stato sviluppato principalmente al fine di porre in termini di equazioni differenziali i principi fondamentali che governano i fenomeni fisici nel tempo. Il calcolo differenziale stocastico è stato introdotto con simili scopi nel caso di sistemi che si evolvono con una componente aleatoria. Evidentemente, questo tipo di modelli fornisce una descrizione più efficiente dei fenomeni reali.

La seguente definizione introduce una classe di equazioni differenziali che ha importanza fondamentale nelle applicazioni.

Definizione 10.4.1. Si consideri uno spazio di probabilità $(\Omega, \mathcal{F}, P)$ con filtrazione $(\mathcal{F}_t)_{t\in[0,\infty[}$. Siano inoltre a e b due funzioni tali che $a : \mathbb{R}^2 \to \mathbb{R}$ e $b : \mathbb{R}^2 \to \mathbb{R}$. Un processo di Itô $X = (X_t)_{t\in[0,T]}$ è detto *soluzione dell'equazione differenziale stocastica*

$$dX_t = a(X_t, t)\, dt + b(X_t, t)\, dB_t$$

con *condizione iniziale* $X_0 = C$, se C è una v.a. misurabile rispetto alla σ-algebra $\mathcal{F}_0$, i p.a. $a = (a(X_t, t))_{t\in[0,T]}$ e $b = (b(X_t, t))_{t\in[0,T]}$ sono tali che $a, b \in \mathcal{M}_T^2$ e risulta

$$X_t = C + \int_0^t a(X_s, s)\, ds + \int_0^t b(X_s, s)\, dB_s\,,$$

per ogni t. □

Il seguente Teorema fornisce le condizioni per l'esistenza e l'unicità di una soluzione di un'equazione differenziale stocastica, che in questo caso viene detta *soluzione forte*.

Teorema 10.4.2. *Si consideri uno spazio di probabilità $(\Omega, \mathcal{F}, P)$ con filtrazione $(\mathcal{F}_t)_{t\in[0,\infty[}$. Siano inoltre a e b due funzioni $a : \mathbb{R}^2 \to \mathbb{R}$ e $b : \mathbb{R}^2 \to \mathbb{R}$ tali che esiste $L > 0$ per cui*

$$|a(x,t) - a(y,t)| + |b(x,t) - b(y,t)| \le L|x - y|\,,$$

per ogni $x, y \in \mathbb{R}$ e $t \in [0, T]$, e esiste $K > 0$ per cui

$$|a(x,t)| + |b(x,t)| \le K(1 + |x|)\,,$$

per ogni $x \in \mathbb{R}$ e $t \in [0, T]$. Inoltre, sia C una v.a. misurabile rispetto alla σ-algebra $\mathcal{F}_0$, tale che $\mathrm{E}[C^2] < \infty$. *In questo caso, l'equazione differenziale stocastica*

$$dX_t = a(X_t, t)\, dt + b(X_t, t)\, dB_t\,,$$

con condizione iniziale $X_0 = C$, possiede un'unica soluzione X con traiettorie continue e tale che $\mathrm{E}[\int_0^T X_t^2 dt] < \infty$.

Dimostrazione. Si veda Da Prato (2014, p.134). □

• **Esempio 10.4.1.** (Modello di Black e Scholes) Si consideri l'equazione differenziale stocastica

$$dX_t = \mu_t X_t\, dt + \sigma_t X_t\, dB_t\,,$$

con condizione iniziale $X_0 = C$, dove $\mu : \mathbb{R} \to \mathbb{R}$ e $\sigma : \mathbb{R} \to \mathbb{R}$ sono funzioni continue in $[0, T]$. Dunque, in questo caso si ha $a(x,t) = \mu_t x$ e $b(x,t) = \sigma_t x$. La prima condizione del Teorema 10.4.2 risulta soddisfatta, in quanto

$$|a(x,t) - a(y,t)| + |b(x,t) - b(y,t)| = (|\mu_t| + |\sigma_t|)|x - y| \leq L|x - y| \,,$$

dal momento che esiste $L > 0$ per cui $|\mu_t| + |\sigma_t| \leq L$ essendo μ e σ funzioni continue in $[0, T]$. Anche la seconda condizione del Teorema 10.4.2 risulta soddisfatta in modo simile, in quanto

$$|a(x,t)| + |b(x,t)| = (|\mu_t| + |\sigma_t|)|x| \leq K(1 + |x|) \,.$$

Dunque, esiste una soluzione forte. Si consideri inoltre i processi di Itô U e V, tali che $dU_t = \sigma_t dB_t$ e $dV_t = dt$ e la trasformata

$$g(u,v) = c\, e^{u + \int_0^v (\mu_s - \frac{1}{2}\sigma_s^2) ds} \,,$$

dove c è una determinazione della v.a. C. Si tenga presente che in questo caso si può applicare la formula ridotta di Itô dell'Esempio 10.3.5, data la struttura dei p.a. U e V. Dunque, sostituendo opportunamente in questa formula, si ottiene

$$dg(U_t, t) = \mu_t\, g(U_t, t)\, dt + \sigma_t\, g(U_t, t)\, dB_t \,.$$

Data l'equivalenza di questa espressione con l'equazione differenziale stocastica iniziale, la soluzione $X = (X_t)_{t \in [0,T]}$ è tale che

$$X_t = g(U_t, t) = C\, e^{\int_0^t (\mu_s - \frac{1}{2}\sigma_s^2) ds + \int_0^t \sigma_s dB_s} \,.$$

Questo p.a. è detto *modello di Black e Scholes*, dal momento che è stato introdotto dal matematico Fischer Sheffey Black (1938-1995) e dall'economista Myron Samuel Scholes (1941-). Il modello è in effetti una generalizzazione del moto Browniano geometrico introdotto nell'Esempio 10.3.6, che si ottiene per $\mu_t = \mu$ e $\sigma_t = \sigma$ funzioni costanti. Il modello di Black and Scholes riveste un'importanza fondamentale nella matematica finanziaria. □

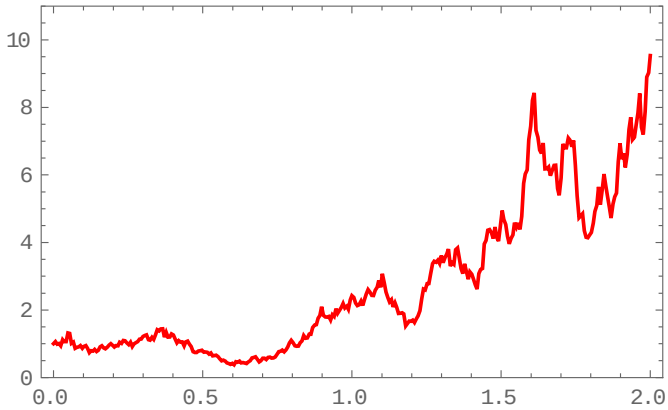


Figura 10.4.1. Una traiettoria del modello di Black e Scholes con $\mu_t = t$ e $\sigma_t = 1$.

• **Esempio 10.4.2.** (Processo di Ornstein e Uhlenbeck) Si consideri l'equazione differenziale stocastica

$$dX_t = -\mu X_t\, dt + \sigma\, dB_t \,,$$

con condizione iniziale $X_0 = C$, dove $\mu, \sigma \in]0, \infty[$. Dunque, in questo caso si ha $a(x,t) = -\mu x$ e $b(x,t) = \sigma$. La prima condizione del Teorema 10.4.2 risulta soddisfatta, in quanto

$$|a(x,t) - a(y,t)| + |b(x,t) - b(y,t)| = |\mu||x - y| \, .$$

Anche la seconda condizione del Teorema 10.4.2 risulta soddisfatta, in quanto

$$|a(x,t)| + |b(x,t)| = |\mu||x| + |\sigma| \, .$$

Quindi, esiste una soluzione forte. Tenendo presente la formula di integrazione stocastica del prodotto introdotta nell'Esempio 10.3.8, con il p.a. Y tale che $Y_t = e^{\mu t}$, si ha

$$d(X_t \, e^{\mu t}) = e^{\mu t} \, dX_t + \mu \, e^{\mu t} \, X_t \, dt = \sigma e^{\mu t} \, dB_t \, .$$

Dalla precedente relazione risulta dunque

$$e^{\mu t} X_t = C + \sigma \int_0^t e^{\mu s} \, dB_s \, ,$$

ovvero, la soluzione $X = (X_t)_{t \in [0,T]}$ dell'equazione differenziale stocastica è tale che

$$X_t = Ce^{-\mu t} + \sigma e^{-\mu t} \int_0^t e^{\mu s} \, dB_s \, .$$

Questo p.a. è detto *processo di Ornstein e Uhlenbeck*, dal momento che è stato introdotto dai fisici Leonard Salomon Ornstein (1880-1941) e George Eugene Uhlenbeck (1900-1988). Anche questo modello ha grande rilevanza nella matematica finanziaria. □

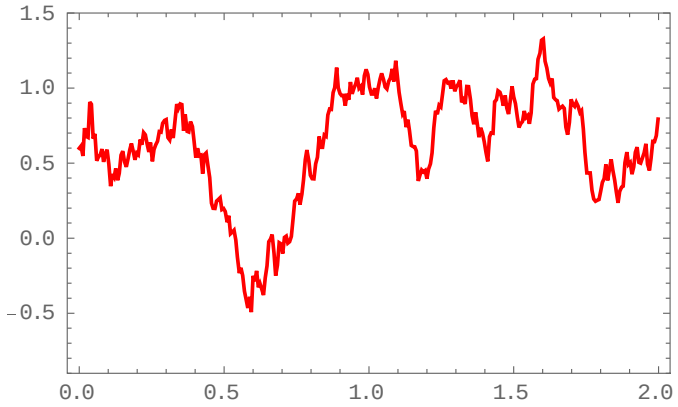


Figura 10.4.2. Una traiettoria del processo di Ornstein e Uhlenbeck con $\mu = 1$ e $\sigma = 1$.

• **Esempio 10.4.3.** (Ponte Browniano) Si consideri l'equazione differenziale stocastica

$$dX_t = \frac{1}{1-t} \, (c - X_t) \, dt + dB_t \, ,$$

con condizione iniziale $X_0 = 0$ e $c \in \mathbb{R}$, dove $t \in [0, 1[$. Dunque, in questo caso si ha $a(x,t) = \frac{c-x}{1-t}$ e $b(x,t) = 1$. La prima condizione del Teorema 10.4.2 risulta soddisfatta, in quanto

$$|a(x,t) - a(y,t)| + |b(x,t) - b(y,t)| = \frac{1}{1-t} \, |x - y| \, .$$

Anche la seconda condizione del Teorema 10.4.2 risulta soddisfatta, in quanto

$$|a(x,t)| + |b(x,t)| = \frac{1}{1-t} \, |c - x| + 1 \, .$$

Dunque, esiste una soluzione forte. Si consideri l'ulteriore processo di Itô $Y = (Y_t)_{t\in[0,1[}$, tale che $Y_t = \frac{1}{1-t}dt$. Evidentemente, Y è un processo deterministico. Dalla formula di integrazione stocastica del prodotto dell'Esempio 10.3.8, si ha

$$d(X_tY_t) = \frac{1}{1-t}\,dX_t + \frac{1}{(1-t)^2}\,X_t\,dt = \frac{c}{(1-t)^2}\,dt + \frac{1}{1-t}\,dB_t\,,$$

ovvero

$$\frac{1}{1-t}\,X_t = X_0 + c\int_0^t \frac{1}{(1-s)^2}\,ds + \int_0^t \frac{1}{1-s}\,dB_s\,.$$

Dunque, la soluzione $X = (X_t)_{t\in[0,1[}$ dell'equazione differenziale stocastica è tale che

$$X_t = ct + (1-t)\int_0^t \frac{1}{1-s}\,dB_s\,.$$

Questo p.a. è detto in modo pittoresco *ponte Browniano*, dal momento che è equivalente ad un moto Browniano le cui traiettorie si originano in $(0,0)$ e terminano in $(1,c)$. □

La seguente Proposizione fornisce un metodo per ottenere la soluzione di un particolare tipo di equazione differenziale stocastica. La tecnica è anche detta *metodo del fattore d'integrazione*.

Proposizione 10.4.3. *Si consideri uno spazio di probabilità $(\Omega,\mathcal{F},P)$ con filtrazione $(\mathcal{F}_t)_{t\in[0,\infty[}$. Sia dato un processo di Itô $X = (X_t)_{t\in[0,T]}$, che è soluzione dell'equazione differenziale stocastica*

$$dX_t = a(X_t,t)\,dt + b(t)X_t\,dB_t\,,$$

dove $b:\mathbb{R}\to\mathbb{R}$ è una funzione continua. Sia dato inoltre il p.a. $Y = (Y_t)_{t\in[0,T]}$ con

$$Y_t = e^{\frac{1}{2}\int_0^t b(s)^2 ds - \int_0^t b(s)dB_s}\,.$$

Si ha

$$d(X_tY_t) = a(X_t,t)Y_t\,dt\,.$$

Dimostrazione. Se $Z = (Z_t)_{t\in[0,T]}$ è il processo di Itô tale che

$$dZ_t = \frac{1}{2}\,b(t)^2\,dt - b(t)\,dB_t$$

e tenendo presente che $Y_t = e^{Z_t}$, dalla formula di Itô (Teorema 10.3.3) si ha

$$dY_t = e^{Z_t}\,dZ_t + \frac{1}{2}\,e^{Z_t}\,b(t)^2\,dt = b(t)^2Y_t\,dt - b(t)Y_t\,dB_t\,.$$

Inoltre, dalla formula di integrazione stocastica del prodotto introdotta nell'Esempio 10.3.8 si ottiene

$$\begin{aligned} d(X_tY_t) &= Y_t(a(X_t,t)\,dt + b(t)X_t\,dB_t) + X_t(b(t)^2Y_t\,dt - b(t)Y_t\,dB_t) - b(t)^2X_tY_t\,dt \\ &= a(X_t,t)Y_t\,dt\,, \end{aligned}$$

da cui segue la tesi. □

Si osservi che il risultato fornito nella Proposizione 10.4.3 permette di ricondurre l'equazione differenziale stocastica originale ad una equazione differenziale ordinaria, che in pratica può essere risolta con l'utilizzo di metodi standard. Nelle applicazioni usuali, molte equazioni differenziali

stocastiche sono esprimibili nella forma considerata nella Proposizione 10.4.3 e dunque questo risultato fornisce un metodo di soluzione molto efficiente.

• **Esempio 10.4.4.** Si consideri l'equazione differenziale stocastica relativa al modello di Black e Scholes dell'Esempio 10.4.1. In questo caso, applicando la Proposizione 10.4.3 si ha

$$Y_t = e^{\frac{1}{2}\int_0^t \sigma_s^2 ds - \int_0^t \sigma_s dB_s} ,$$

mentre

$$d(X_t Y_t) = \mu_t X_t Y_t \, dt .$$

Dunque, tenendo presente che $\{Y_0 = 1\}$ $q.c.$, si ha

$$X_t Y_t = C \, e^{\int_0^t \mu_s ds} ,$$

ovvero

$$X_t = C \, e^{\int_0^t (\mu_s - \frac{1}{2}\sigma_s^2) ds + \int_0^t \sigma_s dB_s} ,$$

che conferma il risultato ottenuto nell'Esempio 10.4.1. □

• **Esempio 10.4.5.** Si consideri l'equazione differenziale stocastica

$$dX_t = \mu \, dt + \sigma X_t \, dB_t ,$$

con condizione iniziale $X_0 = C$ e dove $\mu, \sigma \in \mathbb{R}$. Dunque, in questo caso si ha $a(x,t) = \mu$ e $b(x,t) = \sigma x$. La prima condizione del Teorema 10.4.2 risulta soddisfatta, in quanto

$$|a(x,t) - a(y,t)| + |b(x,t) - b(y,t)| = |\sigma||x - y| .$$

Anche la seconda condizione del Teorema 10.4.2 risulta soddisfatta, in quanto

$$|a(x,t)| + |b(x,t)| = |\mu| + |\sigma x| .$$

Quindi, esiste una soluzione forte. In questo caso, applicando la Proposizione 10.4.3 si ha

$$Y_t = e^{\frac{1}{2}\sigma^2 t - \sigma B_t} ,$$

mentre

$$d(X_t Y_t) = \mu Y_t \, dt ,$$

ovvero, tenendo presente che $\{Y_0 = 1\}$ $q.c.$, risulta

$$X_t Y_t = C + \mu \int_0^t Y_s \, ds .$$

Dunque, la soluzione X dell'equazione differenziale stocastica è tale che

$$X_t = e^{-\frac{1}{2}\sigma^2 t + \sigma B_t} \left(C + \mu \int_0^t e^{\frac{1}{2}\sigma^2 s - \sigma B_s} \, ds \right) .$$ □

Un ulteriore metodo di soluzione è basato sulle cosiddette *equazioni stocastiche esatte*. Si consideri l'equazione differenziale stocastica della Definizione 10.4.1 e si voglia determinare una soluzione forte del tipo $X_t = g(B_t, t)$. In questo caso, l'equazione differenziale stocastica si riduce a

$$dg(B_t, t) = a(g(B_t, t), t) \, dt + b(g(B_t, t), t) \, dB_t .$$

Inoltre, tenendo presente la formula ridotta di Itô (Esempio 10.3.5), l'equazione differenziale stocastica è esatta se esiste una funzione g per cui si ha

$$a(g(x,y),y) = \frac{\partial g(x,y)}{\partial y} + \frac{1}{2}\,\frac{\partial^2 g(x,y)}{\partial x^2}$$

e

$$b(g(x,y),y) = \frac{\partial g(x,y)}{\partial x}\ .$$

Si osservi che in questo caso sussiste la relazione

$$\frac{\partial a(g(x,y),y)}{\partial x} = \frac{\partial b(g(x,y),y)}{\partial y} + \frac{1}{2}\,\frac{\partial^2 b(g(x,y),y)}{\partial x^2}\ .$$

In effetti, dalla formula ridotta di Itô, per questa scelta della funzione g si ha

$$dX_t = dg(B_t,t)\ ,$$

ovvero

$$X_t = g(B_t,t)\ .$$

• **Esempio 10.4.5.** Si consideri il caso specifico dell'equazione differenziale stocastica

$$dX_t = \mu X_t\,dt + \sigma X_t\,dB_t\ .$$

Evidentemente, le condizioni del Teorema 10.4.2 sono soddisfatte e quindi esiste una soluzione forte. In questo caso, essendo $b(g(x,y),y) = \sigma g(x,y)$, si ha

$$\sigma g(x,y) = \frac{\partial g(x,y)}{\partial x}\ ,$$

da cui

$$g(x,y) = e^{\sigma x + h(y)}\ ,$$

dove h è una funzione da determinare. Inoltre, essendo $a(g(x,y),y) = \mu g(x,y)$ risulta

$$\mu\, e^{\sigma x + h(y)} = e^{\sigma x + h(y)} h'(y) + \frac{\sigma^2}{2}\, e^{\sigma x + h(y)}\ ,$$

da cui

$$h(x) = (\mu - \frac{1}{2}\,\sigma^2)t + c$$

con $c \in \mathbb{R}$. Dunque, si ottiene

$$g(x,y) = e^{(\mu - \frac{1}{2}\sigma^2)t + \sigma x + c}\ .$$

Dal momento che $g(0,0) = e^c$, si deve quindi concludere che

$$X_t = X_0\, e^{(\mu - \frac{1}{2}\sigma^2)t + \sigma B_t}\ ,$$

ovvero si è ottenuto il moto Browniano geometrico dell'Esempio 10.3.6. □

• **Esempio 10.4.6.** Si consideri l'equazione differenziale stocastica

$$dX_t = a(X_t)\,dt + b(X_t)\,dB_t$$

e si voglia determinare l'esistenza di una soluzione forte del tipo $X_t = g(B_t)$. In questo caso, la precedente equazione differenziale stocastica si riduce a

$$dg(B_t) = a(g(B_t))\,dt + b(g(B_t))\,dB_t\,.$$

Dalla formula di Itô del Teorema 10.3.3 si ha inoltre

$$dg(B_t) = \frac{1}{2}\,g''(B_t)\,dt + g'(B_t)\,dB_t\,.$$

Dunque, l'equazione differenziale stocastica è esatta se esiste una funzione g per cui si ha

$$a(g(x)) = \frac{1}{2}\,g''(x)$$

e

$$b(g(x)) = g'(x)\,,$$

ovvero sussiste la relazione

$$a(g(x)) = \frac{1}{2}\,b'(g(x))g'(x)\,.$$

In effetti, per questa scelta risulta

$$dX_t = dg(B_t)\,,$$

ovvero

$$X_t = g(B_t)\,.$$

Si consideri il caso specifico dell'equazione differenziale stocastica

$$dX_t = \frac{1}{2}\,X_t\,dt + \sqrt{1 + X_t^2}\,dB_t\,.$$

Le condizioni del Teorema 10.4.2 sono facilmente verificate e quindi esiste una soluzione forte. Sulla base della precedente relazione si ottiene la seguente equazione differenziale

$$g'(x) = \sqrt{1 + g(x)^2}\,,$$

la cui soluzione è data da

$$g(x) = \sinh(x + c)\,,$$

con $c \in \mathbb{R}$. Dunque, si deve concludere che

$$X_t = \sinh(B_t + \operatorname{arcsinh}(X_0))\,.$$ □

10.5. Riferimenti bibliografici

I testi di Applebaum (2009), Baldi (2017), Chung e Williams (1990), Karatzas e Shreve (1991), Klebaner (2005), Kuo (2006), Le Gall (2016), Medvegyev (2007), Øksendal (2003), Pascucci (2011), Protter (2003) e Shreve (2004) sono dedicati al calcolo stocastico. Testi avanzati su questo argomento sono Borodin (2017), Capasso e Bakstein (2015), Cohen e Elliott (2015), Da Prato (2014) e Revuz e

Yor (2005). Per un approccio introduttivo al calcolo stocastico si possono consultare i testi di Brzeźniak e Zastawniak (2002), Calin (2015), Choe (2016) e Wiersema (2008).

10.6. Esercizi svolti

• **Nota.** Al fine di evitare notazioni ridondanti e ripetizioni, nei seguenti esercizi si assume l'esistenza di uno spazio probabilizzato $(\Omega, \mathcal{F}, P)$ e di una filtrazione $(\mathcal{F}_t)_{t\in\mathbb{T}}$, a meno che non venga specificato diversamente.

Sezione 10.1

• **Esercizio 10.1.1.** (Moto Browniano integrato) Se $B = (B_t)_{t\in\mathbb{R}^+}$ è un moto Browniano, si consideri il p.a. $X = (X_t)_{t\in\mathbb{R}^+}$ dove

$$X_t = \int_0^t B_s \, ds \, .$$

Si determini la legge relativa alla v.a. X_t. Il p.a. X è detto moto Browniano integrato.

Soluzione. Dal momento che B ha traiettorie continue $q.c.$, il p.a. X è definito. Inoltre, si consideri la successione di v.a. $(X_{t,n})_{n\geq 1}$ tale che

$$X_{t,n} = \frac{t}{n} \sum_{k=1}^{n} B_{kt/n}$$

e si osservi che $\lim_n X_{t,n} = X_t \; q.c.$ per le proprietà dell'integrale di Riemann. Mediante una semplice manipolazione algebrica $X_{t,n}$ può essere espresso come

$$X_{t,n} = \frac{t}{n} \sum_{k=1}^{n} (n-k+1)(B_{kt/n} - B_{(k-1)t/n}) \, .$$

Per le proprietà del moto Browniano gli incrementi $(B_{kt/n} - B_{(k-1)t/n})$ sono indipendenti con legge Normale $\mathcal{N}(0, \frac{t}{n})$. Quindi, per le proprietà della legge Normale, la v.a. $X_{t,n}$ è distribuita con legge Normale tale che $\mathrm{E}[X_{t,n}] = 0$ e

$$\mathrm{Var}[X_{t,n}] = \frac{t^3}{n^3} \sum_{k=1}^{n} (n-k+1)^2 = \frac{t^3}{n^3} \sum_{k=1}^{n} k^2 = \frac{n(n+1)(2n+1)t^3}{6n^3} \, .$$

Dal momento che $\lim_n \mathrm{E}[X_{t,n}] = 0$ e $\lim_n \mathrm{Var}[X_{t,n}] = \frac{t^3}{3}$, sulla base dell'Esercizio 8.1.1 la v.a. X_t si distribuisce con legge Normale $\mathcal{N}(0, \frac{t^3}{3})$. □

Sezione 10.2

• **Esercizio 10.2.1.** Se $B = (B_t)_{t\in\mathbb{R}^+}$ è un moto Browniano, si consideri il p.a. $X = (X_t)_{t\in[0,T]}$ dove $X_t = \int_0^t e^s dB_s$. Se

$$I(X) = \int_0^T X_t \, dB_t \, ,$$

si calcoli $\mathrm{E}[I(X)]$ e $\mathrm{Var}[I(X)]$.

Soluzione. Dal momento che X_t è un integrale di Wiener, allora si ha $\mathrm{E}[X_t] = 0$ e

$$\mathrm{E}[X_t^2] = \mathrm{Var}[X_t] = \int_0^t e^{2s}\, ds = \frac{1}{2}\,(e^{2t} - 1)\,.$$

Inoltre, risulta

$$\mathrm{E}[\textstyle\int_0^T X_t^2\, dt] = \displaystyle\int_0^T \mathrm{E}[X_t^2]\, dt = \frac{1}{4}\,(e^{2T} - 2T - 1) < \infty$$

e dunque $X \in \mathcal{M}_T^2$. Quindi, si ottiene che $\mathrm{E}[I(X)] = 0$ e

$$\mathrm{Var}[I(X)] = \int_0^T \mathrm{E}[X_t^2]\, dt = \frac{1}{4}\,(e^{2T} - 2T - 1)\,. \qquad \square$$

• **Esercizio 10.2.2.** Se $B = (B_t)_{t\in\mathbb{R}^+}$ è un moto Browniano, si determini $\mathrm{E}[B_s \int_0^T B_t dB_t]$ dove $s \le T$.

Soluzione. Considerato il p.a. degenere $X = (X_t)_{t\in[0,T]}$ dove $X_t = \mathbf{1}_{[0,s]}(t)$, si osservi che risulta $X \in \mathcal{M}_T^2$ e si ponga

$$I(X) = \int_0^T \mathbf{1}_{[0,s]}(t)\, dB_t = B_s\,.$$

Inoltre, se $Y = (Y_t)_{t\in[0,T]}$ dove $Y_t = B_t$, ovviamente si ha $Y \in \mathcal{M}_T^2$ e si assuma che

$$I(Y) = \int_0^T B_t\, dB_t\,.$$

Dunque, si deve concludere che

$$\mathrm{E}[B_s \textstyle\int_0^T B_t\, dB_t] = \mathrm{E}[I(X)I(Y)] = \displaystyle\int_0^T \mathrm{E}[X_t Y_t]\, dt = \int_0^T \mathrm{E}[\mathbf{1}_{[0,s]}(t) B_t]\, dB_t = 0\,. \qquad \square$$

Sezione 10.3

• **Esercizio 10.3.1.** Se $B = (B_t)_{t\in\mathbb{R}^+}$ è un moto Browniano, si consideri il ponte Browniano $X = (X_t)_{t\in[0,1[}$ tale che

$$X_t = (1-t)\int_0^t \frac{1}{1-s}\, dB_s$$

e si calcoli dX_t.

Soluzione. Applicando la formula di integrazione stocastica per il prodotto, ponendo $U_t = \int_0^t (1-s)^{-1} dB_s$ e $V_t = 1 - t$, dal momento che

$$dU_t = \frac{1}{1-t}\, dB_t\,,$$

si ha

$$d(U_t V_t) = V_t\, dU_t + U_t\, dV_t = dB_t - U_t\, dt\,,$$

ovvero

$$dX_t = -\frac{1}{1-t}\, X_t\, dt + dB_t\,.$$

Ovviamente, la precedente espressione rappresenta l'equazione differenziale stocastica che definisce il ponte Browniano. Applicando la formula ridotta di Itô con $g(u,v) = u(1-v)$, si ottiene lo stesso risultato, ovvero

$$dg(U_t,t) = \left.\frac{\partial g(u,v)}{\partial u}\right|_{(U_t,t)} dU_t + \left.\frac{\partial g(u,v)}{\partial v}\right|_{(U_t,t)} dt = (1-t)\,dU_t - U_t\,dt = dB_t - U_t\,dt\,. \qquad \square$$

• **Esercizio 10.3.2.** Se $B = (B_t)_{t\in\mathbb{R}^+}$ è un moto Browniano, si consideri il p.a. $X = (X_t)_{t\in\mathbb{R}^+}$ tale che

$$X_t = tB_t - \int_0^t B_s\,ds$$

e si calcoli dX_t. Si verifichi inoltre che il p.a. X è una martingala.

Soluzione. Si ha

$$dX_t = d(tB_t) - B_t\,dt\,.$$

e, applicando la formula di integrazione stocastica per il prodotto, risulta

$$dX_t = t\,dB_t + B_t\,dt - B_t\,dt = t\,dB_t\,.$$

Dal momento che $\{X_0 = 0\}$ $q.c.$, dalla precedente espressione si ha anche

$$X_t = \int_0^t s\,dB_s\,.$$

Dunque, la v.a. X_t è in effetti un integrale di Wiener distribuito con legge Normale $\mathcal{N}(0, \frac{t^3}{3})$. Questo risultato permette di concludere che il p.a. considerato è una martingala. In effetti, il procedimento considerato è generale, nel senso che se il differenziale stocastico di un p.a. non dipende dalla variazione (ovvero, in termini informali, non contiene l'addendo in dt) allora il p.a. è un integrale di Itô ed è necessariamente una martingala. $\square$

• **Esercizio 10.3.3.** (Processo di Bessel in $\mathbb{R}^2$) Se $B = (B_t)_{t\in\mathbb{R}^+}$ e $W = (W_t)_{t\in\mathbb{R}^+}$ sono due moti Browniani indipendenti, si consideri il p.a. $X = (X_t)_{t\in\mathbb{R}^+}$ tale che

$$X_t = \sqrt{B_t^2 + W_t^2}$$

e si calcoli dX_t. Il p.a. X è comunemente detto processo di Bessel in $\mathbb{R}^2$.

Soluzione. Se $g : \mathbb{R}^2 \to \mathbb{R}$ è una funzione con derivate parziali seconde continue, la generalizzazione della formula di Itô per trasformate bivariate al caso di due moti Browniani indipendenti fornisce l'espressione

$$\begin{aligned} dg(B_t,W_t) = {} & \left.\frac{\partial g(u,v)}{\partial u}\right|_{(B_t,W_t)} dB_t + \left.\frac{\partial g(u,v)}{\partial v}\right|_{(B_t,W_t)} dW_t \\ & + \frac{1}{2}\left(\left.\frac{\partial^2 g(u,v)}{\partial u^2}\right|_{(B_t,W_t)} + \left.\frac{\partial^2 g(u,v)}{\partial v^2}\right|_{(B_t,W_t)}\right) dt\,. \end{aligned}$$

Applicando la precedente formula di Itô con $g(u,v) = \sqrt{u^2+v^2}$, risulta

$$dX_t = \frac{B_t}{X_t}\,dB_t + \frac{W_t}{X_t}\,dW_t + \frac{1}{2}\left(\frac{W_t^2}{X_t^3} + \frac{B_t^2}{X_t^3}\right) dt = \frac{B_t}{X_t}\,dB_t + \frac{W_t}{X_t}\,dW_t + \frac{1}{2X_t}\,dt\,. \qquad \square$$

• **Esercizio 10.3.4.** Se $B = (B_t)_{t\in\mathbb{R}^+}$ è un moto Browniano, si considerino i p.a. $X = (X_t)_{t\in\mathbb{R}^+}$ e $Y = (Y_t)_{t\in\mathbb{R}^+}$ dove $X_t = \int_0^t s dB_s$ e $Y_t = \int_0^t B_s ds$. Si determini il coefficiente di correlazione ρ_{X_t,Y_t}.

Soluzione. Dall'Esercizio 10.3.2 si ha che $X_t + Y_t = tB_t$ e quindi

$$\mathrm{Var}[X_t] + \mathrm{Var}[Y_t] + 2\,\mathrm{Cov}[X_t, Y_t] = t^2\,\mathrm{Var}[B_t]\ .$$

La v.a. X_t è un integrale di Wiener ed è una v.a. con legge Normale $\mathcal{N}(0, \frac{t^3}{3})$. Inoltre, tenendo presente l'Esercizio 10.1.1, il p.a. Y è un moto Browniano integrato e quindi Y_t è una v.a. con legge Normale $\mathcal{N}(0, \frac{t^3}{3})$. Dalla precedente relazione si ha dunque infine che $\mathrm{Cov}[X_t, Y_t] = \frac{t^3}{6}$ e quindi $\rho_{X_t,Y_t} = \frac{1}{2}$. □

• **Esercizio 10.3.5.** Se $B = (B_t)_{t\in\mathbb{R}^+}$ è un moto Browniano, si consideri il p.a. $X = (X_t)_{t\in\mathbb{R}^+}$ dove

$$X_t = \int_0^t e^{\frac{1}{2}s} \cos(B_s)\, dB_s\ .$$

Si determini un'espressione esplicita per il precedente integrale stocastico.

Soluzione. Tenendo presente la formula per l'integrazione stocastica per parti con $h(t) = e^{\frac{1}{2}t}$ e $g(x) = \sin(x)$ si ha

$$\begin{aligned}\int_0^t e^{\frac{1}{2}s} \cos(B_s)\, dB_s &= e^{\frac{1}{2}s} \sin(B_s)\Big|_0^t - \frac{1}{2}\int_0^t e^{\frac{1}{2}s} \sin(B_s)\, ds + \frac{1}{2}\int_0^t e^{\frac{1}{2}s} \sin(B_s)\, ds \\ &= e^{\frac{1}{2}t} \sin(B_t)\ .\end{aligned}$$

□

• **Esercizio 10.3.6.** (Generalizzazione del moto Browniano integrato) Se $B = (B_t)_{t\in\mathbb{R}^+}$ è un moto Browniano, si consideri il p.a. $X = (X_t)_{t\in\mathbb{R}^+}$ con

$$X_t = \int_0^t f(s) B_s\, ds\ ,$$

dove f è una funzione integrabile. Si determini la legge relativa alla v.a. X_t.

Soluzione. Posto $h(t) = \int_0^t f(s) ds$ e $g(s) = s$, manipolando in modo opportuno la formula per l'integrazione stocastica per parti, si ottiene

$$\int_0^t f(s) B_s\, ds = h(t) B_t - \int_0^t h(s)\, dB_s = \int_0^t (h(t) - h(s))\, dB_s\ .$$

Dal momento che la generalizzazione del moto Browniano integrato è espressa come integrale di Wiener, la v.a. X_t è distribuita con legge Normale $\mathcal{N}(0, \int_0^t (h(t) - h(s))^2 ds)$. Nel caso in cui $f(s) = 1$, si ha il moto Browniano integrato, per cui la legge della v.a. X_t è Normale $\mathcal{N}(0, \frac{t^3}{3})$. Si noti che questo risultato è ottenuto in modo più elegante e immediato rispetto all'Esercizio 10.1.1. □

Sezione 10.4

• **Esercizio 10.4.1.** (Processo di Ornstein e Uhlenbeck generalizzato) Si consideri l'equazione differenziale stocastica

$$dX_t = -\mu(t) X_t\, dt + \sigma(t)\, dB_t\,,$$

con condizione iniziale $X_0 = C$, dove $\mu : \mathbb{R} \to \mathbb{R}$ e $\sigma : \mathbb{R} \to \mathbb{R}^+$ sono funzioni localmente limitate, e se ne determini la soluzione.

Soluzione. Si ha $a(x,t) = -\mu(t)x$ e $b(x,t) = \sigma(t)$, per cui

$$|a(x,t) - a(y,t)| + |b(x,t) - b(y,t)| = |\mu(t)||x - y|$$

e

$$|a(x,t)| + |b(x,t)| = |\mu(t)||x| + |\sigma(t)| \,.$$

Dal momento che μ e σ sono localmente limitate, le condizioni per l'esistenza e l'unicità dell'equazione differenziale stocastica sono soddisfatte ed esiste una soluzione forte. Assumendo che $A(t) = \int_0^t \mu(s)ds$ e tenendo presente la formula di integrazione stocastica per il prodotto, si ha

$$d(X_t\, e^{A(t)}) = e^{A(t)}\, dX_t + \mu(t)\, e^{A(t)} X_t\, dt = \sigma(t)\, e^{A(t)}\, dB_t \,.$$

Dalla precedente relazione ed essendo $e^{A(0)} = 1$, risulta dunque

$$e^{A(t)} X_t = C + \int_0^t \sigma(s)\, e^{A(s)}\, dB_s \,,$$

ovvero l'equazione differenziale stocastica ha soluzione $X = (X_t)_{t\in\mathbb{R}^+}$ con

$$X_t = Ce^{-A(t)} + e^{-A(t)} \int_0^t \sigma(s)\, e^{A(s)}\, dB_s \,.$$

Si osservi che $\int_0^t \sigma(s) e^{A(s)} dB_s$ è un integrale di Wiener e quindi la legge della v.a. X_t condizionata all'evento $\{C = c\}$ è Normale $\mathcal{N}(ce^{-A(t)}, e^{-2A(t)} \int_0^t \sigma(s)^2 e^{2A(s)} ds)$. □

• **Esercizio 10.4.2.** Si consideri l'equazione differenziale stocastica

$$dX_t = \, -\mu X_t\, dt + \sigma\, dB_t \,,$$

con condizione iniziale $X_0 = C$, dove $\mu, \sigma \in \,]0, \infty[$, che ha per soluzione il processo di Ornstein e Uhlenbeck. Si determini la legge della v.a. Y tale che $X_t \xrightarrow{\mathcal{L}} Y$ per $t \to \infty$.

Soluzione. Dall'Esercizio 10.4.1 risulta che la soluzione dell'equazione differenziale stocastica considerata è data da $X = (X_t)_{t\in\mathbb{R}^+}$ dove

$$X_t = Ce^{-\mu t} + \sigma e^{-\mu t} \int_0^t e^{\mu s}\, dB_s \,.$$

e la legge della v.a. X_t condizionata all'evento $\{C = c\}$ è Normale $\mathcal{N}(ce^{-\mu t}, \frac{\sigma^2}{2\mu}(1 - e^{-2\mu t}))$. Dal momento che $\mu \in \,]0, \infty[$, si ha

$$\lim_{t\to\infty} \mathrm{E}[X_t] = 0$$

e

$$\lim_{t\to\infty} \mathrm{Var}[X_t] = \frac{\sigma^2}{2\mu} \,.$$

Quindi, in base all'Esercizio 8.1.1, la legge della v.a. Y è Normale $\mathcal{N}(0, \frac{\sigma^2}{2\mu})$. Si osservi che la legge limite non dipende dalla condizione iniziale $X_0 = C$. Inoltre, si assuma che X_0 sia una v.a. indipendente dal moto Browniano e tale che $X_0 \overset{\mathcal{L}}{=} Y$. In questo caso, poichè $\int_0^t e^{\mu s} dB_s$ è una v.a. con legge Normale $\mathcal{N}(0, \frac{1}{2\mu} e^{2\mu t})$, allora è immediato verificare che la legge della v.a. X_t è ancora Normale $\mathcal{N}(0, \frac{\sigma^2}{2\mu})$. □

• **Esercizio 10.4.3.** Assumendo che $t \in [0, 1[$, si consideri l'equazione differenziale stocastica

$$dX_t = -\frac{1}{1-t}\,X_t\,dt + dB_t\,,$$

con condizione iniziale $X_0 = C$, e se ne determini la soluzione.

Soluzione. Tenendo presente l'Esercizio 10.4.1, si ha $\mu(t) = \frac{1}{1-t}$ e $\sigma(t) = 1$, per cui

$$A(t) = \int_0^t \frac{1}{1-s}\,ds = -\log(1-t)\,.$$

L'equazione differenziale stocastica ha soluzione $X = (X_t)_{t\in[0,1[}$ con

$$X_t = C(1-t) + (1-t)\int_0^t \frac{1}{1-s}\,dB_s$$

e la legge della v.a. X_t condizionata all'evento $\{C = c\}$ è Normale $\mathcal{N}(c(1-t), t(1-t))$. Inoltre, se $\{C = 0\}$ $q.c.$, il p.a. X è un ponte Browniano le cui traiettorie si originano in $(0,0)$ e terminano in $(1,0)$. □

• **Esercizio 10.4.4.** Si consideri l'equazione differenziale stocastica

$$dX_t = -\frac{1}{2}\,X_t\,dt + \sqrt{1-X_t^2}\,dB_t\,.$$

con condizione iniziale $X_0 = C$, e se ne determini la soluzione.

Soluzione. Si ha $a(x,t) = -\frac{1}{2}x$ e $b(x,t) = \sqrt{1-x^2}$, per cui

$$|a(x,t) - a(y,t)| + |b(x,t) - b(y,t)| = \frac{1}{2}\,|x-y| + |\sqrt{1-x^2} - \sqrt{1-y^2}| \le 2|x-y|$$

e

$$|a(x,t)| + |b(x,t)| = \frac{1}{2}\,|x| + \sqrt{1-x^2} \le 2(1+|x|)\,.$$

Dunque le condizioni per l'esistenza e l'unicità dell'equazione differenziale stocastica sono soddisfatte ed esiste una soluzione forte. L'equazione differenziale stocastica è esatta e può essere risolta considerando l'equazione differenziale

$$g'(x) = \sqrt{1-g(x)^2}\,,$$

la cui soluzione è data da

$$g(x) = \sin(x+c)\,,$$

con $c \in \mathbb{R}$. Dunque, si deve concludere che la soluzione dell'equazione differenziale stocastica è data da $X = (X_t)_{t\in\mathbb{R}^+}$ con

$$X_t = \sin(B_t + \arcsin(C))\,.$$

In effetti, applicando la formula ridotta di Itô si ottiene

$$d\sin(B_t) = \cos(B_t)\,dB_t - \frac{1}{2}\sin(B_t)\,dt = \sqrt{1-\sin^2(B_t)}\,dB_t - \frac{1}{2}\sin(B_t)\,dt\,.$$ □

• **Esercizio 10.4.5.** Si consideri l'equazione differenziale stocastica

$$dX_t = \frac{1}{X_t}\,dt + X_t\,dB_t\,.$$

con condizione iniziale $X_0 = C$, e se ne determini la soluzione.

Soluzione. L'equazione differenziale stocastica può essere risolta considerando il metodo del fattore d'integrazione. In questo caso si ha

$$Y_t = e^{\frac{1}{2}t - B_t}\,,$$

mentre

$$d(X_tY_t) = \frac{Y_t}{X_t}\,dt\,,$$

ovvero

$$X_tY_t\,d(X_tY_t) = Y_t^2\,dt\,.$$

Tenendo presente che $X_0Y_0 = C$, la precedente equazione differenziale ha come soluzione

$$\frac{1}{2}\,X_t^2Y_t^2 = \frac{1}{2}\,C^2 + \int_0^t Y_s^2\,ds\,,$$

ovvero l'equazione differenziale stocastica ha soluzione $X = (X_t)_{t\in\mathbb{R}^+}$ con

$$X_t = e^{B_t - \frac{1}{2}t}\left(C^2 + 2\int_0^t e^{s-2B_s}\,ds\right)^{\frac{1}{2}}\,. \qquad \square$$

• **Esercizio 10.4.6.** Si consideri l'equazione differenziale stocastica

$$dX_t = \left(\sqrt{1+X_t^2} + \frac{X_t}{2}\right)dt + \sqrt{1+X_t^2}\,dB_t\,.$$

con condizione iniziale $X_0 = C$, e se ne determini la soluzione.

Soluzione. Si può verificare che in modo simile all'Esercizio 10.4.4 le condizioni per l'esistenza e l'unicità dell'equazione differenziale stocastica sono soddisfatte. Dunque, esiste una soluzione forte. L'equazione differenziale stocastica è esatta e può essere risolta considerando le relazioni

$$\sqrt{1+g(x,y)^2} + \frac{g(x,y)}{2} = \frac{\partial g(x,y)}{\partial y} + \frac{1}{2}\,\frac{\partial^2 g(x,y)}{\partial x^2}$$

e

$$\sqrt{1+g(x,y)^2} = \frac{\partial g(x,y)}{\partial x}\,,$$

ovvero risolvendo l'equazione differenziale

$$g(x,y) = -2\,\frac{\partial g(x,y)}{\partial x} + 2\,\frac{\partial g(x,y)}{\partial y} + \frac{1}{2}\,\frac{\partial^2 g(x,y)}{\partial x^2}\,,$$

la cui soluzione è data da

$$g(x,y) = \sinh(x+y+c)\,,$$

con $c \in \mathbb{R}$. Dunque, si deve concludere che l'equazione differenziale stocastica ha soluzione $X = (X_t)_{t\in\mathbb{R}^+}$ con

$$X_t = \sinh(B_t + t + \operatorname{arcsinh}(C)) \, .$$

In effetti, applicando la formula ridotta di Itô e tenendo presente che $\cosh(x) = \sqrt{1 + \sinh^2(x)}$, si ha

$$\begin{aligned} d\sinh(B_t + t) &= \cosh(B_t + t)\, dB_t + \cosh(B_t + t)\, dt + \frac{1}{2} \sinh(B_t + t)\, dt \\ &= \left(\sqrt{1 + \sinh^2(B_t + t)} + \frac{\sinh(B_t + t)}{2} \right) dt + \sqrt{1 + \sinh^2(B_t + t)}\, dB_t \, . \end{aligned}$$

□

Pagina intenzionalmente vuota

Appendice A
Elementi di integrazione

Sia $(\Omega, \mathcal{F})$ uno spazio misurabile. Un'applicazione $g : \Omega \to \mathbb{R}$ è *misurabile* se

$$g^{-1}(B) = \{\omega \in \Omega : g(\omega) \in B\} \in \mathcal{F}$$

per ogni $B \in \mathcal{B}(\mathbb{R})$. In questo caso, g è anche detta funzione Boreliana su $(\Omega, \mathcal{F})$. Per le funzioni misurabili valgono i seguenti risultati.

Teorema A.1. *Se $(\Omega, \mathcal{F})$ è uno spazio misurabile, allora:*
i) g è una funzione misurabile se e solo se $g^{-1}(]-\infty, x]) \in \mathcal{F}$ per ogni $x \in \mathbb{R}$;
ii) se g e h sono funzioni misurabili, la funzione prodotto gh è misurabile (assumendo che sia ben definita in base alla convenzione $0 \cdot \infty = 0$ e $\infty \cdot 0 = 0$);
iii) se g e h sono funzioni misurabili e se $a, b \in \mathbb{R}$, la combinazione lineare $ag + bh$ è una funzione misurabile (assumendo che sia ben definita);
iv) se $(g_n)_{n\geq 1}$ è una successione di funzioni misurabili, allora $\sup_n g_n$, $\inf_n g_n$, $\limsup_n g_n$ e $\liminf_n g_n$ sono funzioni misurabili.
Dimostrazione. Vedi Billingsley (1995). □

Risulta immediato dimostrare che la funzione indicatrice $\mathbf{1}_A$ dell'insieme A è una funzione misurabile se $A \in \mathcal{F}$. Una *funzione semplice* è una combinazione lineare di funzioni indicatrici, ovvero

$$\varphi = \sum_{k=1}^{n} a_k \mathbf{1}_{A_k} ,$$

dove $A_1, \ldots, A_n \in \mathcal{F}$, mentre $a_1, \ldots, a_n \in \mathbb{R}$. Dal Teorema A.1 si ha dunque che una funzione semplice è misurabile.

Teorema A.2. *Se $(\Omega, \mathcal{F})$ è uno spazio misurabile e g è una funzione misurabile, allora esiste una successione di funzioni semplici $(\varphi_n)_{n\geq 1}$ tali che $0 \leq \varphi_1 \leq \varphi_2 \leq \ldots \leq g$ e $\lim_n \varphi_n = g$.*
Dimostrazione. Vedi Billingsley (1995). □

Una successione di funzioni semplici come quella considerata nel precedente Teorema può essere sempre costruita scegliendo φ_n come segue

$$\varphi_n = \sum_{k=1}^{n2^n} \frac{k-1}{2^n} \mathbf{1}_{[(k-1)2^{-n}, k2^{-n}[}(g) + n\mathbf{1}_{[n,\infty[}(g) .$$

La definizione di integrale di una funzione misurabile rispetto ad una misura procede per vari passi, partendo dalla definizione di integrale di una funzione semplice non negativa.

Definizione A.3. Se $(\Omega, \mathcal{F}, \mu)$ è uno spazio misurato e φ è una funzione semplice non negativa, ovvero $a_1, \ldots, a_n \geq 0$, l'integrale di φ è dato da

$$\int \varphi \, d\mu = \sum_{k=1}^{n} a_k \mu(A_k) \, ,$$

con la convenzione che $0 \cdot \infty = 0$. □

Si noti che l'integrale è sempre ben definito, anche se può risultare $\int \varphi d\mu = \infty$. Inoltre, anche se φ non ha una rappresentazione univoca, nel senso che differenti scelte di $A_1, \ldots, A_n$ e $a_1, \ldots, a_n$ possono dare luogo ad una stessa funzione semplice φ, tuttavia è possibile dimostrare che tutte le differenti rappresentazioni producono lo stesso valore per $\int \varphi d\mu$ (si veda Billingsley, 1995). Tenendo presente il Teorema A.2, è quindi immediato dare la definizione di integrale per una funzione misurabile non negativa.

Definizione A.4. Sia $(\Omega, \mathcal{F}, \mu)$ uno spazio misurato e g una funzione misurabile non negativa. Se $(\varphi_n)_{n \geq 1}$ è una successione di funzioni semplici non negative tali che $0 \leq \varphi_1 \leq \varphi_2 \leq \ldots \leq g$, l'integrale di g è dato da

$$\int g \, d\mu = \sup_n \int \varphi_n \, d\mu \, .$$ □

Si noti che l'integrale non dipende dalla scelta di una particolare successione di funzioni elementari. Si definisca inoltre come parte positiva di g la funzione $g^+ = \max(g, 0)$ e come parte negativa di g la funzione $g^- = -\min(g, 0)$. Le funzioni g^+ e g^- sono misurabili e non negative, mentre risulta $g = g^+ - g^-$ e $|g| = g^+ + g^-$. In questo caso, si ha la seguente definizione generale di integrale per una funzione misurabile.

Definizione A.5. Sia $(\Omega, \mathcal{F}, \mu)$ uno spazio misurato e g una funzione misurabile. L'integrale $\int g d\mu$ esiste se e solo se almeno uno degli integrali $\int g^+ d\mu$ o $\int g^- d\mu$ è finito e in questo caso si pone

$$\int g \, d\mu = \int g^+ \, d\mu - \int g^- \, d\mu \, .$$

Se entrambi gli integrali $\int g^+ d\mu$ e $\int g^- d\mu$ sono finiti, la funzione g è detta integrabile. Se entrambi gli integrali non sono finiti (ovvero, si ha la forma del tipo $+\infty - \infty$), allora l'integrale non è definito. □

L'insieme di tutte le funzioni integrabili rispetto a μ viene indicato con $L^1(\mu)$. In generale, l'insieme delle funzioni per cui $\int |g|^p d\mu < \infty$, dove $p > 0$, viene indicato con $L^p(\mu)$. Inoltre, l'integrale

$$\|g\|_p = \int |g|^p \, d\mu$$

è detta norma di ordine p di g, mentre l'integrale

$$\langle g, h \rangle = \int gh \, d\mu$$

è detto prodotto interno. Inoltre, se $A \in \mathcal{F}$ allora l'integrale di $g \in L^1(\mu)$ su A è definito come

$$\int_A g \, d\mu = \int \mathbf{1}_A g \, d\mu \, .$$

Per enfatizzare l'argomento di g, per il precedente integrale si adotta talvolta anche la notazione

$$\int_A g\, d\mu = \int_A g(\omega)\, d\mu(\omega)\ .$$

Nel caso particolare in cui $\Omega = \mathbb{R}$ e $\mathcal{F} = \mathcal{B}(\mathbb{R})$, mentre $\mu = \lambda$ è la misura di Lebesgue, allora l'integrale di g su A (detto integrale di Lebesgue su A) viene scritto semplicemente come

$$\int_A g\, d\mu = \int_A g(x)\, dx\ .$$

Qualora $A = [a, b]$ (dove gli estremi dell'intervallo sono eventualmente non finiti), allora l'integrale di Lebesgue si scrive

$$\int_A g\, d\mu = \int_a^b g(x)\, dx\ .$$

La precedente scrittura può apparire ambigua, dal momento che non distingue l'intervallo $[a, b]$ dall'intervallo $]a, b[$. Tuttavia, poichè $\lambda(\{a\}) = \lambda(\{b\}) = 0$, allora gli integrali estesi ad uno qualsiasi degli intervalli $]a, b[$, $]a, b]$, $[a, b[$ e $[a, b]$ coincidono. Infine, si può dimostrare che l'integrale $\int_a^b g(x)dx$ coincide con quello di Riemann, quando questo è definito (vedi Billingsley, 1995). Tuttavia, vi sono funzioni per cui l'integrale di Lebesgue è definito, mentre l'integrale di Riemann non è definito.

In modo simile, quando $\Omega = \mathbb{R}^n$ e $\mathcal{F} = \mathcal{B}(\mathbb{R}^n)$, mentre $\mu = \lambda^n$ è la misura di Lebesgue in $\mathbb{R}^n$, allora l'integrale di Lebesgue di g su A (detto integrale di Lebesgue su A) viene scritto come

$$\int_A g\, d\mu = \int_A g(x_1, \ldots, x_n)\, dx_1 \ldots dx_n\ .$$

Qualora $A = [a_1, b_1] \times \cdots \times [a_n, b_n]$, allora l'integrale di Lebesgue si scrive

$$\int_A g\, d\mu = \int_{a_1}^{b_1} \ldots \int_{a_n}^{b_n} g(x_1, \ldots, x_n)\, dx_1 \ldots dx_n\ .$$

Di nuovo, si può dimostrare che questo integrale coincide con quello di Riemann, quando questo è definito. Di seguito vengono considerate alcune proprietà dell'integrale.

Teorema A.6. *Sia $(\Omega, \mathcal{F}, \mu)$ uno spazio misurato e $g, h \in L^1(\mu)$:*
i) se $a, b \in \mathbb{R}$, allora

$$\int (ag + bh)\, d\mu = a \int g\, d\mu + b \int h\, d\mu\ ;$$

ii) se $g \leq h$ q.o., allora

$$\int g\, d\mu \leq \int h\, d\mu\ ;$$

iii) se $g \geq 0$ q.o. e $\int g d\mu = 0$, allora $g = 0$ q.o.

Dimostrazione. Vedi Billingsley (1995). □

Dal Teorema A.6 segue che due funzioni $g, h \in L^1(\mu)$ equivalenti $q.o.$ rispetto a μ, ovvero $g \neq h$ solo per un insieme di misura nulla, hanno lo stesso integrale. Inoltre, dal Teorema A.6 è immediato ottenere la disuguaglianza

$$\left| \int g d\mu \right| \leq \int |g| \, d\mu \, ,$$

per cui si ha che una funzione misurabile g è integrabile se e solo se $|g|$ è integrabile. Inoltre, una funzione $g \in L^1(\mu)$ è finita $q.o.$ Il prossimo Teorema fornisce le condizioni per cui le operazioni di limite e di integrazione possono essere scambiate. La prima parte del Teorema è legata al nome del matematico francese Pierre Joseph Louis Fatou (1878-1929), mentre la seconda parte del Teorema è legata al nome del matematico italiano Beppo Levi (1875-1961).

Teorema A.7. *Sia $(\Omega, \mathcal{F}, \mu)$ uno spazio misurato e $(g_n)_{n\geq 1}$ una successione di funzioni di $L^1(\mu)$:*
i) (Lemma di Fatou) se $g_n \geq 0$ per ogni n, allora

$$\int \liminf_n g_n \, d\mu = \liminf_n \int g_n \, d\mu \, ;$$

ii) (Teorema della convergenza dominata di Beppo Levi) Se $\lim_n g_n = g$ $q.o.$ ed esiste una funzione $h \in L^1(\mu)$ tale che $|g_n| \leq h$ $q.o.$, allora $g \in L^1(\mu)$ e

$$\int g \, d\mu = \int \lim_n g_n \, d\mu = \lim_n \int g_n \, d\mu \, ;$$

Dimostrazione. Vedi Billingsley (1995). □

Sia $(\Omega, \mathcal{F}, \mu)$ uno spazio misurato e g una funzione misurabile da $(\Omega, \mathcal{F})$ allo spazio misurabile $(\Lambda, \mathcal{G})$. In questo caso, la misura immagine η di μ mediante g è definita come

$$\eta = \mu(g^{-1}(A))$$

per ogni $A \in \mathcal{G}$. La misura immagine η è anche detta misura indotta da g. Il seguente Teorema fornisce un espressione per l'integrale rispetto alla misura immagine.

Teorema A.8. *Sia $(\Omega, \mathcal{F}, \mu)$ uno spazio misurato e sia η la misura immagine di μ mediante g. Se h è una funzione Boreliana e la funzione $h \circ g = h(g)$ è integrabile, allora*

$$\int h \, d\eta = \int (h \circ g) \, d\mu \, .$$

Dimostrazione. Vedi Billingsley (1995). □

Se $\Omega = \mathbb{R}^n$ e $\mu = \lambda^n$ è la misura di Lebesgue in $\mathbb{R}^n$, si supponga che $g : \mathbb{R}^n \to \mathbb{R}^n$. Sia inoltre g un diffeomorfismo, ovvero un'applicazione biunivoca di un insieme aperto $\mathbb{R}^n$ su un insieme aperto di $\mathbb{R}^n$ differenziabile con continuità e tale che la sua inversa sia differenziabile con continuità. In questo caso, il Teorema A.8 si riduce a

$$\int_B h(y_1, \ldots, y_n) \, dy_1 \ldots dy_n = \int_{g^{-1}(B)} h(g(x_1, \ldots, x_n)) |J(g(x_1, \ldots, x_n))| \, dx_1 \ldots dx_n \, ,$$

dove $J(g(x_1, \ldots, x_n))$ è lo jacobiano relativo a g.

Sia $(\Omega, \mathcal{F}, \nu)$ uno spazio misurato e f una funzione Boreliana non negativa su $(\Omega, \mathcal{F})$. In questo caso, l'applicazione $\eta : \mathcal{F} \to \mathbb{R}$, data da

$$\eta(A) = \int_A f \, d\nu$$

per ogni $A \in \mathcal{F}$, è una misura. In questo contesto, la funzione Boreliana f viene detta *densità*, mentre η è detta *misura definita* dalla densità f.

Teorema A.9. (Teorema di Radon-Nikodym) *Siano η e ν due misure definite sullo spazio misurabile $(\Omega, \mathcal{F})$. Se η è assolutamente continua rispetto a ν, ovvero $\nu(A) = 0$ implica che $\eta(A) = 0$, allora esiste una funzione Boreliana non negativa f tale che η è la misura definita dalla densità f. Inoltre, f è unica $q.o.$ rispetto a ν, ovvero se h è una funzione Boreliana non negativa tale che $\eta(A) = \int_A h d\nu$ per ogni $A \in \mathcal{F}$, allora $f = h$ $q.o.$*

Dimostrazione. Vedi Billingsley (1995). □

In pratica, il Teorema A.9, che prende nome dal matematico austriaco Johann Karl August Radon (1887-1956) e dal matematico polacco Otto Marcin Nikodym (1887-1974), fornisce l'esistenza di una classe di funzioni non negative f che coincidono $q.o.$ rispetto a ν e per questo motivo f viene definita come una versione della densità. Si noti che il Teorema di Radon-Nikodym fornisce sia un modo di costruire una misura, ma anche un metodo per calcolare la misura di ogni $A \in \mathcal{F}$. In effetti, se ν è una misura di cui sono ben note le proprietà (quale la misura di Lebesgue) allora, una volta determinata la densità f, il valore della misura può essere ottenuto per integrazione. In generale, ogni integrazione rispetto a η può essere ricondotta ad un'integrazione rispetto a ν, come si ottiene dal seguente Teorema.

Teorema A.10. *Sia $(\Omega, \mathcal{F}, \nu)$ uno spazio misurato e sia η la misura definita dalla densità f. Se h è una funzione Boreliana e la funzione hf è integrabile, allora*

$$\int h \, d\eta = \int hf \, d\nu \, .$$

Dimostrazione. Vedi Billingsley (1995). □

Il seguente Teorema evidenzia come un calcolare un integrale rispetto ad una misura prodotto per mezzo di una integrazione iterata.

Teorema A.11. (Teorema di Fubini) *Siano $(\Omega_1, \mathcal{F}_1, \mu_1)$ e $(\Omega_2, \mathcal{F}_2, \mu_2)$ due spazi misurati e sia g una funzione misurabile su $(\Omega_1 \times \Omega_2, \mathcal{F}_1 \otimes \mathcal{F}_2)$. Se $g \geq 0$ allora la funzione*

$$h(\omega_2) = \int g(\omega_1, \omega_2) \, d\mu_1(\omega_1)$$

esiste $q.o.$ rispetto a μ_2 ed è misurabile su Ω_2, mentre

$$\int g \, d(\mu_1 \otimes \mu_2) = \int \left(\int g(\omega_1, \omega_2) \, d\mu_1(\omega_1) \right) d\mu_2(\omega_2) \, .$$

Dimostrazione. Vedi Billingsley (1995). □

Il Teorema può essere esteso a n spazi misurati ed è legato al nome del matematico italiano Guido Fubini (1879-1943), anche se più correttamente dovrebbe essere denominato Teorema di Fubini-Tonelli, dal momento che è stato esteso e perfezionato dal matematico italiano Leonida Tonelli (1885-1946).

Riferimenti bibliografici

Il classico testo di riferimento per la Teoria della Misura è Halmos (1974). Ulteriori testi sulla Teoria della Misura e integrazione sono Ambrosio *et al.* (2011), Cohn (2013), Schilling (2005), Taylor (2006), Wheeden e Zygmund (2015). Un testo esaustivo sull'argomento è Bogachev (2007, volume I e II), mentre testi introduttivi sono Capiński e Kopp (2007), Kubrusly (2015) e Ovchinnikov (2013).

Appendice B

Notazioni e abbreviazioni

Capitolo 1

Ω	spazio fondamentale
$\emptyset$	evento impossibile
$\{\omega\}$	evento elementare
E	evento
$\text{card}(E)$	cardinalità di un evento
$\mathcal{E}$	classe di eventi
$\mathcal{P}(\Omega)$	insieme delle parti
$(E_i)_{i\in I}$	classe di eventi I insieme di indici
$(E_i)_{i=1}^{n}$	classe di eventi $I = \{1, \dots, n\}$
$(E_n)_{n\geq 1}$	successione di eventi $I = \{1, 2, \dots\}$
$(E_n)_{n\geq 0}$	successione di eventi $I = \{0, 1, \dots\}$
E^c	evento opposto
$\bigcup_{i\in I} E_i$	evento unione
$E_1 \cup E_2$	evento unione $I = \{1, 2\}$
$\bigcup_{i=1}^{n} E_i$	evento unione $I = \{1, \dots, n\}$
$\bigcup_{n=1}^{\infty} E_n$	evento unione $I = \{1, 2, \dots\}$
$\bigcup_{n=0}^{\infty} E_n$	evento unione $I = \{0, 1, \dots\}$
$\bigcap_{i\in I} E_i$	evento intersezione
$E_1 \cap E_2$	evento intersezione $I = \{1, 2\}$
$\bigcap_{i=1}^{n} E_i$	evento intersezione $I = \{1, \dots, n\}$
$\bigcap_{n=1}^{\infty} E_n$	evento intersezione $I = \{1, 2, \dots\}$
$\bigcap_{n=0}^{\infty} E_n$	evento intersezione $I = \{0, 1, \dots\}$
$E_1 \setminus E_2$	evento differenza
$\liminf_n E_n$	limite inferiore di successione di eventi
$\limsup_n E_n$	limite superiore di successione di eventi
$\lim_n E_n$	limite di successione di eventi
$(\mathcal{E}_i)_{i\in I}$	classe di classi di eventi
$(\mathcal{E}_n)_{n\geq 1}$	successione di classi di eventi $I = \{1, 2, \dots\}$
$(\mathcal{E}_n)_{n\geq 0}$	successione di classi di eventi $I = \{0, 1, \dots\}$
$\bigcup_{i\in I} \mathcal{E}_i$	unione di classi di eventi
$\bigcup_{n=1}^{\infty} \mathcal{E}_n$	unione di classi di eventi $I = \{1, 2, \dots\}$
$\bigcup_{n=0}^{\infty} \mathcal{E}_n$	unione di classi di eventi $I = \{0, 1, \dots\}$
$\bigcap_{i\in I} \mathcal{E}_i$	intersezione di classi di eventi
$\bigcap_{n=1}^{\infty} \mathcal{E}_n$	intersezione di classi di eventi $I = \{1, 2, \dots\}$
$\bigcap_{n=0}^{\infty} \mathcal{E}_n$	intersezione di classi di eventi $I = \{0, 1, \dots\}$
$\mathcal{F}$	σ-algebra
$\sigma(\mathcal{E})$	σ-algebra generata dalla classe $\mathcal{E}$
$(\Omega, \mathcal{F})$	spazio probabilizzabile

$\mathbb{N} = \{0, 1, 2, \dots\}$	insieme degli interi non negativi
$\mathbb{Z} = \{0, \pm 1, \pm 2, \dots\}$	insieme dei numeri interi
$\mathbb{Z}^+ = \{1, 2, \dots\}$	insieme degli interi positivi
$\mathbb{Q} = \{m/n : m, n \in \mathbb{Z}, n \neq 0\}$	insieme dei numeri razionali
$\mathbb{R}$	insieme dei numeri reali
$\mathbb{R}^+$	insieme dei numeri reali positivi
$\mathbb{R}^-$	insieme dei numeri reali negativi
$\mathbb{C}$	insieme dei numeri complessi
$[a, b] = \{x \in \mathbb{R} : a \leq x \leq b\}$	intervallo chiuso
$]a, b[= \{x \in \mathbb{R} : a < x < b\}$	intervallo aperto
$]a, b] = \{x \in \mathbb{R} : a < x \leq b\}$	intervallo aperto a sinistra e chiuso a destra
$[a, b[= \{x \in \mathbb{R} : a \leq x < b\}$	intervallo chiuso a sinistra e aperto a destra
$\mathcal{B}(\mathbb{R})$	σ-algebra di Borel
$(\mathbb{R}, \mathcal{B}(\mathbb{R}))$	spazio misurabile
$\Omega_1 \times \cdots \times \Omega_k$	spazio fondamentale prodotto di k spazi
$E_1 \times \cdots \times E_k$	evento rettangolare di k eventi
$\mathcal{F}_1 \otimes \cdots \otimes \mathcal{F}_k$	σ-algebra prodotto di k spazi
$\mathbb{R}^k$	spazio Euclideo
$\mathcal{B}(\mathbb{R}^k)$	σ-algebra di Borel su $\mathbb{R}^k$
$\prod_{i \in I} \Omega_i$	spazio fondamentale prodotto I insieme indici
$\prod_{i \in I} E_i$	evento rettangolare
$\bigotimes_{i \in I} \mathcal{F}_i$	σ-algebra prodotto

Capitolo 2

$\mu(\,\cdot\,)$	misura
$(\Omega, \mathcal{F}, \mu)$	spazio misurato
$P(\,\cdot\,)$	misura di probabilità
$(\Omega, \mathcal{F}, P)$	spazio probabilizzato
$\mathbf{1}_A(\,\cdot\,)$	funzione indicatrice dell'insieme A
$\lambda(\,\cdot\,)$	misura di Lebesgue su $\mathbb{R}$
$(\mathbb{R}, \mathcal{B}(\mathbb{R}), \lambda)$	spazio misurato
$P(\,\cdot \mid E_0)$	probabilità condizionata
$P_1 \otimes \ldots \otimes P_k$	probabilità prodotto
$\lambda^k(\,\cdot\,)$	misura di Lebesgue su $\mathbb{R}^k$
$(\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k), \lambda^k)$	spazio misurato prodotto
$\bigotimes_{i \in I} P_i$	probabilità prodotto

Capitolo 3

X	variabile aleatoria (v.a.)
$P_X(\,\cdot\,)$	legge (o distribuzione) della v.a. X
$(\mathbb{R}, \mathcal{B}(\mathbb{R}), P_X)$	spazio probabilizzato
$\stackrel{\mathcal{L}}{=}$	uguaglianza in legge
$F_X(\,\cdot\,)$	funzione di ripartizione della v.a. X (f.r.)
$p_X(\,\cdot\,)$	funzione di probabilità della v.a. X (f.p.)
$\lfloor\,\cdot\,\rfloor$	funzione parte intera
$f_X(\,\cdot\,)$	densità di probabilità della v.a. X (d.p.)
$(X_1, \dots, X_k)^{\mathrm{T}}$	vettore di variabili aleatorie (v.v.a.)
X_i	i-esima componente marginale del v.v.a.

$F_X(\cdot)$	funzione di ripartizione del v.v.a. X (f.r.c.)
$F_{X_i}(\cdot)$	funzione di ripartizione della v.a. X_i (f.r.m.)
$p_X(\cdot)$	funzione di probabilità del v.v.a. X (f.p.c.)
$p_{X_i}(\cdot)$	funzione di probabilità della v.a. X_i (f.p.m.)
$f_X(\cdot)$	densità di probabilità del v.v.a. X (d.p.c.)
$f_{X_i}(\cdot)$	densità di probabilità della v.a. X_i (d.p.m.)

Capitolo 4

$\mathrm{E}[X]$	valore atteso della v.a. X
μ_X	valore atteso della v.a. X
X^+	$\max(X, 0)$ parte positiva della v.a. X
X^-	$-\min(X, 0)$ parte negativa della v.a. X
$\mu_{X,r}$	momento di ordine r della v.a. X
$\mathrm{Var}[X]$	varianza della v.a. X
σ_X^2	varianza della v.a. X
$\mathrm{Cov}[X_1, X_2]$	covarianza tra le v.a. X_1 e X_2
σ_{X_1,X_2}	covarianza tra le v.a. X_1 e X_2
ρ_{X_1,X_2}	coefficiente di correlazione tra le v.a. X_1 e X_2
$\mathrm{sgn}(\cdot)$	$\mathbf{1}_{]0,\infty[}(\cdot) - \mathbf{1}_{]-\infty,0[}(\cdot)$ funzione segno
Σ_X	matrice di varianza-covarianza
$\det(\Sigma_X)$	varianza generalizzata

Capitolo 5

$\mathrm{E}[X \mid E_0]$	valore atteso condizionato all'evento E_0
$\mathrm{E}[X \mid \mathcal{F}_0]$	valore atteso condizionato alla σ-algebra $\mathcal{F}_0$
$\mathcal{F}_X$	σ-algebra indotta dalla v.a. X
$\mathrm{E}[X_2 \mid X_1]$	valore atteso condizionato di X_2 a X_1
$P_{X_2\|X_1=x_1}(\cdot)$	legge condizionata di X_2 all'evento $\{X_1 = x_1\}$
$p_{X_2\|X_1=x_1}$	f.p. condizionata di X_2 all'evento $\{X_1 = x_1\}$
$f_{X_2\|X_1=x_1}$	d.p. condizionata di X_2 all'evento $\{X_1 = x_1\}$

Capitolo 6

$\mathcal{B}(n, p)$	legge Binomiale di parametri n e p
$\mathcal{P}(\lambda)$	legge di Poisson di parametro λ
$\mathcal{BN}(k, p)$	legge Binomiale Negativa di parametri k e p
$\mathcal{I}(n, D, N)$	legge Ipergeometrica di parametri n, D e N
$\mathcal{M}(n, p)$	legge Multinomiale di parametri n e p
$\mathcal{N}(\mu, \sigma^2)$	legge Normale di parametri μ e σ
$\phi(\cdot)$	d.p. della v.a. Z con legge Normale ridotta
$\Phi(\cdot)$	f.r. della v.a. Z con legge Normale ridotta
$\mathcal{G}(a, b, k)$	legge Gamma di parametri a, b e k
$\Gamma(k)$	funzione Gamma di Eulero
$\gamma(k, z)$	funzione Gamma incompleta
χ_n^2	legge Chi-quadrato con n gradi di libertà
$\mathcal{BE}(a, b, \alpha, \beta)$	legge Beta di parametri a, b, α e β
$B(\alpha, \beta)$	funzione Beta di Eulero
$I_z(\alpha, \beta)$	funzione Beta incompleta regolarizzata

F_{n_1,n_2}	legge di Snedecor con n_1 e n_2 gradi di libertà
t_n	legge t di Student con n gradi di libertà
$\mathcal{N}_k(\mu_X, \Sigma_X)$	legge Normale Multivariata
$\mathcal{D}(\alpha)$	legge di Dirichlet di parametro α

Capitolo 7

i	$\sqrt{-1}$ unità immaginaria
$\Re(\cdot)$	parte reale
$\Im(\cdot)$	parte immaginaria
$\lvert\cdot\rvert$	modulo
$\varphi_X(\cdot)$	funzione caratterisitica della v.a. X (f.c.)
$G_X(\cdot)$	funzione generatrice della v.a. X (f.g.)
$\zeta(s)$	funzione zeta di Riemann
$\varphi_X(\cdot)$	funzione caratterisitica multivariata (f.c.m.)
$G_X(\cdot)$	funzione generatrice multivariata (f.g.m.)

Capitolo 8

$(X_n)_{n\geq 1}$	successione di v.a.
$(X_n)_{n\geq 0}$	successione di v.a.
$\xrightarrow{\mathcal{L}}$	convergenza in legge
$\xrightarrow{P}$	convergenza in probabilità
$\xrightarrow{q.c.}$	convergenza quasi certa
$\xrightarrow{L^p}$	convergenza in media di ordine p

Capitolo 9

$(X_n)_{t\in\mathbb{T}}$	processo aleatorio (p.a.)
$(\mathcal{F}_t)_{t\in\mathbb{T}}$	filtrazione
$(M_t)_{t\in\mathbb{T}}$	martingala
τ	tempo di arresto
$\mathcal{F}_\tau$	σ-algebra associata al tempo di arresto τ
$(M_{\tau\wedge t})_{t\in\mathbb{T}}$	martingala arrestata al tempo di arresto τ
$a \vee b$	$\max(a, b)$
$a \wedge b$	$\min(a, b)$
$(B_t)_{t\in[0,\infty[}$	moto Browniano
τ_a	tempo di primo passaggio per a

Capitolo 10

$I(X)$	integrale stocastico di Itô del p.a. X
$\mathcal{S}_T^2$	classe dei p.a. semplici
$\mathcal{M}_T^2$	classe dei p.a. integrabili
dX_t	differenziale stocastico del p.a. X

Bibliografia

Ambrosio, L., Da Prato, G. e Mennucci, A. (2011) *Introduction to Measure Theory and Integration*, Edizioni della Normale, Scuola Normale Superiore, Pisa.

Applebaum, D. (2009) *Lévy Processes and Stochastic Calculus*, Cambridge University Press, Cambridge.

Ash, R.B. (2008) *Basic Probability Theory*, Dover, New York.

Ash, R.B. e Doléans-Dade, C.A. (2000) *Probability and Measure Theory*, Academic Press, New York.

Athreya, K.B. e Lahiri, S.N. (2006) *Measure Theory and Probability Theory*, Springer, New York.

Baldi, P. (2017) *Stochastic Calculus*, Springer, New York.

Bass, R.F. (2011) *Stochastic Processes*, Cambridge University Press, Cambridge.

Bhattacharya, R.N. e Waymire, E.C. (2016) *Stochastic Processes with Applications*, seconda edizione, Society for Industrial and Applied Mathematics, Philadelphia.

Billingsley, P. (1995) *Probability and Measure*, terza edizione, Wiley, New York.

Blom, G., Holst, L. e Sandell, D. (1994) *Problems and Snapshots from the World of Probability*, Springer, New York.

Bogachev, V.I. (2007) *Measure Theory*, volume I, Springer, Berlin.

Bogachev, V.I. (2007) *Measure Theory*, volume II, Springer, Berlin.

Borodin, A.N. (2017) *Stochastic Processes*, Birkhäuser, Basel.

Borovkov, A.A. (2013) *Probability Theory*, quinta edizione, Springer, London.

Boucheron, S., Lugosi, G. e Massart, P. (2013) *Concentration Inequalities*, Oxford University Press, Oxford.

Brémaud, P. (2020) *Probability Theory and Stochastic Processes*, Springer, New York.

Brzeźniak, Z. e Zastawniak, T. (2002) *Basic Stochastic Processes*, Springer, London.

Calin, O. (2015) *An Informal Introduction to Stochastic Calculus with Applications*, World Scientific Publishing, Singapore.

Capasso, V. e Bakstein, D. (2015) *An Introduction to Continuous-Time Stochastic Processes*, terza edizione, Birkhäuser, New York.

Capiński, M. e Kopp, P. (2007) *Measure, Integral and Probability*, seconda edizione, Springer, NewYork.

Charalambides, C.A. (2005) *Combinatorial Methods in Discrete Distribution*, Wiley, New York.

Chaumont, L. e Yor, M. (2005) *Exercises in Probability*, Cambridge University Press, Cambridge.

Choe, G.H. (2016) *Stochastic Analysis for Finance with Simulations*, Springer, New York.

Chung, K.L. (2001) *A Course in Probability Theory*, terza edizione, Academic Press, San Diego.

Chung, K. e Williams, R. (1990) *Introduction to Stochastic Integration*, seconda edizione, Birkhäuser, Boston.

Çinlar, E. (2010) *Probability and Stochastics*, Springer, New York.

Cohen, S.N. e Elliott, R.J. (2015) *Stochastic Calculus and Applications*, seconda edizione, Birkhäuser, New York.

Cohn, D.L. (2013) *Measure Theory*, seconda edizione, Birkhäuser, New York.

Da Prato, G. (2014) *Introduction to Stochastic Analysis and Malliavin Calculus*, terza edizione, Edizioni della Normale, Scuola Normale Superiore, Pisa.

Devroye, L. (1986) *Non-uniform Random Variate Generation*, Springer, New York.

Dorogovtsev, A.Y., Silvestrov, D.S., Skorokhod, A.V. and Yadrenko, M.I. (1997) *Probability Theory: Collection of Problems*, Translations of Mathematical Monographs **163**, American Mathematical Society, Providence.

Dubins, L.E. e Savage, L.J. (1965) *How to Gamble If You Must*, McGraw-Hill, New York.

Dudley, R.M. (2004) *Real Analysis and Probabiliy*, Cambridge University Press, Cambridge.

Durrett, R. (2012) *Essentials of Stochastic Processes*, seconda edizione, Springer, New York.

Erickson, M. (2010) *Pearls of Discrete Mathematics*, CRC Press, Boca Raton.

Ethier, S.N. (2010) *The Doctrine of Chances*, Springer, Berlin.

Feller, W. (1968) *An Introduction to Probability Theory and its Applications*, volume I, terza edizione, Wiley, New York.

Feller, W. (1971) *An Introduction to Probability Theory and its Applications*, volume II, seconda edizione, Wiley, New York.

Ferguson, T.S. (1996) *A Course in Large-sample Theory*, Chapman & Hall, New York.

de Finetti, B. (1970) *Theory of Probability*, volume 1, Wiley, New York.

de Finetti, B. (1970) *Theory of Probability*, volume 2, Wiley, New York.

Foata, D. e Fuchs, A. (1998) *Calcul des Probabilités*, seconda edizione, Dunod, Paris.

Forbes, C., Evans, M., Hastings, N. e Peacock, B. (2011) *Statistical Distributions*, quarta edizione, Wiley, New York.

Gordon, H. (1997) *Discrete Probability*, Springer, New York.

Gorroochurn, P. (2012) *Classic Problems of Probability*,Wiley, New York.

Graham, R.L., Knuth, D.E. e Patashnik, O. (1994) *Concrete Mathematics*, seconda edizione, Addison-Wesley, Reading.

Grimmett, G.R. e Stirzaker, D.R. (2001) *One Thousand Exercises in Probability*, Oxford University Press, Oxford.

Gut, A. (2005) *Probability: a Graduate Course*, seconda edizione, Springer, New York.

Hald, A. (1998) *A History of Mathematical Statistics from 1750 to 1930*, Wiley New York.

Hald, A. (2003) *History of Probability and Statistics and Their Applications before 1750*, Wiley, New York.

Halmos, P.R. (1974) *Measure Theory*, Springer, New York.

Hamming, R.W. (1991) *The Art of Probability*, Addison-Wesley, Reading.

Isaac, R. (1995) *The Pleasures of Probability*, Springer, New York.

Jacod, J. e Protter, P. (2004) *Probability Essentials*, seconda edizione, Springer, New York.

Johnson, N., Kemp, A. e Kotz, S. (2005) *Univariate Discrete Distributions*, terza edizione, New York, Wiley.

Johnson, N. e Kotz, S. (1977) *Urn Models and Their Applications*, New York, Wiley.

Johnson, N., Kotz, S. e Balakrishnan, N. (1994) *Continuous Univariate Distributions*, volume 1, seconda edizione, New York, Wiley.

Johnson, N., Kotz, S. e Balakrishnan, N. (1995) *Continuous Univariate Distributions*, volume 2, seconda edizione, New York, Wiley.

Johnson, N., Kotz, S. e Balakrishnan, N. (1997) *Discrete Multivariate Distributions*, New York, Wiley.

Kallenberg, O. (2002) *Foundations of Modern Probability*, seconda edizione, Springer, New York.

Karatzas, I. e Shreve, S.E. (1991) *Brownian Motion and Stochastic Calculus*, seconda edizione, Springer, New York.

Khoshnevisan, D. (2007) *Probability*, American Mathematical Society, Providence.

Klafter, J. e Sokolov, I.M. (2011) *First Steps in Random Walks*, Oxford University Press, Oxford.

Klebaner, F. (2005) *Introduction to Stochastic Calculus with Applications*, seconda edizione, World Scientific Publishing, Singapore.

Knill, O. (2009) *Probability and Stochastic Processes with Applications*, Overseas Press, New Dehli.

Kolmogorov, A.N. (1933) *Grundbegriffe der Wahrscheinlichkeitsrechnung*, Springer, Berlin. Traduzione inglese: Kolmogorov, A.N. (1956) *Foundations of the Theory of Probability*, Chelsea, New York.

Koralov, L.B. e Sinai, Y.G. (2007) *Theory of Probability and Random Processes*, seconda edizione, Springer, New York.
Kotz, S., Balakrishnan, N. e Johnson, N.L. (2000) *Continuous Multivariate Distributions*, volume 1, seconda edizione, New York, Wiley.
Kubrusly, C.S. (2015) *Essentials of Measure Theory*, Springer, New York.
Kuo, H. (2006) *Introduction to Stochastic Integration*, Springer, New York.
Le Gall, J.F. (2016) *Brownian Motion, Martingales, and Stochastic Calculus*, Springer, Berlin.
Leadbetter, R., Cambanis, S. e Pipiras, V. (2014) *A Basic Course in Measure and Probability Theory for Applications*, Cambridge University Press, Cambridge.
Lehmann, E.L. e Casella, G. (1998) *The Theory of Point Estimation*, seconda edizione, Springer, New York.
Letta, G. (1993) *Probabilità Elementare*, Zanichelli, Bologna.
Lin, Z. e Bai, Z. (2010) *Probability Inequalities*, Springer, New York.
Lovász, L., Pelikán, J. e Vesztergombi, K. (2003) *Discrete Mathematics: Elementary and Beyond*, Springer, New York.
Lukacs, E. (1970) *Characteristic Functions*, seconda edizione, Griffin, London.
Medvegyev, P. (2007) *Stochastic Integration Theory*, Oxford University Press, Oxford.
Mishura, Y. e Shevchenko, G. (2017) *Theoretical and Statistical Aspects of Stochastic Processes*, Elsevier, Amsterdam.
Mörters, P. e Peres, Y. (2010) *Brownian Motion*, Cambridge University Press, Cambridge.
Mosteller, F. (1987) *Fifty Challenging Problems in Probability with Solutions*, seconda edizione, Dover Publications, New York.
Olofsson, P. (2015) *Probabilities - The Little Numbers that Rule Our Lives*, seconda edizione, Wiley, New York.
Ore, Ø. (1953) *Cardano: The Gambling Scholar*, Princeton University Press, Princeton.
Ovchinnikov, S. (2013) *Measure, Integral, Derivative*, Springer, New York.
Øksendal, B. (2003) *Stochastic Differential Equations*, sesta edizione, Springer, Berlin.
Pascucci, A. (2011) *PDE and Martingale Methods in Option Pricing*, Bocconi & Springer Series, Springer Italia, Milan.
Petkovšek, M., Wilf, H.S. e Zeilberger, D. (1996) *A = B*, Peters, Wellesley.
Petkovic, M.S. (2009) *Famous Puzzles of Great Mathematicians*, American Mathematical Society, Providence.
Petrov, V.V. (1995) *Limit Theorems of Probability Theory*, Clarendon Press, Oxford.
von Plato, J. (1994) *Creating Modern Probability*, Cambridge University Press, Cambridge.
Protter, P. (2003) *Stochastic Integration and Differential Equations*, seconda edizione, Springer, New York.
Rao, M.M. (2005) *Conditional Measures and Applications*, seconda edizione, Chapman & Hall, New York.
Rao, M.M. e Swift, R.J. (2006) *Probability Theory with Applications*, seconda edizione, Springer, New York.
Resnick, S.I. (2005) *Adventures in Stochastic Processes*, Birkhäuser, Boston.
Resnick, S.I. (2014) *A Probability Path*, Birkhäuser, Boston.
Revuz, D. e Yor, M. (2005) *Continuous Martingales and Brownian Motion*, terza edizione, Springer, New York.
Romano, J.P. e Siegel, A.F. (1986) *Counterexamples in Probability and Statistics*, Wadsworth, London.
Rosenthal, J. (2006) *A First Look at Rigorous Probability Theory*, seconda edizione, World Scientific Publishing, Singapore.
Sasvári, Z. (2013) *Multivariate Characteristic and Correlation Functions*, De Gruyter, Berlin.
Sato, K. (1999) *Lévy Processes and Infinitely Divisible Distributions*, Cambridge University Press, Cambridge.
Schilling, R. (2005) *Measures, Integrals and Martingales*, Cambridge University Press, Cambridge.

Schilling, R.L. e Partzsch, L. (2012) *Brownian Motion*, De Gruyter, Berlin.
Serfling R.J. (1980) *Approximation Theorems of Statistics*, Wiley, New York.
Shiryaev, A.N. (2016) *Probability-1*, terza edizione, Springer, New York.
Shiryaev, A.N. (2019) *Probability-2*, terza edizione, Springer, New York.
Shreve, S. (2004) *Stochastic Calculus for Finance II: Continuous-Time Models*, Springer, New York.
Spivey, M.Z. (2019) *The Art of Proving Binomial Identities*, CRC Press, Boca Raton.
Steutel, F.W. e van Harn, K. (2004) *Infinite Divisibility of Probability Distributions on the Real Line*, Dekker, New York.
Steyer, R. e Nagel, W. (2017) *Probability and Conditional Expectation*, Wiley, Chichester.
Stroock, D.W. (2013) *Mathematics of Probability*, American Mathematical Society, Providence.
Székely, G.J. (1986) *Paradoxes in Probability Theory and Mathematical Statistics*, Kluwer, Dordrecht.
Taylor, M.E. (2006) *Measure Theory and Integration*, American Mathematical Society, Providence.
Tijms, H.C. (2003) *A First Course in Stochastic Models*, Wiley, Chichester.
Tijms, H.C. (2012) *Understanding Probability*, terza edizione, Cambridge University Press, Cambridge.
Uchaikin, V.V. e Zolotarev, V.M. (1999) *Chance and Stability*, De Gruyter, Utrecht.
van der Vaart, A.W. (1998) *Asymptotic Statistics*, Cambridge University Press, Cambridge.
Venkatesh, S.S. (2012) *The Theory of Probability: Explorations and Applications*, Cambridge University Press, Cambridge.
Walsh, J.B. (2012) *Knowing the Odds: An Introduction to Probability*, American Mathematical Society, Providence.
Wheeden, R.L. e Zygmund, A. (2015) *Measure and Integral*, seconda edizione, CRC Press, Boca Raton.
Whittle, P. (2000) *Probability via Expectation*, quarta edizione, Springer, New York.
Wiersema, U. (2008) *Brownian Motion Calculus*, Wiley, Hoboken.
Wilf, H.S. (2006) *Generatingfunctionology*, Peters, Wellesley.
Williams, D. (1992) *Probability with Martingales*, Cambridge University Press, Cambridge.
Wise, G.L. e Hall, E.B. (1993) *Counterexamples in Probability and Real Analysis*, Oxford University Press, Oxford.
Zolotarev, V.M. (1986) *One-dimensional Stable Distributions*, American Mathematical Society, Providence.