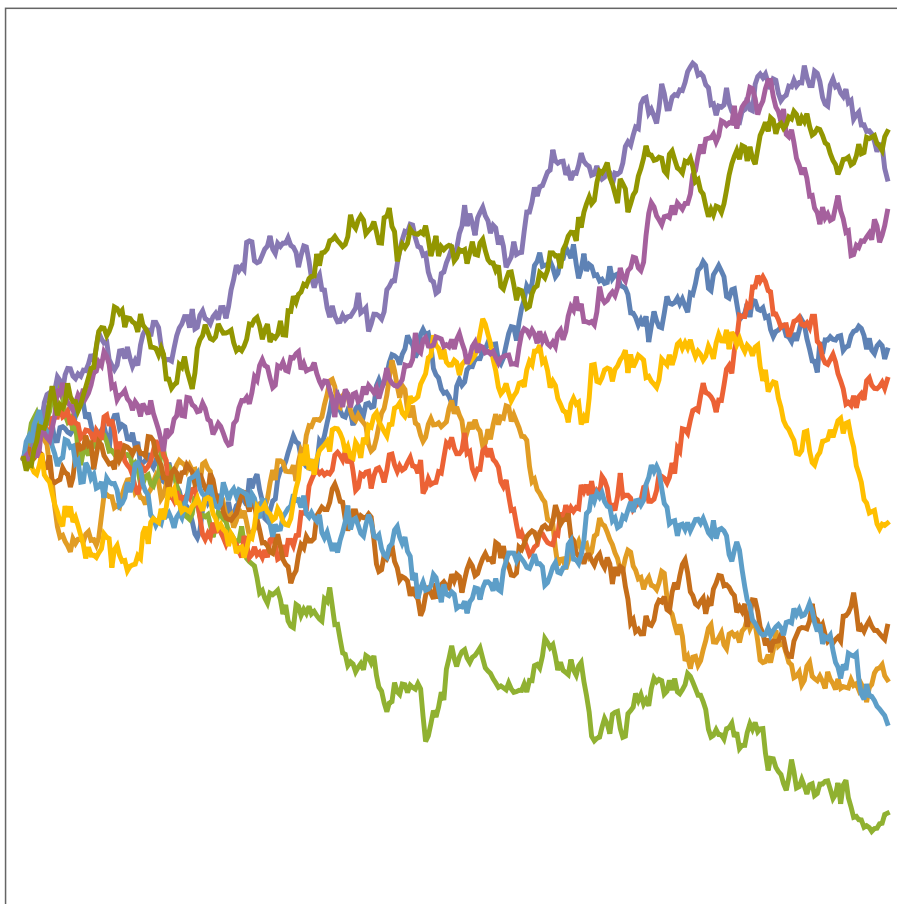


Fondamenti di Probabilità

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 $\Omega = \{t, c\}$. \square

• **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 $k = 1, \dots, 6$, 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.

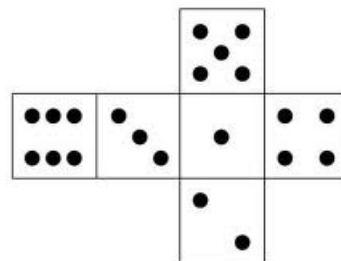


Figura 1.1.3. Comuni dadi da gioco e “sviluppo” di un singolo dado.

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. \square

• **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. \square

• **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. \square

• **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. \square

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

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

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

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 $E_2 \setminus E_1 = \{3, 4\}$. \square

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 $k \neq j = 1, \dots, n$. 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 $n \neq k = 1, 2, \dots$, 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 . \square

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 $n = 1, 2, \dots$, la successione è detta *crescente*, mentre se $E_{n+1} \subset E_n$ per ogni $n = 1, 2, \dots$, 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

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

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

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

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

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. \square

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

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 -algre, 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 $\bigcup_{n=1}^{\infty} E_n \in \mathcal{F}$. \square

Anche se la precedente Proposizione assicura che l'intersezione di σ -algre 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 -algre 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

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

\square

1.4. Spazi probabilizzabili

Se \mathcal{F} è una σ -algebra su Ω , la coppia (Ω, \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. \square

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 $k = 1, \dots, n$, e si combinino in modo da ottenere un unico esperimento aleatorio. In questo caso, lo *spazio fondamentale prodotto* risulta essere il prodotto cartesiano

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

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

$$E = E_1 \times \dots \times E_n .$$

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

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

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

$$\mathcal{F} = \mathcal{F}_1 \otimes \dots \otimes \mathcal{F}_n .$$

La coppia (Ω, \mathcal{F}) è detta *spazio probabilizzabile prodotto*. Inoltre, per quanto riguarda la costruzione dello spazio probabilizzabile (Ω, \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 $\mathbb{R}^n = \mathbb{R} \times \dots \times \mathbb{R}$. In questo caso, la σ -algebra di Borel su \mathbb{R}^n è data da $\mathcal{B}(\mathbb{R}^n) = \mathcal{B}(\mathbb{R}) \otimes \dots \otimes \mathcal{B}(\mathbb{R})$ e può essere costruita a partire dagli insiemi rettangolari di \mathbb{R}^n del tipo $]a_1, b_1] \times \dots \times]a_n, b_n]$ (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 (Ω, \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 . \square

• **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 $\Omega = \Omega_1 \times \dots \times \Omega_n$ relativo alla combinazione degli n esperimenti aleatori risulta allora composto da 2^n elementi del tipo $(\omega_{j_1}, \dots, \omega_{j_n})$, dove $j_k = 1, 2$ e $k = 1, \dots, n$. In effetti, $(\omega_{j_1}, \dots, \omega_{j_n})$ è il risultato dell'esperimento combinato tale che “ ω_{j_1} si è verificato nel primo esperimento, \dots , 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 $\mathcal{F} = \mathcal{F}_1 \otimes \dots \otimes \mathcal{F}_n = \mathcal{P}(\Omega)$ 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. \square



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 . \square

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), Klenke (2020), Le Gall (2022), Rao e Swift (2006), Resnick (2014), Roussas (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 (2021) 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, \dots , 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, \dots , 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. \square

• **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, \dots , l' m -esima pallina nelle ultime $(M - m + 1)$ celle disponibili. Dunque, si ha

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

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

• **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

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

• **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. \square

• **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. \square

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

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

• **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 $\mathcal{E} = \{E_1, \dots, E_n\}$, 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 $I = \{1, \dots, n\}$. In questo caso risulta $\text{card}(\mathcal{F}) = 2^{2^n}$. \square

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. \square

Sezione 1.5

• **Esercizio 1.5.1.** Si consideri lo spazio probabilizzabile prodotto (Ω, \mathcal{F}) con $\Omega = \Omega_1 \times \Omega_2$ e la classe di eventi \mathcal{F}_1 tale che se $E \in \mathcal{F}_1$, 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. \square

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

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

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 $P : \mathcal{F} \rightarrow [0, \infty[$ 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

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

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 (Ω, \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 (Ω, \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 (Ω, \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). \square

Proposizione 2.2.3. *Sia (Ω, \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. \square

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 (Ω, \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. \square

• **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). \square

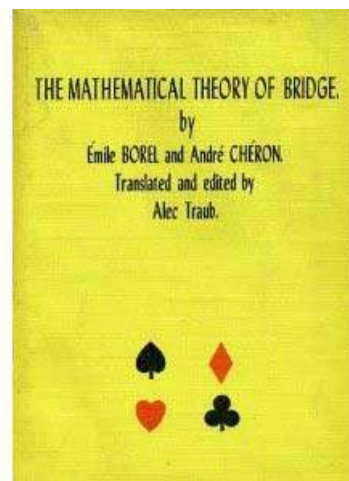
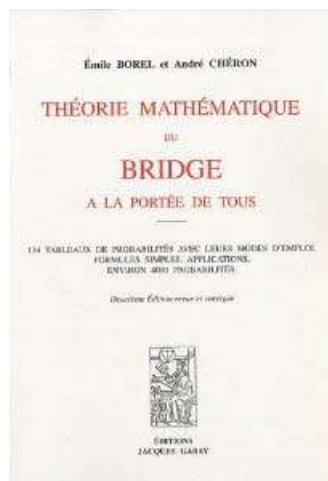


Figura 2.2.1. Frontespizio dell'edizione originale e dell'edizione inglese di *Théorie Mathématique du Bridge à la Portée de Tous* (1940).

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.

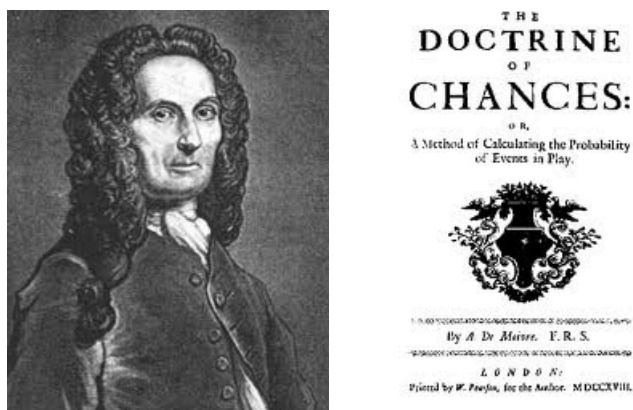


Figura 2.2.2. Abraham de Moivre (1667-1754) e frontespizio di *The Doctrine of Chances* (1718).

Teorema 2.2.5. (Formula di Inclusione ed Esclusione) Sia (Ω, \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

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

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. \square

• **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 $k = 1, \dots, n$ risulta

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

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

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

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. \square



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

Proposizione 2.2.6. Sia (Ω, \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. \square

Proposizione 2.2.7. Sia (Ω, \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. \square

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 (Ω, \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. \square

Proposizione 2.2.9. Sia (Ω, \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. \square

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 (Ω, \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

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

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. \square

Teorema 2.2.11. (Lemma di Fatou) *Se (Ω, \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

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

\square

Proposizione 2.2.12. Se (Ω, \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

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

2.3. Costruzione di spazi probabilizzati

Dato uno spazio probabilizzabile (Ω, \mathcal{F}) , al fine di ottenere il relativo spazio probabilizzato (Ω, \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 $\Omega = \{\omega_1, \dots, \omega_n\}$ ed è sufficiente assegnare le probabilità a ciascuno degli eventi elementari, ovvero

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

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 $k = 1, \dots, n$ 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 $k = 1, \dots, n$, 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).

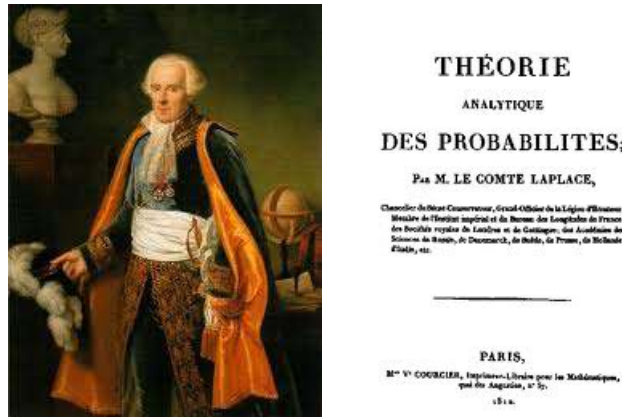


Figura 2.3.1. Pierre-Simon, marchese de Laplace (1749-1827) e frontespizio di *Théorie analytique des probabilités* (1812).

• **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

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

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 $\mu : \mathcal{F} \rightarrow [0, \infty[$ 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

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

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 $\mu(E) = \frac{1}{3}$. \square

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} . \square

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 (Ω, \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.* \square

• **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

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

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

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

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.* \square

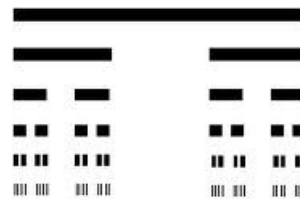


Figura 2.3.3. Georg Cantor (1845-1918) e le prime cinque iterazioni per la costruzione dell'insieme di Cantor.

2.4. Probabilità condizionata

Dato lo spazio probabilizzato (Ω, \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 (Ω, \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

$$P(E | E_0) = \frac{P(E \cap E_0)}{P(E_0)}. \quad \square$$

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

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

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

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

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

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

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

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

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

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

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

o, alternativamente, come

$$P(E_1 \cap E_2) = P(E_1 | E_2)P(E_2).$$

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

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

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

$$P\left(\bigcap_{n=1}^{\infty} E_n\right) = P(E_n | E_1 \cap E_2 \cap \dots \cap E_{n-1}) \times \dots \times P(E_2 | E_1)P(E_1).$$

Teorema 2.4.3. (Legge delle Probabilità Totali) Se (Ω, \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 $P(E_k) > 0$ per ogni $k = 1, \dots, n$, allora per ogni $E \in \mathcal{F}$ si ha

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

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

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

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

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

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).

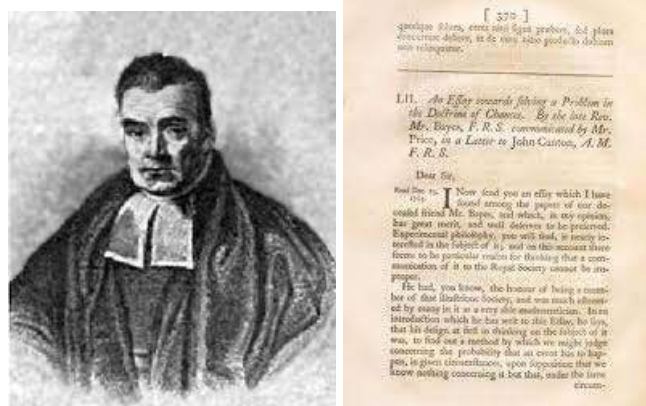


Figura 2.4.1. Probabile ritratto di Thomas Bayes (1702-1761) e frontespizio di *Essay Towards Solving a Problem in the Doctrine of Chances* (1763).

Teorema 2.4.4. (Formula di Bayes) Sia (Ω, \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

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

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. \square

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 $\Omega = \{\omega_1, \dots, \omega_{1000}\}$, 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 $P(E | E_1) = \frac{1}{10}$, $P(E | E_2) = \frac{1}{20}$ e $P(E | E_3) = 0$. 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

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

Applicando inoltre la Formula di Bayes, si è in grado di ottenere la probabilità $P(E_k | E)$, 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

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

e

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

mentre

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

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. \square

• **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 $P(E | E_1) = 1$ e

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

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

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. \square

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

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

Se questo non avviene, si deve verificare necessariamente che

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

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

Definizione 2.5.1. Sia (Ω, \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

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

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 $k = 1, \dots, 6$, 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 $p_2 = \dots = p_6 = \frac{1}{10}$, 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

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

Proposizione 2.5.2. Sia (Ω, \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 . □

Proposizione 2.5.3. Se (Ω, \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 $P(\Omega) = 1$. □

Proposizione 2.5.4. Se (Ω, \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. □

Definizione 2.5.5. Sia (Ω, \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 $k \neq j = 1, \dots, n$, mentre sono detti *completamente indipendenti* se per ogni possibile scelta $1 \leq j_1 < \dots < j_k \leq n$ di k eventi con $k = 2, \dots, n$, si ha

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

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 $\Omega = \{1, \dots, 8\}$, 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 $k = 1, \dots, 8$, 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

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

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 $k = 1, \dots, n$. Sia dunque $\Omega = \Omega_1 \times \dots \times \Omega_n$ 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 $E_1 \in \mathcal{F}_1, \dots, E_n \in \mathcal{F}_n$. 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 (Ω, \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 $k = 1, \dots, 6$, 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. \square

• **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 $\Omega = \Omega_1 \times \dots \times \Omega_{24}$, 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

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

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. □

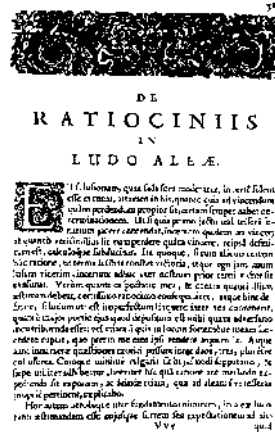


Figura 2.6.3. Christiaan Huygens (1629-1695) e frontespizio di *De ratiociniis in ludo aleae* (1657).

• **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.

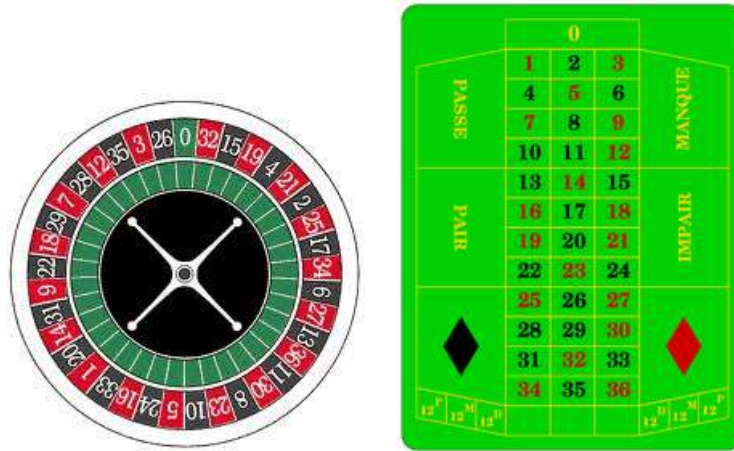


Figura 2.6.4. Disposizione dei numeri sulla *roulette* e sul tavolo verde per una *roulette* di tipo Monte Carlo.

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 $\{(\omega_{j_1}, \dots, \omega_{j_n})\}$ di Ω , ovvero

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

dove $j_k = 1, 2$ e $k = 1, \dots, n$. 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. \square

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 (Ω, \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 (Ω, \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. \square

Teorema 2.7.2. (Legge zero-uno di Kolmogorov) Dato lo spazio probabilizzato (Ω, \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). \square

• **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 . \square

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' semplificazione 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 (Ω, \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, (Ω, \mathcal{F}, P) è uno spazio probabilizzato. \square

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 $m \in \{0, 1, \dots, n\}$ 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 $1 \leq j_1 < \dots < j_k \leq n$ si ha

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

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. \square

• **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 $m \in \{1, \dots, n\}$ 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

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

• **Esercizio 2.2.3.** Dato lo spazio probabilizzato (Ω, \mathcal{F}, P) e considerati due eventi $E_1, E_2 \in \mathcal{F}$, si determini la probabilità che si verificano 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

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

• **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 $P(E_j) = 5^n 6^{-n}$, $P(E_{j_1} \cap E_{j_2}) = 4^n 6^{-n}$, ..., $P(\bigcap_{k=1}^5 E_{j_k}) = 6^{-n}$ e $P(\bigcap_{k=1}^6 E_k) = 0$. In base alla Formula di Inclusione ed Esclusione, la probabilità richiesta è quindi data da

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

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

Sezione 2.3

• **Esercizio 2.3.1.** (Dimostrazione probabilistica à la Paul Erdős) 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 Erdős (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. \square

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 $P(E_2 | E_1) = 1$ e $P(E_2 | E_1^c) = \frac{1}{2}$. 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

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

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

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

Dunque, in modo controintuitivo per un neofita risulta $P(E_1 | E_2) = \frac{2}{3}$ 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 $P(E_1)$, $P(E_2 | E_1)$ e $P(E_2 | E_1^c)$ 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. \square

• **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, $P(E_2 | E_1) = p_1$ e $P(E_2 | E_1^c) = p_2$. 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

$$P(E_2) = P(E_2 | E_1)P(E_1) + P(E_2 | E_1^c)P(E_1^c) = p_1p + p_2(1 - p).$$

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

$$P(E_1 | E_2) = \frac{p_1p}{p_1p + p_2(1 - p)} = \frac{1}{1 + \left(\frac{1}{p} - 1\right)\frac{p_2}{p_1}}.$$

Si osservi che la probabilità $P(E_1 | E_2)$ decresce al diminuire di p e all'aumentare del rapporto $\frac{p_2}{p_1}$. Dal momento che si ha $P(E_1^c | E_2) = 1 - P(E_1 | E_2)$, 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 $P(E_1 | E_2) = \frac{11}{122} \simeq 0.09$, 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 $P(E_1 | E_2) \geq \alpha$ 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

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

• **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

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

e

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

anche se

$$P(E_1 | E_2) < P(E_1 | E_2^c) .$$

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

$$\Omega = \{E_1 \cap E_2 \cap E_3, E_1^c \cap E_2 \cap E_3, E_1 \cap E_2^c \cap E_3, E_1^c \cap E_2^c \cap E_3, \\ E_1 \cap E_2 \cap E_3^c, E_1^c \cap E_2 \cap E_3^c, E_1 \cap E_2^c \cap E_3^c, E_1^c \cap E_2^c \cap E_3^c\}$$

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

$$P(E_1 \cap E_2 \cap E_3) = p_1, P(E_1^c \cap E_2 \cap E_3) = p_2, P(E_1 \cap E_2^c \cap E_3) = p_3, P(E_1^c \cap E_2^c \cap E_3) = p_4, \\ P(E_1 \cap E_2 \cap E_3^c) = p_5, P(E_1^c \cap E_2 \cap E_3^c) = p_6, P(E_1 \cap E_2^c \cap E_3^c) = p_7, P(E_1^c \cap E_2^c \cap E_3^c) = p_8,$$

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

$$P(E_2 \cap E_3) = p_1 + p_2, P(E_2^c \cap E_3) = p_3 + p_4, P(E_2 \cap E_3^c) = p_5 + p_6, P(E_2^c \cap E_3^c) = p_7 + p_8, \\ P(E_1 \cap E_2) = p_1 + p_5, P(E_1^c \cap E_2) = p_2 + p_6, P(E_1 \cap E_2^c) = p_3 + p_7, P(E_1^c \cap E_2^c) = p_4 + p_8,$$

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). \square

• **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

$$P(E_1 | E_2) < P(E_1 | E_1 \cup E_2).$$

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 $P(E_1 | E_2) = P(E_1)$. Inoltre, dal momento che

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

risulta

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

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. \square

Sezione 2.5

• **Esercizio 2.5.1.** Si consideri lo spazio fondamentale $\Omega = \{1, \dots, n\}$, 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). \square

• **Esercizio 2.5.2.** Sia dato lo spazio probabilizzato (Ω, \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 $n_1 < \dots < n_k$ 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. \square

• **Esercizio 2.5.3.** (Paradosso dei tre prigionieri) In uno stato dove esiste ancora la pena capitale, tre prigionieri rinchiusi in un carcere sono condannati a morte. Uno dei tre prigionieri viene graziato mediante una scelta casuale. Il direttore del carcere conosce l'identità del prigioniero graziato, anche se non la può rendere nota. Il primo prigioniero prega il direttore di rivelargli l'identità di uno degli altri due prigionieri che verrà giustiziato, affermando che in questo modo non avrà informazioni sul proprio destino. Inoltre, nel caso che sia lui stesso il prigioniero che riceve la grazia, chiede al direttore di lanciare una moneta bilanciata per scegliere l'identità del condannato fra gli altri due prigionieri. Ingenuamente, quando il primo prigioniero ha conosciuto questa identità, suppone che la propria probabilità di sopravvivenza sia aumentata da $\frac{1}{3}$ a $\frac{1}{2}$, in quanto vi sono solo due condannati rimanenti che possono essere graziati. Si verifichi che la risposta fornita dal direttore non cambia la probabilità di sopravvivenza.

Soluzione. Se si indicano con E_1, E_2 e E_3 rispettivamente gli eventi che il primo, il secondo e il terzo prigioniero siano graziati, si ha $P(E_1) = P(E_2) = P(E_3) = \frac{1}{3}$. Inoltre, se E rappresenta l'evento che il direttore riveli l'identità del secondo prigioniero, si ha $P(E | E_1) = \frac{1}{2}$, $P(E | E_2) = 0$ e $P(E | E_3) = 1$. Dunque, applicando la Legge delle Probabilità Totali si ha

$$P(E) = \sum_{k=1}^3 P(E | E_k)P(E_k) = \frac{1}{2}.$$

La probabilità di sopravvivenza del primo condannato, data la risposta del direttore, risulta

$$P(E_1 | E) = \frac{P(E | E_1)P(E_1)}{P(E)} = \frac{1}{3}.$$

Dunque, gli eventi E_1 e E sono indipendenti essendo $P(E_1 | E) = P(E_1)$. Quindi, la risposta del direttore non aumenta le informazioni del primo prigioniero. \square

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 $j = 1, \dots, N$.

Soluzione. L'esperimento aleatorio analizzato è relativo ad uno spazio fondamentale prodotto dato da $\Omega = \Omega_1 \times \dots \times \Omega_n$, dove $\Omega_k = \{1, \dots, N\}$ rappresenta lo spazio fondamentale relativo alla k -esima estrazione per $k = 1, \dots, n$. Dal momento che 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 $P = P_1 \otimes \dots \otimes P_n$ con un'assegnazione equiprobabile su ogni spazio fondamentale, ovvero $P_k(\{j\}) = N^{-1}$ per $k = 1, \dots, n$. Se $E_{k,j}$ rappresenta l'evento che venga estratto un numero minore o uguale a j nella k -esima estrazione, ovvero $E_{k,j} = \{1, \dots, j\}$, si ha $P_k(E_{k,j}) = \frac{j}{N}$. Dunque, se $E_j = E_{1,j} \times \dots \times E_{n,j}$, 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

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

• **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 $\Omega = \Omega_1 \times \dots \times \Omega_n$, 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 $P = P_1 \otimes \dots \otimes P_n$ con l'assegnazione $P_k(\{1\}) = p$ e $P_k(\{0\}) = 1 - p$ per $k = 1, \dots, n$. Dunque, la probabilità di non avere nessuna corrispondenza è data da $P(\{(0, \dots, 0)\}) = (1 - p)^n$ e dunque la probabilità richiesta risulta

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

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

• **Esercizio 2.6.3.** (Gioco del Craps) Il gioco del Craps è basato sul lancio di due dadi bilanciati. Il giocatore vince se la somma dei punti ottenuti nel lancio è pari a 7 o 11, mentre perde se la somma è pari a 2, 3 o 12. Nel caso in cui la somma sia pari a qualsiasi altro valore j (dove $j = 4, 5, 6, 8, 9, 10$), si continua a lanciare i dadi fino a quando la somma dei punti è pari a j o 7. Il giocatore vince se la somma dei punti è pari a j , mentre il giocatore perde se la somma dei punti è pari a 7. Si determini la probabilità di vincere al gioco del Craps.

Soluzione. L'esperimento aleatorio analizzato è relativo allo spazio fondamentale prodotto dato da $\Omega = \Omega_1 \times \Omega_2$, dove $\Omega_k = \{1, 2, 3, 4, 5, 6\}$ è lo spazio fondamentale relativo al lancio del k -esimo dado per $k = 1, 2$, supponendo che gli eventi elementari rappresentino il numero dei punti. La σ -algebra prodotto \mathcal{F} può essere scelta come $\mathcal{P}(\Omega)$. Inoltre, dal momento che i dadi sono bilanciati e i lanci indipendenti, si considera la probabilità prodotto $P = P_1 \otimes P_2$ con l'assegnazione $P_k(\{j\}) = \frac{1}{6}$ per $k = 1, 2$. Dunque, si ha $P(\{(j, l)\}) = \frac{1}{36}$. Quindi, essendo gli eventi elementari nello spazio fondamentale prodotto equiprobabili, sulla base di una opportuna enumerazione, la probabilità che la somma dei punti ottenuti nel lancio sia pari a j è data da

$$p_j = \frac{\min(j-1, 13-j)}{36}$$

per $j = 2, 3, \dots, 12$. Si indichi con E_j l'evento che il giocatore vinca il gioco se la somma dei punti al primo lancio è pari a j . Evidentemente, si ha

$$P(E_2) = P(E_3) = P(E_{12}) = 0,$$

mentre

$$P(E_7) = p_7 = \frac{1}{6}$$

e

$$P(E_{11}) = p_{11} = \frac{1}{18}.$$

Nel caso in cui $j = 4$, per vincere con n lanci si deve ottenere una somma dei punti pari a 4 all' n -esimo lancio, in modo tale che nei precedenti $(n-2)$ lanci la somma dei punti non sia stata pari a 4 o 7. Dal momento che i lanci dei dadi sono indipendenti, la probabilità di vittoria con esattamente n lanci è data da $p_4^2(1-p_4-p_7)^{n-2}$. Quindi, tenendo presente le proprietà della serie geometrica, la probabilità che il giocatore vinca il gioco avendo ottenuto una somma dei punti pari a 4 nel primo lancio è data da

$$P(E_4) = \sum_{n=2}^{\infty} p_4^2(1-p_4-p_7)^{n-2} = p_4^2 \sum_{n=0}^{\infty} (1-p_4-p_7)^n = \frac{p_4^2}{p_4+p_7} = \frac{1}{36},$$

dal momento che gli eventi di ottenere la vittoria in $n = 2, 3, \dots$ lanci sono incompatibili. In modo analogo, si ha

$$P(E_5) = \frac{p_5^2}{p_5+p_7} = \frac{2}{45}$$

e

$$P(E_6) = \frac{p_6^2}{p_6+p_7} = \frac{25}{396}.$$

Inoltre, dal momento che $p_8 = p_6$, $p_9 = p_5$ e $p_{10} = p_4$, allora si ha $P(E_8) = P(E_6)$, $P(E_9) = P(E_5)$ e $P(E_{10}) = P(E_4)$. Dunque, essendo gli eventi E_2, E_3, \dots, E_{12} incompatibili, la probabilità di vittoria al gioco del Craps è data da

$$P(E_2 \cup E_3 \cup \dots \cup E_{12}) = P(E_2) + P(E_3) + \dots + P(E_{12}) = \frac{244}{495} \simeq 0.4929 .$$

Il gioco del Craps è molto popolare nelle sale da gioco degli Stati Uniti. □

Sezione 2.7

• **Esercizio 2.7.1.** Dato lo spazio probabilizzato (Ω, \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 $P(\limsup_n E_n) = 0$. □

• **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. \square

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 (Ω, \mathcal{F}, P) è uno spazio probabilizzato, un'applicazione $X : \Omega \rightarrow \mathbb{R}$ 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 (Ω, \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. \square

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 (Ω, \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 $X : \Omega \rightarrow S$, dove $S \subset \mathbb{R}$ è un insieme numerabile, affinché 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 (Ω, \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. \square

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. □

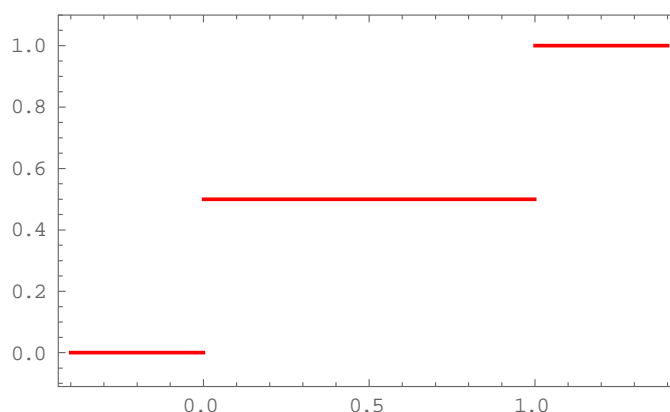


Figura 3.1.1. Funzione di ripartizione relativa alla legge di Bernoulli.

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 (Ω, \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 (Ω, \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

$$F_X(x) = P(X \leq x) \leq P(X \leq y) = F_X(y). \quad \square$$

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

Dimostrazione. Si ha

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

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

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

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

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

Proposizione 3.2.4. Se F_X è la f.r. di una v.a. X definita sullo spazio probabilizzato (Ω, \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

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

ovvero che

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

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

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

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

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

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

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

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

$$\Delta F_X(x) = P(X = x).$$

Dimostrazione. Risulta

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

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

$$\Delta F_X(x) = P(X \leq x) - P(X < x) = P(X = x). \quad \square$$

Proposizione 3.2.6. *Sia F_X la f.r. di una v.a. X definita sullo spazio probabilizzato (Ω, \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

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

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

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

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. \square

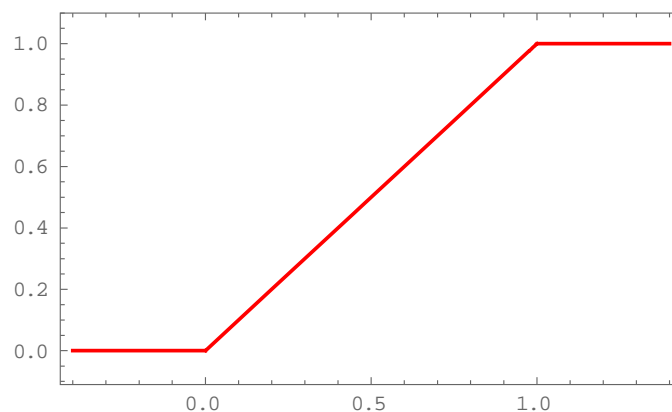


Figura 3.2.1. Funzione di ripartizione relativa alla legge Uniforme.

• **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. \square

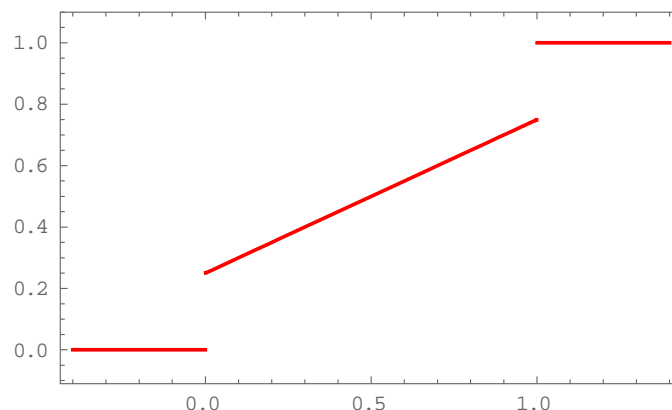


Figura 3.2.2. Funzione di ripartizione relativa alla legge mista.

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

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

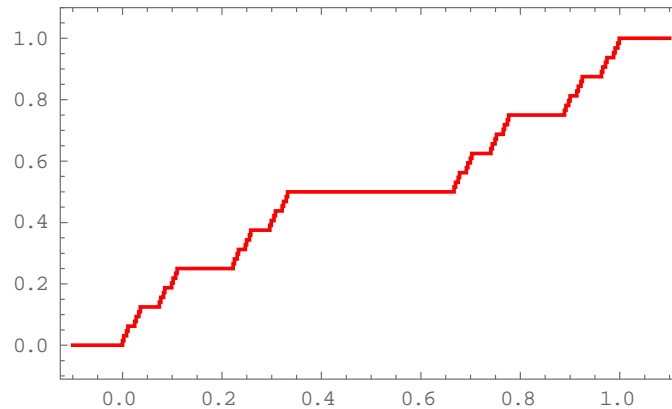
dove $c_n = 0, 1, 2$. Se x appartiene all'insieme di Cantor analizzato nell'Esempio 2.3.5 risulta $c_n = 0, 2$, ovvero x ha un'espansione ternaria che non contiene la cifra 1. Sia inoltre $N = \min\{n \in \mathbb{N} : c_n = 1\}$, mentre sia $N = \infty$ se non esiste tale n , e si ponga

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

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. \square



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

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 $f_X : \mathbb{R} \rightarrow [0, \infty[$ 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

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

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). \square

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 (Ω, \mathcal{F}, P) è uno spazio probabilizzato, un'applicazione $X : \Omega \rightarrow \mathbb{R}^n$ 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 $X = (X_1, \dots, X_n)^T$, dove X_1, \dots, 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 $B = B_1 \times \dots \times B_n$ con $B_k \in \mathcal{B}(\mathbb{R})$, si definisca inoltre l'evento

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

In particolare, si definisce

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

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

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

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

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

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

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 $x = (x_1, \dots, x_n)^T$ e $y = (y_1, \dots, y_n)^T$ sono tali che $x_k \leq y_k$ per ogni $k = 1, \dots, n$. Inoltre, risulta

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

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

$$\lim_{x_1 \rightarrow \infty, \dots, x_n \rightarrow \infty} F_X(x_1, \dots, x_n) = 1.$$

Infine, per $\varepsilon > 0$ si ha

$$\lim_{\varepsilon \rightarrow 0^+} F_X(x_1, \dots, x_k + \varepsilon, \dots, x_n) = F_X(x),$$

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

$$F_X(x) = \sum_{(u_1, \dots, u_n) \in S} \mathbf{1}_{]-\infty, x_1] \times \dots \times]-\infty, x_n]}(u_1, \dots, u_n) p_X(u_1, \dots, u_n).$$

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 $f_X : \mathbb{R}^n \rightarrow [0, \infty[$ 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 $X = (X_1, X_2)^T$ 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). \square

• **Esempio 3.5.2.** Si consideri il v.v.a. assolutamente continuo $X = (X_1, X_2)^T$ 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). \square

• **Esempio 3.5.3.** Si consideri il v.v.a. $X = (X_1, X_2)^T$ 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

$$F_{X_1}(x_1) = P(X_1 \leq x_1, X_2 \in \mathbb{R}) = x_1 \mathbf{1}_{[0,1]}(x_1) + \mathbf{1}_{]1,\infty[}(x_1),$$

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 $X_1 \stackrel{\mathcal{L}}{=} X_2$. Dunque, il fatto che le componenti marginali siano assolutamente continue non implica che il v.v.a. sia assolutamente continuo. \square

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 (Ω, \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 $X = (X_1, \dots, X_n)^T$, 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 (Ω, \mathcal{F}, P) sono indipendenti se e solo se la f.p.c. p_X del v.v.a. $X = (X_1, \dots, X_n)^T$ 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 (Ω, \mathcal{F}, P) sono indipendenti se e solo se la d.p.c. f_X del v.v.a. $X = (X_1, \dots, X_n)^T$ 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. $X = (X_1, X_2)^T$ 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. $X = (X_1, X_2)^T$ che ammette d.p.c.

$$f_X(x_1, x_2) = 4x_1 x_2 \mathbf{1}_{]0,1[}(x_1) \mathbf{1}_{]0,1[}(x_2).$$

La v.a. X_1 ammette d.p.m. data da

$$f_{X_1}(x_1) = \int_0^1 4x_1 x_2 \mathbf{1}_{]0,1[}(x_1) dx_2 = 2x_1 \mathbf{1}_{]0,1[}(x_1)$$

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). \square

• **Esempio 3.6.3.** Si consideri il v.v.a. assolutamente continuo $X = (X_1, X_2)^T$ 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

$$f_{X_1}(x_1) = \int_{-\sqrt{1-x_1^2}}^{\sqrt{1-x_1^2}} \frac{1}{\pi} \mathbf{1}_{]-1,1[}(x_2) dx_2 = \frac{2}{\pi} \sqrt{1-x_1^2} \mathbf{1}_{]-1,1[}(x_1)$$

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

$$f_X(x_1, x_2) \neq f_{X_1}(x_1)f_{X_2}(x_2). \quad \square$$

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 (Ω, \mathcal{F}, P) e la v.a. $X : \Omega \rightarrow \mathbb{R}$, se $g : \mathbb{R} \rightarrow \mathbb{R}$ 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 $g : \mathbb{R}^n \rightarrow \mathbb{R}^m$ 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

$$f_Y(y) = 2f_X(y) \mathbf{1}_{[0, \infty[}(y).$$

Nel caso particolare in cui

$$f_X(x) = \frac{1}{2} \mathbf{1}_{]-1, 1[}(x),$$

si ha dunque

$$f_Y(y) = \mathbf{1}_{]0, 1[}(y). \quad \square$$

• **Esempio 3.7.2.** Dato il v.v.a. $X = (X_1, X_2)^T$, 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. \square

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 (Ω, \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 . \square

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, \dots, 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

$$f_Y(y) = \frac{1}{2\sqrt{y}} \mathbf{1}_{]0, 1[}(y). \quad \square$$

Proposizione 3.7.2. Sia dato il v.v.a. X assolutamente continuo definito sullo spazio probabilizzato (Ω, \mathcal{F}, P) e sia $g : \mathbb{R}^n \rightarrow \mathbb{R}^n$ 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

$$P_Y(B) = P_X(g^{-1}(B)) = \int_{g^{-1}(B)} f_X(x_1, \dots, x_n) dx_1 \dots dx_n.$$

Tenendo presente il commento successivo al Teorema A.8, allora risulta

$$P_Y(B) = \int_B f_X(g^{-1}(y_1, \dots, y_n)) |J(g^{-1}(y_1, \dots, y_n))| dy_1 \dots dy_n,$$

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, \dots, 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. $X = (X_1, X_2)^T$ assolutamente continuo, si vuole determinare la d.p. della v.a. somma $(X_1 + X_2)$. A questo fine è conveniente considerare la trasformata $Y = (Y_1, Y_2)^T = g(X)$, dove $g(x) = (x_1, x_1 + x_2)^T$. Dal momento che $g^{-1}(y) = (y_1, y_2 - y_1)^T$, 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 $Y = (Y_1, Y_2)^T = g(X)$, dove $g(x) = (x_1, x_1 - x_2)^T$. 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

$$f_{Y_2}(y_2) = \int_{-\infty}^{\infty} f_{X_1}(y_1) f_{X_2}(y_1 + y_2) dy_1. \quad \square$$

• **Esempio 3.7.5.** Dato il v.v.a. assolutamente continuo $X = (X_1, X_2)^T$, si vuole determinare la d.p. della v.a. prodotto $X_1 X_2$. A questo fine è conveniente considerare la trasformata $Y = (Y_1, Y_2)^T = g(X)$, dove $g(x) = (x_1, x_1 x_2)^T$. Dal momento che si ha $g^{-1}(y) = (y_1, \frac{y_2}{y_1})^T$, 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

$$f_{Y_2}(y_2) = \int_0^1 \mathbf{1}_{]0,1[}(y_1 y_2) y_1 dy_1 = \frac{1}{2} \mathbf{1}_{]0,1[}(y_2) + \frac{1}{2y_2^2} \mathbf{1}_{]1,\infty[}(y_2). \quad \square$$

• **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 $Y = (Y_1, Y_2)^T = g(X)$, dove $g^{-1}(y) = (y_1 \cos(y_2), y_1 \sin(y_2))^T$. 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

$$f_{Y_1}(y_1) = 2y_1 \mathbf{1}_{]0,1[}(y_1),$$

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. $X = (X_1, \dots, X_n)^\top$ con componenti indipendenti definito sullo spazio probabilizzato (Ω, \mathcal{F}, P) . Se $Y = (Y_1, \dots, Y_n)^\top = g(X)$ 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 $k = 1, \dots, n$, 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 (2012), 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 (Ω, \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 $P(E)$. \square

• **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

$$\mathbf{1}_{E_1 \cup E_2} = (\mathbf{1}_{E_1 \setminus E_2} - \mathbf{1}_{E_1 \cap E_2}) + (\mathbf{1}_{E_2 \setminus E_1} + \mathbf{1}_{E_1 \cap E_2}) - \mathbf{1}_{E_1 \cap E_2} = \mathbf{1}_{E_1} + \mathbf{1}_{E_2} - \mathbf{1}_{E_1} \mathbf{1}_{E_2} . \quad \square$$

• **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. \square

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

$$\begin{aligned} F_X(x) &= P(X \leq x) = P(-X \geq -x) = P(X \geq -x) \\ &= 1 - P(X < -x) = 1 - P(X \leq -x) + P(X = -x) \\ &= 1 - P(X \leq -x) + P(-X = x) = 1 - F_X(-x) + P(X = x). \end{aligned}$$

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

$$f_X(x) = \frac{d}{dx} F_X(x) = \frac{d}{dx} (1 - F_X(-x)) = f_X(-x). \quad \square$$

• **Esercizio 3.2.2.** (Mistura di leggi) Si assuma che $w_k \in [0, 1]$ per $k = 1, \dots, n$ e $\sum_{k=1}^n w_k = 1$. Se F_{X_1}, \dots, 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 $0 \leq F_{X_k}(x) \leq 1$ per ogni $x \in \mathbb{R}$ e $k = 1, \dots, n$, si ha $F_X(x) \geq 0$ e $F_X(x) \leq \sum_{k=1}^n w_k = 1$. Inoltre, si osservi che risulta

$$\lim_{x \rightarrow -\infty} F_X(x) = \sum_{k=1}^n w_k \lim_{x \rightarrow -\infty} F_{X_k}(x) = 0$$

e

$$\lim_{x \rightarrow \infty} F_X(x) = \sum_{k=1}^n w_k \lim_{x \rightarrow \infty} F_{X_k}(x) = \sum_{k=1}^n w_k = 1,$$

essendo $\lim_{x \rightarrow -\infty} F_{X_k}(x) = 0$ e $\lim_{x \rightarrow \infty} F_{X_k}(x) = 1$ per ogni $k = 1, \dots, n$. Inoltre, se $x < y$ si ha $F_{X_k}(x) \leq F_{X_k}(y)$ e dunque

$$F_X(x) = \sum_{k=1}^n w_k F_{X_k}(x) \leq \sum_{k=1}^n w_k F_{X_k}(y) = F_X(y).$$

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 $x \in \{1, \dots, 9\}$ 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. \square

• **Esercizio 3.3.4.** Si consideri la v.a. discreta X con f.p. data da

$$p_X(x) = \binom{n+x}{n} 2^{-n-x} \mathbf{1}_{\{0, 1, \dots, n\}}(x),$$

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 $x \in \{0, 1, \dots, n\}$ 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. \square

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

$$\int \sqrt{r^2 - x^2} dx = \frac{1}{2} x \sqrt{r^2 - x^2} + \frac{r^2}{2} \arcsin\left(\frac{x}{r}\right),$$

si ha

$$\begin{aligned} F_X(x) &= \int_{-\infty}^x f_X(u) du = \frac{2}{\pi r^2} \mathbf{1}_{[-r,r]}(x) \int_{-r}^x \sqrt{r^2 - u^2} du + \mathbf{1}_{[r,\infty[}(x) \\ &= \left(\frac{1}{2} + \frac{1}{\pi r^2} x \sqrt{r^2 - x^2} + \frac{1}{\pi} \arcsin\left(\frac{x}{r}\right) \right) \mathbf{1}_{[-r,r]}(x) + \mathbf{1}_{[r,\infty[}(x). \end{aligned}$$

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. \square

• **Esercizio 3.4.2.** (Legge di Cauchy circolare) Si consideri la v.a. X con f.r. data da

$$F_X(x) = \frac{1}{2} + \frac{1}{\pi} \arctan\left(\frac{1+\rho}{1-\rho} \tan\left(\frac{x}{2}\right)\right) \mathbf{1}_{]-\pi,\pi[}(x) + \mathbf{1}_{[\pi,\infty[}(x),$$

dove $\rho \in]0, 1]$. Si verifichi che F_X è effettivamente una f.r.

Soluzione. Evidentemente, si ha $\lim_{x \rightarrow -\infty} F_X(x) = 0$ e $\lim_{x \rightarrow \infty} F_X(x) = 1$. 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. \square

Sezione 3.5

• **Esercizio 3.5.1.** Si consideri il v.v.a. discreto $X = (X_1, X_2)^T$ con f.p.c.

$$p_X(x_1, x_2) = \frac{4}{n^2(n+1)^2} x_1 x_2 \mathbf{1}_{\{1,\dots,n\}}(x_1) \mathbf{1}_{\{1,\dots,n\}}(x_2),$$

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

$$p_{X_1}(x_1) = \sum_{x_2=1}^n p_X(x_1, x_2) = \frac{4}{n^2(n+1)^2} x_1 \mathbf{1}_{\{1, \dots, n\}}(x_1) \sum_{x_2=1}^n x_2 = \frac{2}{n(n+1)} x_1 \mathbf{1}_{\{1, \dots, n\}}(x_1).$$

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

$$p_{X_2}(x_2) = \frac{2}{n(n+1)} x_2 \mathbf{1}_{\{1, \dots, n\}}(x_2). \quad \square$$

• **Esercizio 3.5.2.** (Legge di Farlie-Gumbel-Morgenstern) Si consideri il v.v.a. discreto $X = (X_1, X_2)^T$ 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. □

• **Esercizio 3.5.3.** Si consideri il v.v.a. assolutamente continuo $X = (X_1, X_2)^T$ 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

$$f_{X_2}(x_2) = e^{-x_2} \mathbf{1}_{[0, \infty[}(x_2). \quad \square$$

Sezione 3.6

• **Esercizio 3.6.1.** Si consideri il v.v.a. assolutamente continuo $X = (X_1, X_2)^T$ 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 $Y = (Y_1, Y_2)^T = g(X)$, ovvero $g^{-1}(y) = (y_1 \cos(y_2), y_1 \sin(y_2))^T$, 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. \square

• **Esercizio 3.6.2.** Se $X = (X_1, X_2, X_3)^T$ 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. $(X_1, X_2)^T$ 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. $(X_1, X_2)^T$ 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. $(X_1, X_3)^T$ e il v.v.a. $(X_2, X_3)^T$. 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. \square

Sezione 3.7

• **Esercizio 3.7.1.** (Legge Triangolare) Si consideri il v.v.a. $X = (X_1, X_2)^T$ 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 $X_1 - X_2 \stackrel{\mathcal{L}}{=} X_1 + X_2 - 1$. \square

• **Esercizio 3.7.2.** (Legge Triangolare, seconda parte) Si consideri il v.v.a. $X = (X_1, X_2)^T$ 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

$$\begin{aligned} f_Z(z) &= \frac{1}{4} \int_{-1}^{1+z} \mathbf{1}_{]-2,0[}(z) dx + \frac{1}{4} \int_{1-z}^1 \mathbf{1}_{]0,2[}(z) dx \\ &= \frac{2+z}{4} \mathbf{1}_{]-2,0[}(z) + \frac{2-z}{4} \mathbf{1}_{]0,2[}(z) = \frac{2-|z|}{4} \mathbf{1}_{]-2,2[}(z). \end{aligned}$$

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. \square

• **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

$$g(x) = g_1(x) + g_2(x),$$

dove $g_1(x) = \cos(2\pi x) \mathbf{1}_{]0, \frac{1}{2}[}(x)$ e $g_2(x) = \cos(2\pi(1-x)) \mathbf{1}_{] \frac{1}{2}, 1[}(x)$. 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

$$f_Y(y) = \frac{1}{\pi \sqrt{1-y^2}} \mathbf{1}_{] -1, 1[}(y). \quad \square$$

• **Esercizio 3.7.4.** Dato il v.v.a. assolutamente continuo $X = (X_1, X_2, X_3)^T$ 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 $Y = (Y_1, Y_2, Y_3)^T = g(X)$, con

$$g^{-1}(y) = (y_1 \sin(y_2) \cos(y_3), y_1 \sin(y_2) \sin(y_3), y_1 \cos(y_2))^T$$

e dove $y_1 \in]0, \infty[$, $y_2 \in [0, \pi[$ e $y_3 \in [0, 2\pi[$.

Soluzione. Dal momento che $|J(g^{-1}(y))| = y_1^2 \sin(y_2)$, la d.p.c. è data da

$$f_Y(y) = y_1^2 h(y_1^2) \sin(y_2) \mathbf{1}_{]0, \infty[}(y_1) \mathbf{1}_{]0, \pi[}(y_2) \mathbf{1}_{]0, 2\pi[}(y_3).$$

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

$$f_{Y_1}(y_1) = y_1^2 h(y_1^2) \mathbf{1}_{]0, \infty[}(y_1) \int_0^\pi \int_0^{2\pi} \sin(y_2) dy_2 dy_3 = 4\pi y_1^2 h(y_1^2) \mathbf{1}_{]0, \infty[}(y_1)$$

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

$$f_{Y_2}(y_2) = \sin(y_2) \mathbf{1}_{]0, \pi[}(y_2) \int_0^\infty \int_0^{2\pi} y_1^2 h(y_1^2) dy_1 dy_3 = \frac{1}{2} \sin(y_2) \mathbf{1}_{]0, \pi[}(y_2).$$

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 marginali del v.v.a. Y sono indipendenti. \square

• **Esercizio 3.7.5.** (Convoluzione di v.a. a valori interi) Si consideri il v.v.a. $X = (X_1, X_2)^T$ 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. \square

• **Esercizio 3.7.6.** Si consideri il v.v.a. discreto $X = (X_1, X_2)^T$ con f.p.c.

$$p_X(x_1, x_2) = 2^{-x_1 - x_2} \mathbf{1}_{\{1,2,\dots\}}(x_1) \mathbf{1}_{\{1,2,\dots\}}(x_2).$$

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 $n = 2, 3, \dots$ 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. \square

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 (Ω, \mathcal{F}, P) , si dice *valore atteso* l'integrale

$$E[X] = \int X dP$$

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

$$E[\mathbf{1}_E] = \int_E dP = P(E),$$

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

$$E[X] = \int_{\mathbb{R}} x dP_X(x).$$

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

$$E[X] = \int_{\mathbb{R}} x p_X(x) d\nu(x) = \sum_{x \in S} x p_X(x),$$

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

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

Infine, se X e Y sono v.a. tali che $X \stackrel{\mathcal{L}}{=} Y$, allora si ottiene immediatamente che $E[X] = E[Y]$.

• **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

$$E[X] = \sum_{x=0}^{\infty} x p_X(x) = \sum_{x=0}^{\infty} x 2^{-x-1}.$$

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

$$E[X] = \frac{1}{2} \sum_{x=0}^{\infty} x 2^{-x} = 1. \quad \square$$

• **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

$$E[X] = \int_0^1 x f_X(x) dx = 3 \int_0^1 x^3 dx = \frac{3}{4}. \quad \square$$

Denotando con $X^+ = \max(X, 0)$ e $X^- = -\min(X, 0)$, il valore atteso può essere scritto come

$$E[X] = E[X^+] - E[X^-].$$

Dunque, il valore atteso esiste finito se si ha $E[X^+] < \infty$ e $E[X^-] < \infty$. Nel caso in cui uno dei valori attesi $E[X^+]$ o $E[X^-]$ non sia finito, allora il valore atteso $E[X]$ non è finito. Infine, se entrambi i valori attesi $E[X^+]$ e $E[X^-]$ non sono finiti, allora il valore atteso $E[X]$ 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

$$f_X(x) = \frac{1}{\sqrt{2\pi x^3}} e^{-\frac{1}{2x}} \mathbf{1}_{]0, \infty[}(x).$$

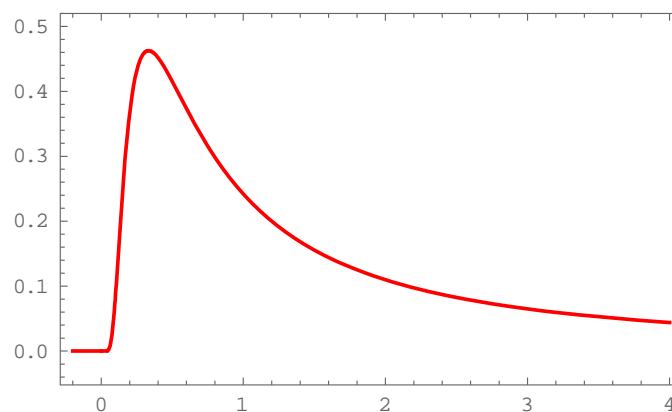


Figura 4.1.1. Densità di probabilità relativa alla legge ridotta di Lévy.

La legge associata a questa v.a. prende nome dal probabilmista 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 $E[X^-] = 0$, mentre

$$E[X^+] = \frac{1}{\sqrt{2\pi}} \int_0^\infty \frac{1}{\sqrt{x}} e^{-\frac{1}{2x}} dx \geq \frac{1}{\sqrt{2\pi}} \int_1^\infty \frac{1}{\sqrt{x}} e^{-\frac{1}{2x}} dx \geq \frac{e^{-\frac{1}{2}}}{\sqrt{2\pi}} \int_1^\infty \frac{1}{\sqrt{x}} dx = \infty$$

e quindi $E[X]$ non è finito. □

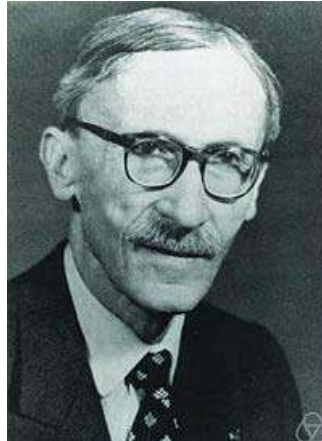


Figura 4.1.2. Paul Pierre Lévy (1886-1971).

• **Esempio 4.1.4.** (Legge di Cauchy) Si consideri la v.a. assolutamente continua X che ammette d.p. data da

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

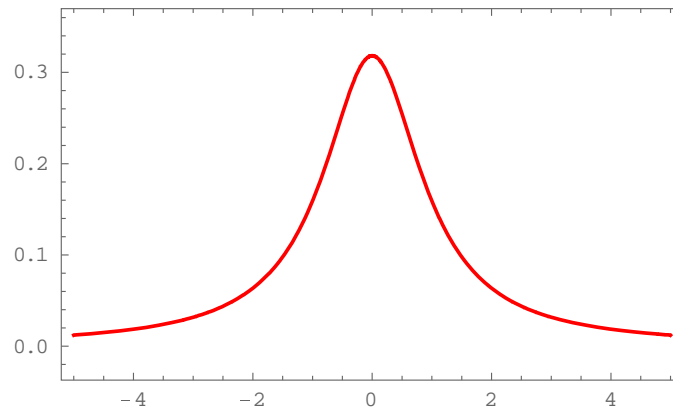


Figura 4.1.3. Densità di probabilità relativa alla legge ridotta di Cauchy.

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

$$E[X^+] = E[X^-] = \int_0^\infty \frac{x}{\pi(1+x^2)} dx = \infty$$

e quindi $E[X]$ 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 (Ω, \mathcal{F}, P) . Si ha*

$$|E[X]| \leq E[|X|]$$

e quindi se $E[|X|]$ esiste finito, allora anche $E[X]$ esiste finito.

Dimostrazione. Dal momento che $\{X \leq |X|\}$ q.c., allora dal Teorema A.6 si ha

$$E[X] \leq |E[X]| = \left| \int X dP \right| \leq \int |X| dP = E[|X|],$$

da cui segue immediatamente anche la seconda parte. □

Proposizione 4.2.2. *Si consideri la v.a. X definita sullo spazio probabilizzato (Ω, \mathcal{F}, P) . Se $P(X \in [a, b]) = 1$ dove a e b sono costanti, allora si ha $a \leq E[X] \leq b$. Quindi, se $a, b < \infty$, allora $E[X]$ esiste finito.*

Dimostrazione. Dal momento che $\{X \geq a\}$ q.c., dal Teorema A.6 si ha

$$E[X] = \int X dP \geq a \int dP = a,$$

mentre, dal momento che $\{X \leq b\}$ q.c., allora

$$E[X] = \int X dP \leq b \int dP = b,$$

da cui segue anche la seconda parte. □

Se $g: \mathbb{R}^n \rightarrow \mathbb{R}$ 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. $X = (X_1, \dots, X_n)^T$. Più esattamente, si ha che

$$|E[g(X)]| \leq E[|g(X)|]$$

e quindi se $E[|g(X)|]$ esiste finito, allora anche $E[g(X)]$ esiste finito. Inoltre, se $P(g(X) \in [a, b]) = 1$, si ha $a \leq E[g(X)] \leq b$. 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 $g : \mathbb{R}^n \rightarrow \mathbb{R}$, si consideri la trasformata $Y = g(X)$ del v.v.a. $X = (X_1, \dots, X_n)^T$ definito sullo spazio probabilitizzato (Ω, \mathcal{F}, P) . Si ha*

$$E[Y] = E[g(X)] = \int_{\mathbb{R}^n} g(x_1, \dots, x_n) dP_X(x_1, \dots, x_n)$$

se l'integrale esiste finito.

Dimostrazione. Dal Teorema A.8 si ha

$$E[Y] = \int g(X) dP = \int_{\mathbb{R}^n} g(x_1, \dots, x_n) dP_X(x_1, \dots, x_n)$$

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

$$E[Y] = E[g(X)] = \sum_{(x_1, \dots, x_n) \in S} g(x_1, \dots, x_n) p_X(x_1, \dots, x_n),$$

mentre se X è un v.v.a. assolutamente continuo che ammette d.p.c. f_X allora si ha

$$E[Y] = E[g(X)] = \int_{\mathbb{R}^n} g(x_1, \dots, x_n) f_X(x_1, \dots, x_n) dx_1 \dots dx_n.$$

• **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

$$E[Y] = \int_0^1 y f_Y(y) dy = \frac{1}{2} \int_0^1 \sqrt{y} dy = \frac{1}{3}.$$

Il medesimo risultato può essere ottenuto in modo immediato applicando direttamente la Proposizione 4.2.3, in quanto si ha

$$E[Y] = E[X^2] = \int_0^1 x^2 f_X(x) dx = \int_0^1 x^2 dx = \frac{1}{3}. \quad \square$$

• **Esempio 4.2.2.** Si consideri il v.v.a. assolutamente continuo $X = (X_1, X_2)^T$ 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

$$E[Y] = - \int_0^1 y \log(y) dy = \frac{1}{4}.$$

Tuttavia, applicando la Proposizione 4.2.3, si ha immediatamente

$$E[Y] = E[X_1 X_2] = \int_0^1 \int_0^1 x_1 x_2 dx_1 dx_2 = \frac{1}{4}. \quad \square$$

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 $X = (X_1, \dots, X_n)^T$ è un v.v.a. definito sullo spazio probabilizzato (Ω, \mathcal{F}, P) , si consideri la trasformata $Y = a + \sum_{k=1}^n b_k X_k$ dove $a \in \mathbb{R}$ e $(b_1, \dots, b_n)^T \in \mathbb{R}^n$ sono costanti. Se $E[X_k]$ esiste finito per ogni $k = 1, \dots, n$, si ha*

$$E[Y] = a + \sum_{k=1}^n b_k E[X_k].$$

Dimostrazione. Tenendo presente il Teorema A.6 si ha

$$E[Y] = \int \left(a + \sum_{k=1}^n b_k X_k \right) dP = a \int dP + b_k \sum_{k=1}^n \int X_k dP = a + \sum_{k=1}^n b_k E[X_k],$$

da cui segue la tesi. □

In particolare, quando si considera la trasformata $Y = a + bX$, dove $a, b \in \mathbb{R}$, dalla Proposizione 4.2.4 risulta

$$E[Y] = a + b E[X].$$

Inoltre, se si ha la trasformata $Y = X - \mu$, con $\mu = E[X]$, allora segue che

$$E[Y] = E[X] - \mu = 0.$$

Infine, se si considera la trasformata $Y = \sum_{k=1}^n X_k$, nelle ipotesi della Proposizione 4.2.4 risulta

$$E[Y] = \sum_{k=1}^n E[X_k].$$

• **Esempio 4.2.3.** Si consideri il v.v.a. assolutamente continuo $X = (X_1, X_2)^T$ introdotto nell'Esempio 3.5.2. Si ha

$$E[X_1] = 24 \int_0^1 \int_0^{1-x_1} x_1^2 (1 - x_1 - x_2) dx_1 dx_2 = \frac{2}{5}$$

e

$$E[X_2] = 24 \int_0^1 \int_0^{1-x_1} x_1 x_2 (1 - x_1 - x_2) dx_1 dx_2 = \frac{1}{5}.$$

Dunque, la media della v.a. $Y = X_1 + X_2$ è data da

$$E[Y] = E[X_1 + X_2] = E[X_1] + E[X_2] = \frac{3}{5}. \quad \square$$

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 $X = (X_1, \dots, X_n)^T$ è un v.v.a. definito sullo spazio probabilitizzato (Ω, \mathcal{F}, P) a componenti indipendenti, si consideri la trasformata $Y = \prod_{k=1}^n X_k$. Se $E[X_k]$ esiste finito per ogni $k = 1, \dots, n$, si ha*

$$E[Y] = \prod_{k=1}^n E[X_k].$$

Dimostrazione. Tenendo presente la Proposizione 4.2.3 e il Teorema di Fubini (Teorema A.11) si ha

$$E[Y] = \int \prod_{k=1}^n X_k dP = \int_{\mathbb{R}^n} \prod_{k=1}^n x_k d(P_{X_1} \otimes \dots \otimes P_{X_n}) = \prod_{k=1}^n \int_{\mathbb{R}} x_k dP_{X_k} = \prod_{k=1}^n E[X_k],$$

che è quanto si voleva dimostrare. □

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 $X = (X_1, X_2)^T$ 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 $E[X_1] = E[X_2] = \frac{1}{2}$ e quindi la media della v.a. $Y = X_1 X_2$ è data da

$$E[Y] = E[X_1 X_2] = E[X_1] E[X_2] = \frac{1}{4}. \quad \square$$

• **Esempio 4.2.5.** Si consideri il v.v.a. assolutamente continuo $X = (X_1, X_2)^T$ dell'Esempio 3.6.3. Risulta

$$E[X_1 X_2] = \frac{1}{\pi} \int_{-1}^1 \int_{-\sqrt{1-x_1^2}}^{\sqrt{1-x_1^2}} x_1 x_2 dx_1 dx_2 = 0,$$

mentre, data la simmetria rispetto all'origine della d.p. $f_{X_1}(x_1)$, si ha

$$E[X_1] = \frac{2}{\pi} \int_{-1}^1 x_1 \sqrt{1-x_1^2} dx_1 = 0.$$

Inoltre, dal momento che X_1 e X_2 posseggono la medesima legge, allora si ha anche $E[X_2] = 0$. Risulta quindi $E[X_1 X_2] = E[X_1] E[X_2]$ anche se le v.a. X_1 e X_2 non sono indipendenti. □

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 $g : \mathbb{R}^n \rightarrow \mathbb{R}$, si consideri la trasformata $Y = g(X)$ del v.v.a. $X = (X_1, \dots, X_n)^T$ definito sullo spazio probabilizzato (Ω, \mathcal{F}, P) . Si ha*

$$g(E[X]) \leq E[g(X)].$$

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 $v = E[X]$ e le proprietà del valore atteso, si ha

$$E[g(X)] \geq E[g(E[X]) + c(X - E[X])] = g(E[X]) + cE[X - E[X]] = g(E[X]),$$

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 $|E[X]| \leq E[|X|]$, 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 (Ω, \mathcal{F}, P) , si dice *momento di ordine r* , dove $r = 0, 1, \dots$, l'integrale

$$E[X^r] = \int X^r dP$$

se esiste finito. Si dice inoltre *momento assoluto di ordine r* l'integrale

$$E[|X|^r] = \int |X|^r dP$$

se esiste finito. Infine, si dicono rispettivamente *momento centrale di ordine r* l'integrale

$$E[(X - \mu)^r] = \int (X - \mu)^r dP$$

e *momento centrale assoluto di ordine r* l'integrale

$$E[|X - \mu|^r] = \int |X - \mu|^r dP$$

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 $\mu_1 = \mu = E[X]$. 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

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

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

$$E[(X - \mu)^r] = E\left[\sum_{k=0}^r \binom{r}{k} (-1)^{r-k} \mu^{r-k} X^k\right] = \sum_{k=0}^r \binom{r}{k} (-1)^{r-k} \mu^{r-k} \mu_k.$$

In particolare, si ha

$$\text{Var}[X] = E[X^2] - E[X]^2 = \mu_2 - \mu^2.$$

Nel linguaggio della Teoria della Misura, i momenti sono evidentemente funzioni di norme, ovvero $E[|X|^r] = \|X\|_r^r$ e $E[|X - \mu|^r] = \|X - \mu\|_r^r$.

• **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

$$E[X^r] = \int_0^1 x^r dx = \frac{1}{r+1}.$$

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

$$E[(X - \mu)^r] = \int_0^1 \left(x - \frac{1}{2}\right)^r dx = \frac{2^{-(r+1)}}{r+1} (1 + (-1)^r)$$

e quindi $E[(X - \mu)^r] = 0$ se r è dispari, mentre

$$E[(X - \mu)^r] = \frac{2^{-r}}{r+1}$$

se r è pari. Dunque, si ha $\text{Var}[X] = \frac{1}{12}$. Infine, risulta

$$E[|X - \mu|^r] = \int_0^1 \left|x - \frac{1}{2}\right|^r dx = \int_0^{\frac{1}{2}} \left(\frac{1}{2} - x\right)^r dx + \int_{\frac{1}{2}}^1 \left(x - \frac{1}{2}\right)^r dx = \frac{2^{-r}}{r+1}. \quad \square$$

• **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 $E[X^r] = E[Y^r]$ per ogni $r = 0, 1, \dots$ □

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 (Ω, \mathcal{F}, P) . Se $0 < s \leq r$, allora si ha*

$$E[|X|^s]^{1/s} \leq E[|X|^r]^{1/r}.$$

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 $E[Y]^a \leq E[Y^a]$. Dal momento che $\frac{r}{s} \geq 1$, ponendo $a = \frac{r}{s}$ e $Y = |X|^s$ nella precedente disuguaglianza, si ottiene

$$E[|X|^s]^{r/s} \leq E[|X|^r],$$

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 (Ω, \mathcal{F}, P) . Se si ha $E[|X|^r] < \infty$, allora $E[X^k]$ esiste finito per ogni $k = 1, \dots, r$.*

Dimostrazione. Dal momento che $k \leq r$, dalla disuguaglianza di Lyapunov (Teorema 4.3.2) si ha

$$E[|X|^k]^{1/k} \leq E[|X|^r]^{1/r}.$$

Dunque, se $E[|X|^r] < \infty$, allora si ha $E[|X|^k] < \infty$. Tuttavia, dalla Proposizione 4.2.1 risulta che se $E[|X|^k] < \infty$, allora si ha anche $E[X^k] < \infty$. \square

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 $\text{Var}[X] \leq E[X^2]$, 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 (Ω, \mathcal{F}, P) . Se $E[X^2]$ esiste finito, si ha*

$$\text{Var}[X] \leq \min_{a \in \mathbb{R}} E[(X - a)^2].$$

Dimostrazione. Tenendo presente le proprietà del valore atteso, si ha

$$\begin{aligned} E[(X - a)^2] &= E[(X - \mu + \mu - a)^2] = E[(X - \mu)^2 + 2(X - \mu)(\mu - a) + (\mu - a)^2] \\ &= \text{Var}[X] + (\mu - a)^2 \end{aligned}$$

e, dal momento che $\text{Var}[X] \geq 0$ non dipende da a , il minimo della precedente quantità si ottiene per $a = \mu$. \square

Proposizione 4.3.5. *Si consideri la v.a. X definita sullo spazio probabilizzato (Ω, \mathcal{F}, P) . Se $E[X^2]$ esiste finito, la varianza della trasformata $Y = a + bX$, dove $a, b \in \mathbb{R}$, è data da*

$$\text{Var}[Y] = \text{Var}[a + bX] = b^2 \text{Var}[X].$$

Dimostrazione. Tenendo presente la Proposizione 4.2.4 si ha $E[Y] = a + b\mu$, e quindi

$$\text{Var}[Y] = E[(a + bX - a - b\mu)^2] = b^2 E[(X - \mu)^2] = b^2 \text{Var}[X],$$

che è quanto si voleva dimostrare. \square

In particolare, se $Y = a + X$, dalla Proposizione 4.3.5 si ha

$$\text{Var}[Y] = \text{Var}[a + X] = \text{Var}[X].$$

Inoltre, se si considera la trasformazione

$$Z = \frac{X - \mu}{\sigma},$$

allora risulta

$$E[Z] = \frac{1}{\sigma} E[X - \mu] = 0,$$

mentre

$$\text{Var}[Z] = \frac{1}{\sigma^2} \text{Var}[X - \mu] = \frac{1}{\sigma^2} \text{Var}[X] = 1.$$

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 (Ω, \mathcal{F}, P) . Se $c > 0$, allora si ha*

$$P(|X| \geq c) \leq \frac{1}{c} E[|X|].$$

Dimostrazione. Si ha

$$E[|X|] = \int |X| dP \geq \int c \mathbf{1}_{\{|X| \geq c\}} dP = cP(|X| \geq c),$$

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 (Ω, \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

$$P(|X - \mu| \geq c) = P((X - \mu)^2 \geq c^2) \leq \frac{1}{c^2} E[(X - \mu)^2],$$

da cui segue la prima parte del Teorema. La seconda parte si ottiene in modo simile considerando la v.a. trasformata $\frac{(X - \mu)^2}{\sigma^2}$. □

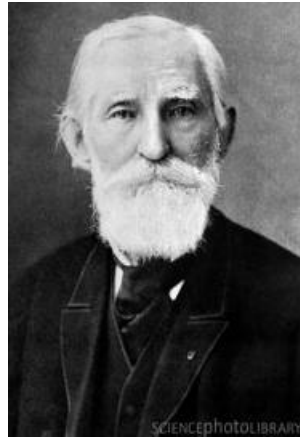


Figura 4.3.3. Pafnuty Lvovich Chebyshev (1821-1894).

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

$$P(|X - \mu| \geq c) \leq \frac{1}{c^r} E[|X - \mu|^r].$$

Proposizione 4.3.8. *Se X è una v.a. definita sullo spazio probabilitizzato (Ω, \mathcal{F}, P) e se $E[X^2]$ esiste finito, si ha $\text{Var}[X] = 0$ 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

$$E[X] = aP(X = a) = a,$$

mentre

$$\text{Var}[X] = (a - E[X])^2 P(X = a) = 0.$$

Inversamente, se $\text{Var}[X] = 0$, 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. $X = (X_1, \dots, X_n)^T$ definito sullo spazio probabilitizzato (Ω, \mathcal{F}, P) , si dice *momento misto di ordine* (r_1, \dots, r_n) , dove $r_k = 0, 1, \dots$ e $k = 1, \dots, n$, l'integrale

$$E[X_1^{r_1} \dots X_n^{r_n}] = \int X_1^{r_1} \dots X_n^{r_n} dP$$

se esiste finito. Inoltre, si dice *momento centrale misto di ordine* (r_1, \dots, r_n) l'integrale

$$E[(X_1 - \mu_{X_1})^{r_1} \dots (X_n - \mu_{X_n})^{r_n}] = \int (X_1 - \mu_{X_1})^{r_1} \dots (X_n - \mu_{X_n})^{r_n} dP$$

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 $E[X_k]$. In particolare, risulta fondamentale considerare i momenti per vettori bivariati di v.a. Più esattamente, dato il v.v.a. $X = (X_1, X_2)^T$, il *momento misto* è dato dall'integrale

$$E[X_1 X_2] = \int X_1 X_2 dP$$

se esiste finito. Analogamente, il *momento misto centrale*, detto usualmente *covarianza*, è dato dall'integrale

$$\text{Cov}[X_1, X_2] = E[(X_1 - \mu_{X_1})(X_2 - \mu_{X_2})] = \int (X_1 - \mu_{X_1})(X_2 - \mu_{X_2}) dP$$

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

$$\text{Cov}[X_1, X_2] = E[X_1 X_2] - E[X_1]E[X_2],$$

mentre ovviamente $\text{Cov}[X_1, X_1] = \text{Var}[X_1]$. In Teoria della Misura, il momento misto e la covarianza sono prodotti interni, ovvero $E[X_1 X_2] = \langle X_1, X_2 \rangle$ e $\text{Cov}[X_1 X_2] = \langle X_1 - \mu_{X_1}, X_2 - \mu_{X_2} \rangle$.

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. $X = (X_1, X_2)^T$ definito sullo spazio probabilizzato (Ω, \mathcal{F}, P) . Se $r, s \in]1, \infty[$ sono costanti tali che $\frac{1}{r} + \frac{1}{s} = 1$ e se $E[|X_1|^r] < \infty$ e $E[|X_2|^s] < \infty$, allora*

$$|E[X_1 X_2]| \leq E[|X_1 X_2|] \leq E[|X_1|^r]^{1/r} E[|X_2|^s]^{1/s}.$$

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

$$Y_1 = \frac{|X_1|^r}{E[|X_1|^r]}$$

e

$$Y_2 = \frac{|X_2|^s}{E[|X_2|^s]},$$

tenendo presente che $E[Y_1] = E[Y_2] = 1$, allora si ha

$$E[Y_1^{1/r} Y_2^{1/s}] \leq \frac{1}{r} E[Y_1] + \frac{1}{s} E[Y_2] = \frac{1}{r} + \frac{1}{s} = 1.$$

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

$$\frac{|X_1|^r}{E[|X_1|^r]} = \frac{|b_1 X_1|^r}{E[|b_1 X_1|^r]}$$

e

$$\frac{|X_2|^s}{E[|X_2|^s]} = \frac{|b_2 X_2|^s}{E[|b_2 X_2|^s]},$$

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

$$E[X_1 X_2]^2 \leq E[X_1^2] E[X_2^2],$$

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 $\text{Cov}[X_1, X_2]$ esiste finita se esistono finiti i momenti $E[X_1^2]$ e $E[X_2^2]$. 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. $X = (X_1, X_2)^T$ definito sullo spazio probabilizzato (Ω, \mathcal{F}, P) . Se $r \geq 1$ e se $E[|X_1|^r] < \infty$ e $E[|X_2|^r] < \infty$, allora*

$$E[|X_1 + X_2|^r]^{1/r} \leq E[|X_1|^r]^{1/r} + E[|X_2|^r]^{1/r}.$$

Dimostrazione. Se $E[|X_1 + X_2|^r] = 0$ la disuguaglianza è immediatamente verificata. Si consideri quindi il caso in cui $E[|X_1 + X_2|^r] > 0$. Tenendo presente che per ogni $u, v \in \mathbb{R}$ vale la disuguaglianza $|u + v| \leq |u| + |v|$, si ha

$$\begin{aligned} E[|X_1 + X_2|^r] &= E[|X_1 + X_2| \cdot |X_1 + X_2|^{r-1}] \leq E[(|X_1| + |X_2|) \cdot |X_1 + X_2|^{r-1}] \\ &= E[|X_1| \cdot |X_1 + X_2|^{r-1}] + E[|X_2| \cdot |X_1 + X_2|^{r-1}]. \end{aligned}$$

Inoltre, dalla disuguaglianza di Hölder applicata alle v.a. $|X_1|$ e $|X_1 + X_2|^{r-1}$ risulta

$$E[|X_1| \cdot |X_1 + X_2|^{r-1}] \leq E[|X_1|^r]^{1/r} E[|X_1 + X_2|^{(r-1)s}]^{1/s} = E[|X_1|^r]^{1/r} E[|X_1 + X_2|^r]^{1-1/r},$$

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

$$E[|X_2| \cdot |X_1 + X_2|^{r-1}] \leq E[|X_2|^r]^{1/r} E[|X_1 + X_2|^r]^{1-1/r}.$$

Dunque, si ha

$$E[|X_1 + X_2|^r] \leq (E[|X_1|^r]^{1/r} + E[|X_2|^r]^{1/r}) E[|X_1 + X_2|^r]^{1-1/r},$$

da cui discende immediatamente la tesi. \square

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. $X = (X_1, X_2)^T$ definito sullo spazio probabilitizzato (Ω, \mathcal{F}, P) , se X_1 e X_2 sono indipendenti e se $E[X_1]$ e $E[X_2]$ esistono finiti, allora si ha $\text{Cov}[X_1, X_2] = 0$.*

Dimostrazione. Tenendo presente la Proposizione 4.2.5 si ha

$$E[X_1 X_2] = E[X_1] E[X_2],$$

da cui segue immediatamente la tesi. \square

Si deve notare che non è in generale valido l'inverso della Proposizione 4.4.4, ovvero se si verifica $\text{Cov}[X_1, X_2] = 0$ non si può concludere che le due componenti del v.v.a. $X = (X_1, X_2)^T$ sono indipendenti. In effetti, come sarà visto in seguito, la covarianza dipende dall'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 $X = (X_1, X_2)^T$. 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

$$E[X_1^r] = \frac{1}{2} \int_{-1}^1 x^r dx = \frac{1}{2(r+1)} (1 + (-1)^r)$$

e quindi $E[X_1^r] = 0$ se r è dispari, mentre

$$E[X_1^r] = \frac{1}{r+1}$$

se r è pari. Dunque, si ha

$$\text{Cov}[X_1, X_2] = E[X_1^3] - E[X_1] E[X_1^2] = 0$$

anche se X_2 è una trasformata di X_1 e dunque X_1 e X_2 non risultano ovviamente indipendenti. \square

Proposizione 4.4.5. *Dato un v.v.a. $X = (X_1, X_2)^T$ definito sullo spazio probabilitizzato (Ω, \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 $\text{Cov}[X_1, X_2]$ esiste finita, allora*

$$\text{Cov}[Y_1, Y_2] = \text{Cov}[a_1 + b_1 X_1, a_2 + b_2 X_2] = b_1 b_2 \text{Cov}[X_1, X_2].$$

Dimostrazione. Tenendo presente la Proposizione 4.2.4 si ha

$$\begin{aligned} \text{Cov}[Y_1, Y_2] &= E[(a_1 + b_1 X_1 - E[a_1 + b_1 X_1])(a_2 + b_2 X_2 - E[a_2 + b_2 X_2])] \\ &= b_1 b_2 E[(X_1 - E[X_1])(X_2 - E[X_2])] = b_1 b_2 \text{Cov}[X_1, X_2], \end{aligned}$$

che è quanto si voleva dimostrare. \square

In particolare, per $Y_1 = a_1 + X_1$ e $Y_2 = a_2 + X_2$, dalla Proposizione 4.4.5 si ha

$$\text{Cov}[Y_1, Y_2] = \text{Cov}[a_1 + X_1, a_2 + X_2] = \text{Cov}[X_1, X_2].$$

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. $X = (X_1, X_2)^T$ definito sullo spazio probabilizzato (Ω, \mathcal{F}, P) , se X_1 e X_2 sono v.a. non degeneri tali che $E[X_1^2]$ e $E[X_2^2]$ 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

$$E[Z_1 Z_2] = \sigma_{Z_1, Z_2} = \rho_{Z_1, Z_2} = \rho_{X_1, X_2},$$

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. $X = (X_1, X_2)^T$ definito sullo spazio probabilizzato (Ω, \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 $E[X_1^2]$ e $E[X_2^2]$ esistono finiti, allora

$$\rho_{Y_1, Y_2} = \text{sgn}(b_1 b_2) \rho_{X_1, X_2}.$$

Dimostrazione. Tenendo presente la Proposizione 4.4.5 e la Proposizione 4.3.5 si ha

$$\rho_{Y_1, Y_2} = \frac{b_1 b_2 \sigma_{X_1, X_2}}{|b_1| \sigma_{X_1} |b_2| \sigma_{X_1}} = \text{sgn}(b_1 b_2) \rho_{X_1, X_2}. \quad \square$$

Proposizione 4.4.7. Dato un v.v.a. $X = (X_1, X_2)^T$ definito sullo spazio probabilizzato (Ω, \mathcal{F}, P) , se X_1 e X_2 sono v.a. non degeneri tali che $E[X_1^2]$ e $E[X_2^2]$ esistono finiti, si ha $|\rho_{X_1, X_2}| \leq 1$. Inoltre, $|\rho_{X_1, X_2}| = 1$ se e solo se $\{X_2 = a + b X_1\}$ q.c. o $\{X_2 = -a - b X_1\}$ 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 $E[Z_1^2] = \sigma_{Y_1}^2 = 1$ e $E[Z_2^2] = \sigma_{Y_2}^2 = 1$, dalla disuguaglianza di Schwarz si ha $E[Z_1 Z_2]^2 \leq 1$. La prima parte della Proposizione è verificata dal momento che $E[Z_1 Z_2] = \rho_{X_1, X_2}$. 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. \square

• **Esempio 4.4.2.** Si consideri il v.v.a. assolutamente continuo $X = (X_1, X_2)^T$ dell'Esempio 3.5.2. Si ha

$$E[X_1 X_2] = 24 \int_0^1 \int_0^{1-x_1} x_1^2 x_2 (1 - x_1 - x_2) dx_1 dx_2 = \frac{1}{15}.$$

Inoltre, è stato visto nell'Esempio 4.2.3 che $E[X_1] = \frac{2}{5}$ e $E[X_2] = \frac{1}{5}$, mentre risulta

$$E[X_1^2] = 24 \int_0^1 \int_0^{1-x_1} x_1^3 (1 - x_1 - x_2) dx_1 dx_2 = \frac{1}{5},$$

e

$$E[X_2^2] = 24 \int_0^1 \int_0^{1-x_1} x_1 x_2^2 (1 - x_1 - x_2) dx_1 dx_2 = \frac{1}{15},$$

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 $\rho_{X_1, X_2} = \frac{\sqrt{6}}{6}$. \square

In generale, considerato il v.v.a. $X = (X_1, \dots, X_n)^T$ definito sullo spazio probabilizzato (Ω, \mathcal{F}, P) , se $E[X_k]$ esiste finito per ogni $k = 1, \dots, n$, il *vettore delle medie* è definito come $\mu_X = (\mu_{X_1}, \dots, \mu_{X_n})^T$, mentre se $\text{Cov}[X_k, X_j]$ esiste finito per ogni $k = j = 1, \dots, n$, la *matrice di varianza-covarianza* è definita come

$$\Sigma_X = \begin{pmatrix} \sigma_{X_1}^2 & \sigma_{X_1, X_2} & \dots & \sigma_{X_1, X_n} \\ \sigma_{X_2, X_1} & \sigma_{X_2}^2 & \dots & \sigma_{X_2, X_n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{X_n, X_1} & \sigma_{X_n, X_2} & \dots & \sigma_{X_n}^2 \end{pmatrix}.$$

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 $X = (X_1, X_2, X_3)^T$ 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

$$E[X_1 X_2] = \int_0^1 \int_0^1 \int_0^1 x_1 x_2 (1 + (1 - 2x_1)(1 - 2x_2)) dx_1 dx_2 dx_3 = \frac{5}{18},$$

mentre

$$E[X_1] = \int_0^1 \int_0^1 \int_0^1 x_1 (1 + (1 - 2x_1)(1 - 2x_2)) dx_1 dx_2 dx_3 = \frac{1}{2}$$

e

$$E[X_1^2] = \int_0^1 \int_0^1 \int_0^1 x_1^2 (1 + (1 - 2x_1)(1 - 2x_2)) dx_1 dx_2 dx_3 = \frac{1}{3}.$$

Inoltre, per simmetria si ha $E[X_2] = E[X_1]$ e $E[X_2^2] = E[X_1^2]$. 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. $(X_1, X_2)^T$ e la v.a. X_3 sono indipendenti. Inoltre, si ha $E[X_3] = \frac{1}{2}$ e $E[X_3^2] = \frac{1}{3}$. Dunque, si ha infine che $\mu = (\frac{1}{2}, \frac{1}{2}, \frac{1}{2})^T$, mentre

$$\Sigma = \begin{pmatrix} \sigma_{X_1}^2 & \sigma_{X_1, X_2} & \sigma_{X_1, X_3} \\ \sigma_{X_2, X_1} & \sigma_{X_2}^2 & \sigma_{X_2, X_3} \\ \sigma_{X_3, X_1} & \sigma_{X_3, X_2} & \sigma_{X_3}^2 \end{pmatrix} = \begin{pmatrix} \frac{1}{12} & \frac{1}{36} & 0 \\ \frac{1}{36} & \frac{1}{12} & 0 \\ 0 & 0 & \frac{1}{12} \end{pmatrix}. \quad \square$$

La seguente Proposizione fornisce un importante risultato sulla varianza di una combinazione lineare di v.a.

Proposizione 4.4.8. *Se $X = (X_1, \dots, X_n)^T$ è un v.v.a. definito sullo spazio probabilizzato (Ω, \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 $b = (b_1, \dots, b_n)^T \in \mathbb{R}^n$ sono costanti. Se $\text{Cov}[X_k, X_j]$ esiste finita per ogni $k = 1, \dots, n$, si ha*

$$\text{Var}[Y] = b^T \Sigma b.$$

Dimostrazione. Tenendo presente il Teorema A.6 e la Proposizione 4.2.4, risulta

$$\begin{aligned} \text{Var}[Y] &= E[(a + \sum_{k=1}^n b_k X_k - a - \sum_{k=1}^n b_k E[X_k])^2] = E[(\sum_{k=1}^n b_k (X_k - E[X_k]))^2] \\ &= E[\sum_{k=1}^n b_k^2 (X_k - E[X_k])^2 + \sum_{k=1}^n \sum_{j \neq k=1}^n b_k b_j (X_k - E[X_k])(X_j - E[X_j])] \\ &= \sum_{k=1}^n b_k^2 \text{Var}[X_k] + \sum_{k=1}^n \sum_{j \neq k=1}^n b_k b_j \text{Cov}[X_k, X_j] = b^T \Sigma b, \end{aligned}$$

che è quanto si voleva dimostrare. □

Dalla Proposizione 4.4.8 si deduce che la matrice di varianza-covarianza è semidefinita positiva. In effetti, dal momento che $b^T \Sigma b$ è una varianza di una v.a., per ogni b diverso dal vettore a componenti nulle si ha $b^T \Sigma b \geq 0$, 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

$$\text{Var}[Y] = \sum_{k=1}^n \sigma_{X_k}^2 + \sum_{k=1}^n \sum_{j \neq k=1}^n \sigma_{X_k, X_j}$$

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 consentono 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 $\text{Var}[X_1 - X_2] = \frac{7}{75}$. \square

Un v.v.a. $X = (X_1, \dots, X_n)^T$ definito sullo spazio probabilizzato (Ω, \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

$$\text{Var}[Y] = b^T \Sigma b = 0,$$

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 (Ω, \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

$$E[X] = \int_0^{\infty} P(X \geq x) dx .$$

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

$$E[X] = \sum_{x=1}^{\infty} P(X \geq x) .$$

Soluzione. Dalla definizione di valore atteso si ha

$$E[X] = \int X dP = \int \int_0^{\infty} \mathbf{1}_{]-\infty, X]}(x) dx dP = \int \int_0^{\infty} \mathbf{1}_{[x, \infty[}(X) dx dP .$$

Sulla base del Teorema di Fubini l'ordine di integrazione può essere invertito e quindi

$$E[X] = \int_0^{\infty} \int \mathbf{1}_{[x, \infty[}(X) dP dx = \int_0^{\infty} P(X \geq x) dx .$$

Si osservi che, essendo $E[X] = E[X^+] - E[X^-]$ e applicando il precedente risultato rispettivamente a $E[X^+]$ e $E[X^-]$, per una v.a. X si ha in generale

$$E[X] = \int_0^{\infty} P(X \geq x) dx - \int_{-\infty}^0 P(X \leq x) dx .$$

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 $E[X]$ si ha

$$E[X] = \sum_{k=0}^{\infty} \int_k^{k+1} P(X \geq x) dx = \sum_{k=0}^{\infty} P(X \geq k + 1) = \sum_{x=1}^{\infty} P(X \geq x) . \quad \square$$

- **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

$$p_X(x) = \frac{1}{\zeta(s)} x^{-s} \mathbf{1}_{\{1,2,\dots\}}(x) ,$$

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 $E[X]$.

Soluzione. Si ha

$$E[X] = \sum_{x=1}^{\infty} x \frac{1}{\zeta(s)} x^{-s} = \frac{1}{\zeta(s)} \sum_{x=1}^{\infty} x^{-(s-1)} = \frac{\zeta(s-1)}{\zeta(s)} ,$$

e dunque $E[X]$ esiste finito solo se $s \in]2, \infty[$, mentre risulta $E[X] = \infty$ se $s \in]1, 2]$. □

• **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 $E[X]$.

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. \square

• **Esercizio 4.1.4.** (Analisi del sangue raggruppate) Un numero elevato di soggetti deve effettuare analisi del sangue per verificare la presenza di una patologia che si presenta su ogni individuo con una probabilità pari a p . Al fine di ridurre il costo delle analisi, i soggetti vengono suddivisi in gruppi di r individui. Successivamente, i campioni di sangue vengono raggruppati e analizzati con un unico test. Se l'analisi fornisce un risultato negativo per il gruppo, tutti i singoli campioni che lo compongono sono ovviamente negativi. In caso contrario, tutti i campioni del gruppo vengono di nuovo analizzati singolarmente. Si determini il valore atteso del numero analisi che si devono effettuare su ogni gruppo di r soggetti.

Soluzione. Sia X la v.a. discreta che descrive il numero di analisi che devono essere effettuate su ogni gruppo. La v.a. X è a valori sull'insieme $S = \{1, r+1\}$ e, assumendo che i campioni di sangue siano ottenuti in modo indipendente, la corrispondente f.p. è data da

$$p_X(x) = (1-p)^r \mathbf{1}_{\{1\}} + (1-(1-p)^r) \mathbf{1}_{\{r+1\}}.$$

Dunque, il valore atteso è dato da

$$E[X] = r+1 - r(1-p)^r.$$

Quindi, la strategia risulta efficiente, nel senso che $E[X] < r+1$. La procedura è stata adottata per determinare una patologia nei soldati statunitensi durante la Seconda Guerra Mondiale. \square

Sezione 4.2

• **Esercizio 4.2.1.** Si consideri una v.a. X simmetrica rispetto all'origine e la funzione misurabile $g: \mathbb{R} \rightarrow \mathbb{R}$. 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 $E[g(X)] = 0$.

Soluzione. Tenendo presente che $X \stackrel{\mathcal{L}}{=} -X$, si ha

$$E[g(X)] = E[g(-X)] = E[-g(X)] = -E[g(X)],$$

da cui si ottiene immediatamente la relazione $E[g(X)] = 0$. \square

• **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 $n = 1, 2, \dots$, 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

$$E[X] = \sum_{n=1}^{\infty} \log(x_n) p_n$$

e

$$E[e^X] = \sum_{n=1}^{\infty} x_n p_n .$$

Dal momento che la funzione esponenziale è convessa, dalla disuguaglianza di Jensen si ottiene

$$e^{E[X]} = \prod_{n=1}^{\infty} x_n^{p_n} \leq E[e^X] = \sum_{n=1}^{\infty} x_n p_n ,$$

ovvero si ha la nota disuguaglianza fra la media geometrica e la media aritmetica. \square

• **Esercizio 4.2.3.** Si consideri il v.v.a. $X = (X_1, X_2)^T$ 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 $E[\frac{1}{X_1}] < \infty$. Si verifichi che $E[\frac{X_1}{X_2}] \geq 1$ e $E[\frac{X_2}{X_1}] \geq 1$.

Soluzione. Dal momento che la funzione $g(x) = x^{-1}$ è convessa su \mathbb{R}^+ , dalla disuguaglianza di Jensen si ottiene

$$\frac{1}{E[X_2]} \leq E\left[\frac{1}{X_2}\right].$$

Dunque, dall'assunzione di indipendenza si ha

$$E\left[\frac{X_1}{X_2}\right] = E[X_1]E\left[\frac{1}{X_2}\right] \geq E[X_1] \frac{1}{E[X_2]} = 1 .$$

Inoltre, dal momento che $X_1 \stackrel{\mathcal{L}}{=} X_2$, per simmetria si ha anche $E[\frac{X_2}{X_1}] \geq 1$. 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. \square

• **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 $E[\lfloor X \rfloor]$, 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

$$\begin{aligned} E[\lfloor X \rfloor] &= \int_0^{\infty} \lfloor x \rfloor e^{-x} dx = \sum_{n=0}^{\infty} \int_n^{n+1} n e^{-x} dx = (1 - e^{-1}) \sum_{n=0}^{\infty} n e^{-n} \\ &= (1 - e^{-1}) \frac{e^{-1}}{(1 - e^{-1})^2} = \frac{1}{e - 1} . \end{aligned}$$

Si osservi che $\lfloor x \rfloor \leq x$ per $x \in \mathbb{R}$ e per le proprietà del valore atteso deve risultare $E[\lfloor X \rfloor] \leq E[X]$. In effetti, essendo $E[X] = 1$ si ha $\frac{1}{e-1} < 1$. \square

• **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

$$\begin{aligned} \mathbf{1}_{\bigcup_{k=1}^n E_k} &= 1 - \prod_{k=1}^n (1 - \mathbf{1}_{E_k}) = \sum_{k=1}^n (-1)^{k+1} \sum_{1 \leq j_1 < \dots < j_k \leq n} \mathbf{1}_{E_{j_1}} \cdots \mathbf{1}_{E_{j_k}} \\ &= \sum_{k=1}^n (-1)^{k+1} \sum_{1 \leq j_1 < \dots < j_k \leq n} \mathbf{1}_{E_{j_1} \cap \dots \cap E_{j_k}}, \end{aligned}$$

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

$$E[\mathbf{1}_{\bigcup_{k=1}^n E_k}] = \sum_{k=1}^n (-1)^{k+1} \sum_{1 \leq j_1 < \dots < j_k \leq n} E[\mathbf{1}_{E_{j_1} \cap \dots \cap E_{j_k}}].$$

Dal momento che $E[\mathbf{1}_E] = P(E)$, la Formula di Inclusione ed Esclusione risulta immediatamente dalla precedente relazione. \square

• **Esercizio 4.2.6.** Se X è v.a. discreta con f.p. data da $p_X(x) = 2^{-x} \mathbf{1}_{\{1,2,\dots\}}(x)$, si consideri la trasformata

$$Y = g(X) = (-1)^{X+1} \frac{2^X}{X}.$$

Si verifichi che $E[Y]$ 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

$$E[Y^+] = \sum_{k=1}^{\infty} (-1)^{2k} \frac{2^{2k-1}}{2k-1} p_X(2k-1) = \sum_{k=1}^{\infty} \frac{1}{2k-1} > \frac{1}{2} \sum_{k=1}^{\infty} \frac{1}{k} = \infty,$$

mentre per $x = 2k$ si ha

$$E[Y^-] = - \sum_{k=1}^{\infty} (-1)^{2k+1} \frac{2^{2k}}{2k} p_X(2k) = \frac{1}{2} \sum_{k=1}^{\infty} \frac{1}{k} = \infty,$$

ovvero $E[Y]$ non è definito. \square

Sezione 4.3

• **Esercizio 4.3.1.** (Disuguaglianza di Cantelli) Si consideri la v.a. X tale che $E[X] = \mu$ e $\text{Var}[X] = \sigma^2 < \infty$. 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 $E[Y] = 0$ e $\text{Var}[Y] = \sigma^2$. Tenendo presente la disuguaglianza di Markov, per $a > 0$ si ha

$$P(Y \geq c) = P(Y + a \geq c + a) \leq P((Y + a)^2 \geq (c + a)^2) \leq \frac{1}{(c + a)^2} E[(Y + a)^2] = \frac{\sigma^2 + a^2}{(c + a)^2}.$$

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

$$\psi_X(c) = \sup_{t \geq 0} (ct - \log(\mathbb{E}[e^{tX}])).$$

Soluzione. Tenendo presente la disuguaglianza di Markov, per $t \geq 0$ si ha

$$P(X \geq c) = P(e^{tX} \geq e^{tc}) \leq \frac{\mathbb{E}[e^{tX}]}{e^{tc}}.$$

Dalla precedente espressione risulta dunque

$$P(X \geq c) \leq \inf_{t \geq 0} \frac{\mathbb{E}[e^{tX}]}{e^{tc}} = e^{-\psi_X(c)}.$$

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

$$\mathbb{E}[e^{tX}] = \int_0^\infty e^{tx} e^{-x} dx = \frac{1}{1-t},$$

per cui $\psi_X(c) = c - 1 - \log(c)$ se $c > 1$, ovvero la disuguaglianza di Chernoff è data da

$$P(X \geq c) \leq ce^{-c+1}.$$

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 $E[X^2] < \infty$. Si verifichi che

$$P(X > 0) \geq \frac{E[X]^2}{E[X^2]}.$$

Soluzione. Considerando la v.a. $Y = \mathbf{1}_{]0, \infty[}(X)$, si ha

$$E[Y^2] = E[\mathbf{1}_{]0, \infty[}(X)^2] = E[\mathbf{1}_{]0, \infty[}(X)] = P(X > 0)$$

e

$$E[XY] = E[X \mathbf{1}_{]0, \infty[}(X)] = E[X \mathbf{1}_{]0, \infty[}(X)] = E[X].$$

Dunque, dalla disuguaglianza di Schwarz si ottiene

$$E[X^2]E[Y^2] = E[X^2]P(X > 0) \geq E[XY]^2 = E[X]^2,$$

da cui segue la disuguaglianza richiesta. \square

• **Esercizio 4.4.2.** (Disuguaglianza di Chebyshev bivariata) Si consideri il v.v.a. $X = (X_1, X_2)^T$. Assumendo che $\sigma_{X_1}^2, \sigma_{X_2}^2 < \infty$, si verifichi che per $c > 0$ risulta

$$P(\{|X_1 - \mu_{X_1}| \geq c\sigma_{X_1}\} \cup \{|X_2 - \mu_{X_2}| \geq c\sigma_{X_2}\}) \leq \frac{1}{c^2} (1 + \sqrt{1 - \rho_{X_1, X_2}^2}).$$

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

$$P(\{|Z_1| \geq c\} \cup \{|Z_2| \geq c\}) \leq \frac{1}{c^2} (1 + \sqrt{1 - \rho_{Z_1, Z_2}^2}).$$

Tenendo presente la disuguaglianza di Markov, si ha

$$\begin{aligned} P(\{|Z_1| \geq c\} \cup \{|Z_2| \geq c\}) &= P(\{Z_1^2 \geq c^2\} \cup \{Z_2^2 \geq c^2\}) \\ &= P(\max(Z_1^2, Z_2^2) \geq c^2) \leq \frac{1}{c^2} E[\max(Z_1^2, Z_2^2)]. \end{aligned}$$

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 $E[Z_1^2] = E[Z_2^2] = 1$ e $E[Z_1 Z_2] = \rho_{Z_1, Z_2}$, dalla disuguaglianza di Schwarz si ottiene

$$\begin{aligned} E[\max(Z_1^2, Z_2^2)] &= 1 + \frac{1}{2} E[|Z_1 + Z_2| \cdot |Z_1 - Z_2|] \leq 1 + \frac{1}{2} \sqrt{E[(Z_1 + Z_2)^2]E[(Z_1 - Z_2)^2]} \\ &= 1 + \sqrt{1 - \rho_{Z_1, Z_2}^2}, \end{aligned}$$

da cui segue immediatamente la disuguaglianza richiesta. \square

• **Esercizio 4.4.3.** Si consideri il v.v.a. assolutamente continuo $X = (X_1, X_2)^T$ 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

$$E[X_1] = \int_0^\infty \int_0^\infty x_1 ((1 + cx_1)(1 + cx_2) - c) e^{-x_1 - x_2 - cx_1x_2} dx_1 dx_2 = \int_0^\infty x_1 e^{-x_1} dx_1 = 1$$

e dunque per simmetria risulta anche $E[X_2] = 1$. Inoltre, risulta

$$E[X_1^2] = \int_0^\infty \int_0^\infty x_1^2 ((1 + cx_1)(1 + cx_2) - c) e^{-x_1 - x_2 - cx_1x_2} dx_1 dx_2 = \int_0^\infty x_1^2 e^{-x_1} dx_1 = 2$$

e quindi $E[X_2^2] = 2$. Dalle precedenti espressioni si ha $\text{Var}[X_1] = \text{Var}[X_2] = 1$. Inoltre, se $c \neq 0$ risulta

$$\begin{aligned} E[X_1 X_2] &= \int_0^\infty \int_0^\infty x_1 x_2 ((1 + cx_1)(1 + cx_2) - c) e^{-x_1 - x_2 - cx_1x_2} dx_1 dx_2 \\ &= \int_0^\infty \frac{x_1(1 + cx_1 + c)}{(1 + cx_1)^2} e^{-x_1} dx_1 = -\frac{e^{1/c}}{c} \text{Ei}\left(-\frac{1}{c}\right), \end{aligned}$$

dove

$$\text{Ei}(x) = \int_{-\infty}^x \frac{e^u}{u} du$$

rappresenta la funzione integrale esponenziale. In questo caso, si ha

$$\rho_{X_1, X_2} = -\frac{e^{1/c}}{c} \text{Ei}\left(-\frac{1}{c}\right) - 1.$$

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 $\rho_{X_1, X_2} \in [-e \text{Ei}(-1) - 1, 0]$. \square

• **Esercizio 4.4.4.** (Identità di Hoeffding) Dato il v.v.a. $X = (X_1, X_2)^T$, si verifichi che

$$\text{Cov}[X] = \int_{-\infty}^\infty \int_{-\infty}^\infty (F_X(x_1, x_2) - F_{X_1}(x_1)F_{X_2}(x_2)) dx_1 dx_2,$$

assumendo che $E[X_1] < \infty$ e $E[X_2] < \infty$.

Soluzione. Si consideri l'ulteriore v.v.a. $Y = (Y_1, Y_2)^T$ tale che X è indipendente da Y e $X \stackrel{\mathcal{L}}{=} Y$. Risulta immediato verificare che

$$E[(X_1 - Y_1)(X_2 - Y_2)] = 2 \text{Cov}[X].$$

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

$$\begin{aligned} E[(\mathbf{1}_{]-\infty, x_1]}(Y_1) - \mathbf{1}_{]-\infty, x_1]}(X_1))(\mathbf{1}_{]-\infty, x_2]}(Y_2) - \mathbf{1}_{]-\infty, x_2]}(X_2))] &= \\ &= 2 E[(\mathbf{1}_{]-\infty, x_1]}(X_1)\mathbf{1}_{]-\infty, x_2]}(X_2) - \mathbf{1}_{]-\infty, x_1]}(X_1)\mathbf{1}_{]-\infty, x_2]}(X_2))] \\ &= 2 (F_X(x_1, x_2) - F_{X_1}(x_1)F_{X_2}(x_2)). \end{aligned}$$

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). \square

• **Esercizio 4.4.5.** (Coefficienti di asimmetria e curtosi) Si osservi che il coefficiente di asimmetria di una v.a. X tale che $E[X^3] < \infty$ è definito come

$$\beta_{1,X} = \frac{E[(X - \mu_X)^3]}{\sigma_X^3},$$

mentre se $E[X^4] < \infty$ il coefficiente di curtosi è definito come

$$\beta_{2,X} = \frac{E[(X - \mu_X)^4]}{\sigma_X^4}.$$

Si consideri il v.v.a. $(X_1, \dots, X_n)^T$ 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

$$\begin{aligned} E[(Y - \mu_Y)^3] &= E[(\sum_{k=1}^n X_k - \sum_{k=1}^n \mu_X)^3] = E[(\sum_{k=1}^n (X_k - \mu_X))^3] \\ &= \sum_{j_1 + \dots + j_n = 3} \binom{3}{j_1 \dots j_n} \prod_{k=1}^n E[(X_k - \mu_X)^{j_k}]. \end{aligned}$$

Tenendo presente che gli addendi nella sommatoria che contengono i fattori $E[X_k - \mu_X] = 0$ si annullano, mentre $E[(X_k - \mu_X)^3] = E[(X - \mu_X)^3]$ per $k = 1, \dots, n$ dal momento che le componenti del v.v.a. X sono equivalenti in legge alla v.a. X , risulta

$$E[(Y - \mu_Y)^3] = \sum_{k=1}^n E[(X_k - \mu_X)^3] = nE[(X - \mu_X)^3],$$

da cui

$$\beta_{1,Y} = \frac{E[(Y - \mu_Y)^3]}{\sigma_Y^3} = \frac{nE[(X - \mu_X)^3]}{\sqrt{n^3}\sigma_X^3} = \frac{\beta_{1,X}}{\sqrt{n}}.$$

Per quanto riguarda il coefficiente di curtosi, si ha

$$\begin{aligned} E[(Y - \mu_Y)^4] &= E\left[\left(\sum_{k=1}^n X_k - \sum_{k=1}^n \mu_X\right)^4\right] = E\left[\left(\sum_{k=1}^n (X_k - \mu_X)\right)^4\right] \\ &= \sum_{j_1 + \dots + j_n = 4} \binom{4}{j_1 \dots j_n} \prod_{k=1}^n E[(X_k - \mu_X)^{j_k}]. \end{aligned}$$

Dal momento che

$$E[(X_k - \mu_X)^2(X_j - \mu_X)^2] = E[(X_k - \mu_X)^2]E[(X_j - \mu_X)^2] = \sigma_X^4$$

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

$$E[(Y - \mu_Y)^4] = nE[(X - \mu_X)^4] + 3n(n-1)\sigma_X^4.$$

Si deve dunque concludere che

$$\beta_{2,Y} = \frac{E[(Y - \mu_Y)^4]}{\sigma_Y^4} = \frac{nE[(X - \mu_X)^4] + 3n(n-1)\sigma_X^4}{n^2\sigma_X^4} = 3 + \frac{\beta_{2,X} - 3}{n}.$$

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). \square

Pagina intenzionalmente vuota

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 (Ω, \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

$$E[X | E_0] = \int X dP(\cdot | E_0) = \frac{1}{P(E_0)} \int_{E_0} X dP$$

se esiste finito. □

Si osservi che dalla Definizione 5.1.1 si ha in effetti

$$E[X | E_0] = \frac{E[X \mathbf{1}_{E_0}]}{P(E_0)}.$$

Se $X = \mathbf{1}_E$ dove $E \in \mathcal{F}$, allora risulta

$$E[\mathbf{1}_E | E_0] = \frac{E[\mathbf{1}_E \mathbf{1}_{E_0}]}{P(E_0)} = \frac{E[\mathbf{1}_{E \cap E_0}]}{P(E_0)} = \frac{P(E \cap E_0)}{P(E_0)} = P(E | E_0),$$

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 $E[X]$ esiste finito dal momento che $E[X \mathbf{1}_{E_0}] \leq E[X]$.

• **Esempio 5.1.1.** Si consideri la v.a. assolutamente continua X analizzata nell'Esempio 3.2.1 e si consideri l'evento $E_0 = X^{-1}(]0, \frac{1}{3}[)$. Dal momento che

$$P(E_0) = \int_0^{\frac{1}{3}} dx = \frac{1}{3}$$

e

$$E[X \mathbf{1}_{E_0}] = \int_0^{\frac{1}{3}} x dx = \frac{1}{18},$$

si ha

$$E[X | E_0] = \frac{1}{6}. \quad \square$$

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 (Ω, \mathcal{F}, P) . Sia inoltre $(E_n)_{n \geq 1}$ una partizione di Ω tale che $P(E_n) > 0$ per ogni $n = 1, 2, \dots$ e sia \mathcal{F}_0 la σ -algebra generata dalla partizione. Se $E[X]$ esiste finito, si dice *valore atteso condizionato* alla σ -algebra \mathcal{F}_0 la v.a.

$$E[X | \mathcal{F}_0] = \sum_{n=1}^{\infty} E[X | E_n] \mathbf{1}_{E_n}. \quad \square$$

Dalla precedente definizione, è immediato verificare che $E[X | \mathcal{F}_0]$ è una v.a. che assume valore $E[X | E_n]$ per ogni $\omega \in E_n$, dove $n = 1, 2, \dots$. Dunque, $E[X | \mathcal{F}_0]$ è una v.a. discreta che assume valori su $S = (E[X | E_n])_{n \geq 1}$ 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 $E[X | E_1] = \frac{1}{6}$ e $E[X | E_2] = \frac{2}{3}$. In questo caso, il valore atteso condizionato $E[X | \mathcal{F}_0]$ 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

$$\begin{aligned} \int_{E_0} E[X | \mathcal{F}_0] dP &= \sum_{n=1}^{\infty} E[X | E_n] \int_{E_0} \mathbf{1}_{E_n} dP = \sum_{n=1}^{\infty} E[X | E_n] P(E_n \cap E_0) \\ &= \sum_{n \in I} E[X | E_n] P(E_n) = \sum_{n \in I} E[X \mathbf{1}_{E_n}] = \sum_{n \in I} \int_{E_n} X dP = \int_{E_0} X dP, \end{aligned}$$

ovvero

$$E[E[X | \mathcal{F}_0] \mathbf{1}_{E_0}] = E[X \mathbf{1}_{E_0}].$$

Si osservi che nel caso in cui $E_0 = \Omega$, la precedente relazione stabilisce che il valore atteso della v.a. $E[X | \mathcal{F}_0]$ è pari al valore atteso della v.a. X , ovvero $E[E[X | \mathcal{F}_0]] = E[X]$. Inoltre, se si ha anche $X = \mathbf{1}_E$ e tenendo presente che

$$E[\mathbf{1}_E] = \int_{\Omega} \mathbf{1}_E dP = P(E),$$

si ottiene la generalizzazione della Legge delle Probabilità Totali (Teorema 2.4.3), ovvero

$$P(E) = \int_{\Omega} \mathbf{1}_E dP = \sum_{n=1}^{\infty} E[\mathbf{1}_E | E_n] P(E_n) = \sum_{n=1}^{\infty} P(E | E_n) P(E_n).$$

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 (Ω, \mathcal{F}, P) e sia $\mathcal{F}_0 \subset \mathcal{F}$ una σ -algebra. Se $E[X]$ 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

$$\int_{E_0} E[X | \mathcal{F}_0] dP = \int_{E_0} X dP . \quad \square$$

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. $X = (X_1, \dots, X_n)^T$ definito sullo spazio probabilizzato (Ω, \mathcal{F}, P) , il valore atteso condizionato alla σ -algebra \mathcal{F}_0 viene ovviamente definito mediante la relazione

$$\int_{E_0} E[g(X) | \mathcal{F}_0] dP = \int_{E_0} g(X) dP$$

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

$$P(E) = \int E[\mathbf{1}_E | \mathcal{F}_0] dP = E[E[\mathbf{1}_E | \mathcal{F}_0]] .$$

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 (Ω, \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 $\{E[X | \mathcal{F}_0] = X\}$ *q.c.*

Dimostrazione. Risulta immediata dalla relazione della Definizione 5.1.3. □

Proposizione 5.2.2. Se $X = (X_1, \dots, X_n)^T$ è un v.v.a. definito sullo spazio probabilizzato (Ω, \mathcal{F}, P) , si consideri la trasformata $Y = a + \sum_{k=1}^n b_k X_k$ dove $a \in \mathbb{R}$ e $(b_1, \dots, b_n)^T \in \mathbb{R}^n$ sono costanti. Se $\mathcal{F}_0 \subset \mathcal{F}$ è una σ -algebra e se $E[Y] < \infty$, si ha $\{E[Y | \mathcal{F}_0] = a + \sum_{k=1}^n b_k E[X_k | \mathcal{F}_0]\}$ *q.c.*

Dimostrazione. Tenendo presente la Definizione 5.1.3, per ogni $E_0 \in \mathcal{F}_0$ si ha

$$\begin{aligned} \int_{E_0} E[Y | \mathcal{F}_0] dP &= \int_{E_0} \left(a + \sum_{k=1}^n b_k X_k \right) dP = a \int_{E_0} dP + b_k \sum_{k=1}^n \int_{E_0} X_k dP \\ &= a \int_{E_0} dP + b_k \sum_{k=1}^n \int_{E_0} E[X_k | \mathcal{F}_0] dP \\ &= \int_{E_0} \left(a + \sum_{k=1}^n b_k E[X_k | \mathcal{F}_0] \right) dP , \end{aligned}$$

da cui segue la tesi. □

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 $g : \mathbb{R}^n \rightarrow \mathbb{R}$, si consideri la trasformata $Y = g(X)$ del v.v.a. $X = (X_1, \dots, X_n)^T$ definito sullo spazio probabilizzato (Ω, \mathcal{F}, P) e sia $\mathcal{F}_0 \subset \mathcal{F}$ una σ -algebra. Risulta $\{g(E[X | \mathcal{F}_0]) \leq E[g(X) | \mathcal{F}_0]\}$ q.c.*

ii) **(Disuguaglianza di Lyapunov condizionata)** *Si consideri la v.a. X definita sullo spazio probabilizzato (Ω, \mathcal{F}, P) e sia $\mathcal{F}_0 \subset \mathcal{F}$ una σ -algebra. Assumendo che $0 < s \leq r$, allora risulta $\{E[|X|^s | \mathcal{F}_0]^{1/s} \leq E[|X|^r | \mathcal{F}_0]^{1/r}\}$ q.c.*

iii) **(Disuguaglianza di Hölder condizionata)** *Si consideri il v.v.a. $X = (X_1, X_2)^T$ definito sullo spazio probabilizzato (Ω, \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 $\{|E[X_1 X_2 | \mathcal{F}_0]| \leq E[|X_1|^r | \mathcal{F}_0]^{1/r} E[|X_2|^s | \mathcal{F}_0]^{1/s}\}$ 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. \square

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 $\{|E[X | \mathcal{F}_0]| \leq E[|X| | \mathcal{F}_0]\}$ 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 (Ω, \mathcal{F}, P) e sia $\mathcal{F}_0 \subset \mathcal{F}$ una σ -algebra. Se $E[X]$ esiste finito, si ha*

$$E[X] = E[E[X | \mathcal{F}_0]],$$

mentre se $E[X^2]$ esiste finito, si ha

$$\text{Var}[X] = E[\text{Var}[X | \mathcal{F}_0]] + \text{Var}[E[X | \mathcal{F}_0]],$$

dove $\text{Var}[X | \mathcal{F}_0] = E[X^2 | \mathcal{F}_0] - E[X | \mathcal{F}_0]^2$.

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 $Y = X - E[X | \mathcal{F}_0]$ e $Z = E[X | \mathcal{F}_0] - E[X]$, si osservi che

$$\text{Var}[X] = E[(Y + Z)^2] = E[Y^2] + E[Z^2] + 2E[YZ].$$

Si può verificare che $\{E[YZ | \mathcal{F}_0] = ZE[Y | \mathcal{F}_0]\}$ 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 $\{E[Y | \mathcal{F}_0] = 0\}$ q.c. e di conseguenza si ha $\{E[YZ | \mathcal{F}_0] = 0\}$ q.c., ovvero dalla prima parte del Teorema si ottiene

$$E[YZ] = E[E[YZ | \mathcal{F}_0]] = 0.$$

Essendo $\{E[Y | \mathcal{F}_0] = 0\}$ q.c., risulta

$$E[Y^2] = E[E[Y^2 | \mathcal{F}_0]] = E[\text{Var}[Y | \mathcal{F}_0]] = E[\text{Var}[X | \mathcal{F}_0]],$$

dal momento che

$$\text{Var}[X | \mathcal{F}_0] = E[X^2 | \mathcal{F}_0] - E[X | \mathcal{F}_0]^2 = E[(X - E[X | \mathcal{F}_0])^2 | \mathcal{F}_0].$$

Infine, dalla prima parte del Teorema si ha

$$E[Z^2] = E[(E[X | \mathcal{F}_0] - E[X])^2] = E[(E[X | \mathcal{F}_0] - E[X])^2 | \mathcal{F}_0] = \text{Var}[E[X | \mathcal{F}_0]],$$

da cui segue la tesi. \square

Si osservi che dal Teorema 5.2.4 si ha

$$X = Y + E[X | \mathcal{F}_0],$$

con $Y = X - E[X | \mathcal{F}_0]$ e dove $E[YE[X | \mathcal{F}_0]] = 0$. In altri termini, la v.a. $E[X | \mathcal{F}_0]$ 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. $X = (X_1, X_2)^T$ definito sullo spazio probabilizzato (Ω, \mathcal{F}, P) e sia $\mathcal{F}_0 \subset \mathcal{F}$ una σ -algebra. Se $E[X_1^2]$ e $E[X_2^2]$ 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*

$$E[(X_2 - X_1)^2] = E[\text{Var}[X_2 | \mathcal{F}_0]] + E[(E[X_2 | \mathcal{F}_0] - X_1)^2].$$

Inoltre, si ha

$$E[(X_2 - X_1)^2] \geq E[\text{Var}[X_2 | \mathcal{F}_0]],$$

dove l'uguaglianza si ottiene se $\{X_1 = E[X_2 | \mathcal{F}_0]\}$ q.c.

Dimostrazione. Posto $Y = X_2 - E[X_2 | \mathcal{F}_0]$ e $Z = E[X_2 | \mathcal{F}_0] - X_1$, 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 $E[(E[X_2 | \mathcal{F}_0] - X_1)^2] \geq 0$ e l'uguaglianza è ottenuta quando $\{E[X_2 | \mathcal{F}_0] - X_1 = 0\}$ q.c. \square

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 (Ω, \mathcal{F}, P) e si consideri le σ -algre \mathcal{F}_0 e \mathcal{F}_1 tali che $\mathcal{F}_0 \subset \mathcal{F}_1 \subset \mathcal{F}$. Se $E[X]$ esiste finito, allora si ha $\{E[E[X | \mathcal{F}_1] | \mathcal{F}_0] = E[X | \mathcal{F}_0]\}$ q.c.*

Dimostrazione. Posto $Y = E[X | \mathcal{F}_1]$, tenendo presente la disuguaglianza di Jensen condizionata (Teorema 5.2.3, i), si ha

$$E[Y] \leq E[|Y|] = E[|E[X | \mathcal{F}_1]|] \leq E[E[|X| | \mathcal{F}_1]] = E[|X|] < \infty.$$

Dunque, per ogni $E_0 \in \mathcal{F}_0$ dalla Definizione 5.1.3 si ha

$$\int_{E_0} E[Y | \mathcal{F}_0] dP = \int_{E_0} Y dP.$$

Inoltre, di nuovo dalla Definizione 5.1.3 e tenendo presente che $E_0 \in \mathcal{F}_0 \subset \mathcal{F}_1$, si ha

$$\int_{E_0} E[X | \mathcal{F}_1] dP = \int_{E_0} X dP.$$

Dalle precedenti espressioni si ha

$$\int_{E_0} E[Y | \mathcal{F}_0] dP = \int_{E_0} X dP,$$

da cui segue la tesi. \square

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 (Ω, \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. $X = (X_1, X_2)^T$ definito sullo spazio probabilizzato (Ω, \mathcal{F}, P) , si può introdurre il *valore atteso condizionato* della v.a. X_2 (alla v.a. X_1) come

$$E[X_2 | X_1] = E[X_2 | \mathcal{F}_{X_1}].$$

Dunque, tenendo presente la Definizione 5.1.3, se $E[X_2]$ esiste finito, $E[X_2 | X_1]$ 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

$$\int_{X_1^{-1}(B)} E[X_2 | X_1] dP = \int_{X_1^{-1}(B)} X_2 dP.$$

Inoltre, dalla precedente relazione, si può verificare che esiste una funzione $h(x_1) = E[X_2 | X_1 = x_1]$, 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

$$\int_{X_1^{-1}(B)} X_2 dP = \int_{X_1^{-1}(B)} E[X_2 | X_1 = x_1] dP = \int_B h(x_1) dP_{X_1}(x_1).$$

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

$$\int_B h(x_1) dP_{X_1}(x_1) = E[X_2 | X_1 = x_1] p_{X_1}(x_1).$$

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

$$E[X_2 | X_1 = x_1] = \sum_{x_2 \in S_2} x_2 \frac{p_X(x_1, x_2)}{p_{X_1}(x_1)}.$$

• **Esempio 5.3.1.** Si consideri il v.v.a. discreto $X = (X_1, X_2)^T$ 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 $E[X_2] = \frac{1}{2} < \infty$, si ha

$$E[X_2 | X_1 = 0] = \sum_{x_2 \in S_2} x_2 \frac{p_X(0, x_2)}{p_{X_1}(0)} = \frac{2}{3}$$

e

$$E[X_2 | X_1 = 1] = \sum_{x_2 \in S_2} x_2 \frac{p_X(1, x_2)}{p_{X_1}(1)} = 0. \quad \square$$

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

$$E[X_2 | X_1 = x_1] = \int_{\mathbb{R}} x_2 \frac{f_X(x_1, x_2)}{f_{X_1}(x_1)} dx_2.$$

• **Esempio 5.3.2.** Si consideri il v.v.a. assolutamente continuo $X = (X_1, X_2)^T$ dell'Esempio 3.5.2. Dall'Esempio 4.2.3 risulta $E[X_2] = \frac{1}{5} < \infty$ e dunque, se $x_1 \in]0, 1[$, si ha

$$E[X_2 | X_1 = x_1] = \int_0^{1-x_1} \frac{2x_2}{1-x_1} \left(1 - \frac{x_2}{1-x_1} \right) dx_2 = \frac{1-x_1}{3},$$

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 $E[X_1] = \frac{2}{5} < \infty$, se $x_2 \in]0, 1[$, si ottiene

$$E[X_1 | X_2 = x_2] = \int_0^{1-x_2} \frac{6x_1^2}{(1-x_2)^2} \left(1 - \frac{x_1}{1-x_2} \right) dx_1 = \frac{1-x_2}{2},$$

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 . □

Nel caso in cui $B = \mathbb{R}$, dal Teorema 5.2.4 si ottiene

$$E[X_2] = E[E[X_2 | X_1]]$$

se $E[X_2]$ esiste finito. Inoltre, posto $\text{Var}[X_2 | X_1] = E[X_2^2 | X_1] - E[X_2 | X_1]^2$, di nuovo dal Teorema 5.2.4 si ha

$$\text{Var}[X_2] = E[\text{Var}[X_2 | X_1]] + \text{Var}[E[X_2 | X_1]],$$

se $E[X_2^2]$ esiste finito. Infine, dal Teorema 5.2.5 si ha

$$E[(X_2 - X_1)^2] = E[\text{Var}[X_2 | X_1]] + E[(E[X_2 | X_1] - X_1)^2]$$

e

$$E[(X_2 - X_1)^2] \geq E[\text{Var}[X_2 | X_1]],$$

dove l'uguaglianza si ottiene se $\{X_1 = E[X_2 | X_1]\}$ q.c.

• **Esempio 5.3.3.** Si consideri il v.v.a. assolutamente continuo $X = (X_1, X_2)^T$ dell'Esempio 3.5.2. Tenendo presente l'Esempio 5.3.2. e l'Esempio 4.2.3, si ha

$$E[X_2] = E[E[X_2 | X_1]] = \frac{1}{3} (1 - E[X_1]) = \frac{1}{5},$$

coerentemente con quanto visto nell'Esempio 4.2.3. In modo simile, si ottiene

$$E[X_1] = E[E[X_1 | X_2]] = \frac{1}{2} (1 - E[X_2]) = \frac{2}{5},$$

in accordo con quanto visto nell'Esempio 4.2.3. Inoltre, si ha

$$E[X_2^2 | X_1 = x_1] = \int_0^{1-x_1} \frac{2x_2^2}{1-x_1} \left(1 - \frac{x_2}{1-x_1}\right) dx_2 = \frac{(1-x_1)^2}{6},$$

da cui

$$\text{Var}[X_2 | X_1] = \frac{(1-X_1)^2}{6} - \frac{(1-X_1)^2}{9} = \frac{(1-X_1)^2}{18}.$$

Inoltre, tenendo presente l'Esempio 4.4.2, si ha

$$E[\text{Var}[X_2 | X_1]] = \frac{1}{18} E[(1-X_1)^2] = \frac{1}{18} (1 - 2E[X_1] + E[X_1^2]) = \frac{1}{45}$$

e

$$\text{Var}[E[X_2 | X_1]] = \frac{1}{9} \text{Var}[1 - X_1] = \frac{1}{9} \text{Var}[X_1] = \frac{1}{225},$$

da cui risulta

$$\text{Var}[X_2] = E[\text{Var}[X_2 | X_1]] + \text{Var}[E[X_2 | X_1]] = \frac{2}{75},$$

in accordo con i risultati ottenuti nell'Esempio 4.4.2. □

• **Esempio 5.3.4.** Si consideri il v.v.a. assolutamente continuo $X = (X_1, X_2)^T$ 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 $E[X_2]$ 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

$$E[X_2 | X_1 = x_1] = \int_{-\infty}^{\infty} x_2 \frac{f_X(x_1, x_2)}{f_{X_1}(x_1)} dx_2 = \int_{-\infty}^{\infty} \frac{x_2}{\sqrt{2\pi}|x_1|} e^{-\frac{1}{2x_1^2}(x_2-x_1)^2} dx_2 = x_1,$$

ovvero

$$E[X_2] = E[E[X_2 | X_1]] = E[X_1] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x_1 e^{-\frac{1}{2}(x_1-1)^2} dx_1 = 1,$$

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. \square

5.4. Leggi condizionate

Si consideri il v.v.a. $X = (X_1, X_2)^T$ definito sullo spazio probabilizzato (Ω, \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 $x_1 \in S_1$, $P(\cdot | X_1 = x_1)$ è in effetti una misura di probabilità (vedi Sezione 2.4) e per $B \in \mathcal{B}(\mathbb{R})$

$$P_{X_2|X_1=x_1}(B) = P(X_2^{-1}(B) | X_1 = x_1) = \frac{P(X_2^{-1}(B) \cap \{X_1 = x_1\})}{P(X_1 = x_1)}$$

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

$$p_{X_2|X_1=x_1}(x_2) = P(X_2 = x_2 | X_1 = x_1) = \frac{p_X(x_1, x_2)}{p_{X_1}(x_1)}.$$

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

$$E[X_2 | X_1 = x_1] = \sum_{x_2 \in S_2} x_2 p_{X_2|X_1=x_1}(x_2).$$

• **Esempio 5.4.1.** Si consideri il v.v.a. discreto $X = (X_1, X_2)^T$ 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

$$p_{X_2|X_1=1}(x_2) = \frac{p_X(1, x_2)}{p_{X_1}(1)} = \mathbf{1}_{\{0\}}(x_2),$$

ovvero le leggi associate alle v.a. condizionate si ottengono per particolari parametrizzazioni della legge Binomiale (si veda Sezione 6.5). \square

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 artificiali

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

$$E[X_2 | X_1 = x_1] = \int_{\mathbb{R}} x_2 f_{X_2|X_1=x_1}(x_2) dx_2.$$

• **Esempio 5.4.2.** Si consideri il v.v.a. assolutamente continuo $X = (X_1, X_2)^T$ 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). \square

• **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). \square

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 (2021) 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 (Ω, \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 (Ω, \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 $\mathcal{E} = \{B_1, \dots, B_n\}$, dove $B_k =]a_{k-1}, a_k]$ e $a_0 = 0 < a_1 < \dots < a_n = 1$, 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 $P(E | B_k) = E[\mathbf{1}_E | B_k]$, dove $k = 1, \dots, n$, e $P(E | \mathcal{F}_0) = E[\mathbf{1}_E | \mathcal{F}_0]$.

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

$$P(E | B_k) = \frac{1}{a_k - a_{k-1}} \int_{a_{k-1}}^{a_k} \mathbf{1}_E(x) dx = \frac{\lambda(E \cap B_k)}{a_k - a_{k-1}} .$$

Inoltre, tenendo presente la definizione di valore atteso condizionato alla σ -algebra \mathcal{F}_0 , $P(E | \mathcal{F}_0)$ è data da

$$P(E | \mathcal{F}_0) = \sum_{k=1}^n P(E | B_k) \mathbf{1}_{B_k} = \sum_{k=1}^n \frac{\lambda(E \cap B_k)}{a_k - a_{k-1}} \mathbf{1}_{B_k} .$$

Quindi, $P(E | \mathcal{F}_0)$ è una v.a. discreta che assume valori su $S = (P(E | B_k))_{k=1}^n$ con legge essenziale data dall'insieme delle probabilità $(P(B_k))_{k=1}^n$. Inoltre, essendo $E[\mathbf{1}_E] = P(B_k)$ si ha

$$\begin{aligned} E[P(E | \mathcal{F}_0)] &= \sum_{k=1}^n \frac{\lambda(E \cap B_k)}{a_k - a_{k-1}} E[\mathbf{1}_{B_k}] = \sum_{k=1}^n \lambda(E \cap B_k) \\ &= \lambda\left(E \cap \bigcup_{k=1}^n B_k\right) = \lambda(E \cap \Omega) = \lambda(E) = P(E) , \end{aligned}$$

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

$$P(\{X \leq x\} \cap E_0) = P(X \leq x)P(E_0)$$

per ogni $x \in \mathbb{R}$ e $E_0 \in \mathcal{F}_0$. Si verifichi che risulta $\{E[X | \mathcal{F}_0] = E[X]\}$ q.c.

Soluzione. Dalla definizione di valore atteso condizionato, per ogni $E_0 \in \mathcal{F}_0$ si ha

$$\int_{E_0} E[X | \mathcal{F}_0] dP = \int_{E_0} X dP = \int_{\Omega} X \mathbf{1}_{E_0} dP = E[X \mathbf{1}_{E_0}] .$$

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

$$E[X \mathbf{1}_{E_0}] = E[X]E[\mathbf{1}_{E_0}],$$

da cui

$$\int_{E_0} E[X | \mathcal{F}_0] dP = E[X] \int_{\Omega} \mathbf{1}_{E_0} dP = \int_{\Omega} E[X] \mathbf{1}_{E_0} dP = \int_{E_0} E[X] dP.$$

Dunque, dalla precedente relazione segue l'identità richiesta. \square

Sezione 5.2

• **Esercizio 5.2.1.** Sia dato lo spazio probabilizzato (Ω, \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 $X = (X_1, X_2)^T$ 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 $E[(X_2 - X_1)^2]$ e i valori attesi $E[\text{Var}[X_2 | \mathcal{F}_0]]$ e $E[(E[X_2 | \mathcal{F}_0] - X_1)^2]$. 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

$$E[(X_2 - X_1)^2] = \int_0^1 (\omega - \mathbf{1}_E(\omega))^2 d\omega = \int_0^1 \omega^2 d\omega - 2 \int_0^{\frac{1}{3}} \omega d\omega + \int_0^{\frac{1}{3}} d\omega = \frac{5}{9}.$$

Inoltre, essendo $P(E) = \frac{1}{3}$ e $P(E^c) = \frac{2}{3}$, segue

$$E[X_2 | E] = \frac{E[X_2 \mathbf{1}_E]}{P(E)} = 3 \int_0^{\frac{1}{3}} \omega d\omega = \frac{1}{6}$$

e

$$E[X_2 | E^c] = \frac{E[X_2 \mathbf{1}_{E^c}]}{P(E^c)} = \frac{3}{2} \int_{\frac{1}{3}}^1 \omega d\omega = \frac{2}{3},$$

mentre

$$E[X_2^2 | E] = \frac{E[X_2^2 \mathbf{1}_E]}{P(E)} = 3 \int_0^{\frac{1}{3}} \omega^2 d\omega = \frac{1}{27}$$

e

$$E[X_2^2 | E^c] = \frac{E[X_2^2 \mathbf{1}_{E^c}]}{P(E^c)} = \frac{3}{2} \int_{\frac{1}{3}}^1 \omega^2 d\omega = \frac{13}{27}.$$

Dal momento che \mathcal{F}_0 è una σ -algebra generata da una partizione, allora risulta

$$E[X_2 | \mathcal{F}_0] = E[X_2 | E] \mathbf{1}_E + E[X_2 | E^c] \mathbf{1}_{E^c} = \frac{1}{6} \mathbf{1}_E + \frac{2}{3} \mathbf{1}_{E^c}$$

e

$$E[X_2^2 | \mathcal{F}_0] = E[X_2^2 | E] \mathbf{1}_E + E[X_2^2 | E^c] \mathbf{1}_{E^c} = \frac{1}{27} \mathbf{1}_E + \frac{13}{27} \mathbf{1}_{E^c},$$

da cui, tenendo presente che $\mathbf{1}_E \mathbf{1}_{E^c} = 0$ per ogni $\omega \in \Omega$,

$$\begin{aligned}\text{Var}[X_2 | \mathcal{F}_0] &= \text{E}[X_2^2 | \mathcal{F}_0] - \text{E}[X_2 | \mathcal{F}_0]^2 \\ &= (\text{E}[X_2^2 | E] - \text{E}[X_2 | E]^2) \mathbf{1}_E + (\text{E}[X_2^2 | E^c] - \text{E}[X_2 | E^c]^2) \mathbf{1}_{E^c} \\ &= \frac{1}{108} \mathbf{1}_E + \frac{1}{27} \mathbf{1}_{E^c}.\end{aligned}$$

Quindi, si ha

$$\text{E}[\text{Var}[X_2 | \mathcal{F}_0]] = \text{E}\left[\frac{1}{108} \mathbf{1}_E + \frac{1}{27} \mathbf{1}_{E^c}\right] = \frac{1}{108} P(E) + \frac{1}{27} P(E^c) = \frac{1}{36}.$$

Inoltre, risulta

$$\begin{aligned}\text{E}[(\text{E}[X_2 | \mathcal{F}_0] - X_1)^2] &= \text{E}\left[\left(\frac{1}{6} \mathbf{1}_E + \frac{2}{3} \mathbf{1}_{E^c} - \mathbf{1}_E\right)^2\right] = \text{E}\left[\frac{25}{36} \mathbf{1}_E + \frac{4}{9} \mathbf{1}_{E^c}\right] \\ &= \frac{25}{36} P(E) + \frac{4}{9} P(E^c) = \frac{19}{36}\end{aligned}$$

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 $X = (X_1, X_2)^T$ 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 $\text{E}[X_2 | X_1 = 0] = \frac{1}{2}$. Si consideri inoltre il v.v.a. $Y = (Y_1, Y_2)^T$ dove $Y_1 = \frac{X_1 + X_2}{X_2}$ e $Y_2 = X_2$ e si verifichi che si ha $\text{E}[Y_2 | Y_1 = 1] = \frac{2}{3}$.

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

$$\begin{aligned}\text{E}[X_2 | X_1 = x_1] &= \int_{-\infty}^{\infty} x_2 \frac{\mathbf{1}_{[0,1]}(x_1 + x_2) \mathbf{1}_{[0,1]}(x_2)}{(1 - |x_1|) \mathbf{1}_{[-1,1]}(x_1)} dx_2 \\ &= \frac{\mathbf{1}_{[-1,0]}(x_1)}{1 + x_1} \int_{x_1}^1 x_2 dx_2 + \frac{\mathbf{1}_{[0,1]}(x_1)}{1 - x_1} \int_0^{1-x_1} x_2 dx_2 = \frac{1 - x_1}{2} \mathbf{1}_{[-1,1]}(x_1).\end{aligned}$$

Quindi, si ha $\text{E}[X_2 | X_1 = 0] = \frac{1}{2}$. Per quanto riguarda la trasformata $Y = g(X)$, si ha

$$g^{-1}(y) = (y_2(y_1 - 1), y_2)^T,$$

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

$$\begin{aligned} E[Y_2 | Y_1 = y_1] &= \int_{-\infty}^{\infty} y_2 \frac{y_2 \mathbf{1}_{[0,1]}(y_1 y_2) \mathbf{1}_{[0,1]}(y_2)}{\frac{1}{2y_1^2} \mathbf{1}_{[1,\infty]}(y_1)} dy_2 \\ &= 2y_1^2 \mathbf{1}_{[1,\infty]}(y_1) \int_0^1 y_2^2 dx_2 = \frac{2}{3} y_1^2 \mathbf{1}_{[1,\infty]}(y_1). \end{aligned}$$

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

$$E[Y_2 | Y_1 = 1] = \frac{2}{3} \neq E[X_2 | X_1 = 0] = \frac{1}{2},$$

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. \square

• **Esercizio 5.3.2.** Si consideri il v.v.a. $X = (X_1, \dots, X_n)^T$ le cui componenti sono indipendenti e con la medesima legge. Si assuma inoltre che $Y = \sum_{k=1}^n X_k$. Si verifichi che $\{E[X_k | Y] = n^{-1}Y\}$ q.c. per ogni $k = 1, \dots, n$.

Soluzione. Dalla definizione di valore atteso condizionato, per ogni $B \in \mathbb{R}$ e $k = 1, \dots, n$ si ha

$$\int_{Y^{-1}(B)} E[X_k | Y] dP = E[X_k \mathbf{1}_{Y^{-1}(B)}].$$

Dunque, dal momento che X_1 e X_k sono equivalenti in legge, risulta $E[X_k \mathbf{1}_{Y^{-1}(B)}] = E[X_1 \mathbf{1}_{Y^{-1}(B)}]$, ovvero tenendo presente la precedente relazione si ottiene

$$P(E[X_k | Y] = E[X_1 | Y]) = 1$$

per ogni $k = 2, \dots, n$. Dunque, si ha q.c.

$$E[X_k | Y] = \frac{1}{n} \sum_{j=1}^n E[X_j | Y] = \frac{1}{n} E\left[\sum_{j=1}^n X_j | Y\right] = \frac{1}{n} E[Y | Y] = \frac{1}{n} Y.$$

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 $E[X_k] = p$ e $\text{Var}[X_k] = pq$. 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 $E[Y] = np$ e $\text{Var}[Y] = npq$. Dunque, dalla precedente relazione si ha

$$E[E[X_k | Y]] = \frac{1}{n} E[Y] = p$$

e

$$\text{Var}[E[X_k | Y]] = \frac{1}{n^2} \text{Var}[Y] = \frac{pq}{n}.$$

Quindi, nella terminologia della statistica inferenziale, sia la v.a. X_k che la v.a. $E[X_k | Y]$ sono stimatori corretti del parametro p . Inoltre, posto $x = (x_1, \dots, x_n)^T$, per $y = 0, 1, \dots, n$ 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 $S_y = \{(x_1, \dots, x_n) : x_k = 0, 1, k = 1, \dots, n, \sum_{k=1}^n x_k = y\}$, 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, $\text{Var}[X_k] > \text{Var}[E[X_k | Y]]$ e quindi lo stimatore $E[X_k | Y]$ è più efficiente dello stimatore X_k . Questo risultato segue dal fatto che lo stimatore $E[X_k | Y]$ 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. \square

• **Esercizio 5.3.3.** Si consideri il v.v.a. assolutamente continuo $X = (X_1, X_2)^T$ 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 $E[X_2 | X_1 = x_1]$ e $\text{Var}[X_2 | X_1 = x_1]$.

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

$$E[X_2 | X_1 = x_1] = \int_{-\infty}^{\infty} x_2 \frac{2\sqrt{(1 + x_1^2)^3}}{\pi(1 + x_1^2 + x_2^2)^2} dx_2 = 0,$$

dal momento che la funzione integranda è dispari. Quindi, si ottiene

$$\text{Var}[X_2 | X_1 = x_1] = \int_{-\infty}^{\infty} x_2^2 \frac{2\sqrt{(1 + x_1^2)^3}}{\pi(1 + x_1^2 + x_2^2)^2} dx_2 = 1 + x_1^2.$$

Si osservi che

$$E[X_2] = \int_{-\infty}^{\infty} x_2 \frac{1}{2\sqrt{(1 + x_2^2)^3}} dx_2 = 0$$

e in effetti risulta $E[E[X_2 | X_1]] = 0$. Tuttavia, si ha

$$E[X_2^2] = \int_{-\infty}^{\infty} x_2^2 \frac{1}{2\sqrt{(1 + x_2^2)^3}} dx_2 = \infty$$

e quindi $\text{Var}[X_2]$ non è finita. In effetti, essendo $X_1 \stackrel{\mathcal{L}}{=} X_2$ per simmetria, si ha

$$E[\text{Var}[X_2 | X_1]] = E[1 + X_1^2] = 1 + E[X_1^2] = \infty.$$

Dunque, può essere $\text{Var}[X_2 | X_1 = x_1] < \infty$ per ogni x_1 , anche se $E[\text{Var}[X_2 | X_1]] = \infty$. \square

Sezione 5.4

• **Esercizio 5.4.1.** Si consideri il v.v.a. assolutamente continuo $X = (X_1, X_2)^T$ 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 $E[X_2 | X_1 = x_1]$ e $\text{Var}[X_2 | X_1 = x_1]$.

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

$$E[X_2 | X_1 = x_1] = \int_{x_1}^{\infty} x_2(x_2 - x_1) e^{-(x_2-x_1)} dx_2 = x_1 + 2,$$

ovvero il valore atteso condizionato è una funzione lineare. Infine, si osservi che risulta

$$E[X_2^2 | X_1 = x_1] = \int_{x_1}^{\infty} x_2^2(x_2 - x_1) e^{-(x_2-x_1)} dx_2 = x_1^2 + 4x_1 + 6,$$

da cui $\text{Var}[X_2 | X_1 = x_1] = 2$. □

• **Esercizio 5.4.2.** Si consideri il v.v.a. assolutamente continuo $X = (X_1, X_2)^T$ 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. $X = (X_1, X_2)^T$ 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

$$p_{X_2|X_1=x_1}(x_2) = \binom{n}{x_2} x_1^{x_2} (1-x_1)^{n-x_2} \mathbf{1}_{\{0,1,\dots,n\}}(x_2).$$

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

$$\begin{aligned} P(X_1 \in B_1, X_2 \in B_2) &= E[\mathbf{1}_{B_1}(X_1) \mathbf{1}_{B_2}(X_2)] = E[E[\mathbf{1}_{B_1}(X_1) \mathbf{1}_{B_2}(X_2) \mid X_1]] \\ &= E[\mathbf{1}_{B_1}(X_1) E[\mathbf{1}_{B_2}(X_2) \mid X_1]] . \end{aligned}$$

Dunque, posto $S_2 = \{0, 1, \dots, n\}$, 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. \square

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 $\Omega = \Omega_1 \times \dots \times \Omega_n$, 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 $\{(\omega_{j_1}, \dots, \omega_{j_n})\}$ di Ω come

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

dove $j_k = 1, 2$ e $k = 1, \dots, n$.

Si consideri la v.a. $X : \Omega \rightarrow S$, 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 $(\omega_{j_1}, \dots, \omega_{j_n}) \in \Omega$. La v.a. X è evidentemente di tipo discreto e si ha $S = \{0, 1, \dots, n\}$. 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 $x = 0, 1, \dots, n$. 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

$$p_X(x) = \binom{n}{x} p^x q^{n-x} \mathbf{1}_{\{0,1,\dots,n\}}(x).$$

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 $E[X_k] = p$ e $\text{Var}[X_k] = pq$ e, tenendo presente le proprietà relative alla media e varianza di somme di v.a. indipendenti, si ottiene

$$E[X] = \sum_{k=1}^n E[X_k] = np$$

e

$$\text{Var}[X] = \sum_{k=1}^n \text{Var}[X_k] = npq.$$

Questi risultati saranno anche verificati con un differente metodo nell'Esempio 7.2.4.

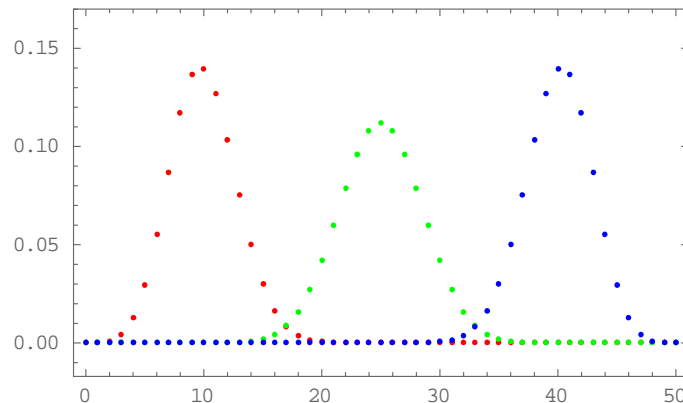


Figura 6.1.1. Funzione di probabilità per la legge Binomiale e $(n, p) = (50, \frac{1}{5}), (50, \frac{1}{2}), (50, \frac{4}{5})$ (rispettivamente in rosso, verde e blu).

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

$$g_n(x) = \begin{cases} \binom{n}{x} p_n^x q_n^{n-x} & x \in \{0, 1, \dots, n\} \\ 0 & x \in \mathbb{N} \setminus \{0, 1, \dots, n\} \end{cases}$$

In altre parole, $(g_n(x))_{x \geq 0}$ è il prolungamento su \mathbb{N} della legge Binomiale $\mathcal{B}(n, p_n)$. Se $x \in \{0, 1, \dots, n\}$ 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

$$E[X] = e^{-\lambda} \sum_{x=0}^{\infty} x \frac{\lambda^x}{x!} = e^{-\lambda} \sum_{x=1}^{\infty} \frac{\lambda^x}{(x-1)!} = \lambda e^{-\lambda} \sum_{x=0}^{\infty} \frac{\lambda^x}{x!} = \lambda.$$

Inoltre, tenendo presente che $E[X(X-1)] = E[X^2] - E[X]$ ed essendo

$$E[X(X-1)] = e^{-\lambda} \sum_{x=0}^{\infty} x(x-1) \frac{\lambda^x}{x!} = e^{-\lambda} \sum_{x=2}^{\infty} \frac{\lambda^x}{(x-2)!} = \lambda^2 e^{-\lambda} \sum_{x=0}^{\infty} \frac{\lambda^x}{x!} = \lambda^2,$$

si ha $E[X^2] = \lambda^2 + \lambda$. Dunque, risulta $\text{Var}[X] = \lambda$. Questi risultati saranno anche verificati con un differente metodo nell'Esempio 7.2.5.

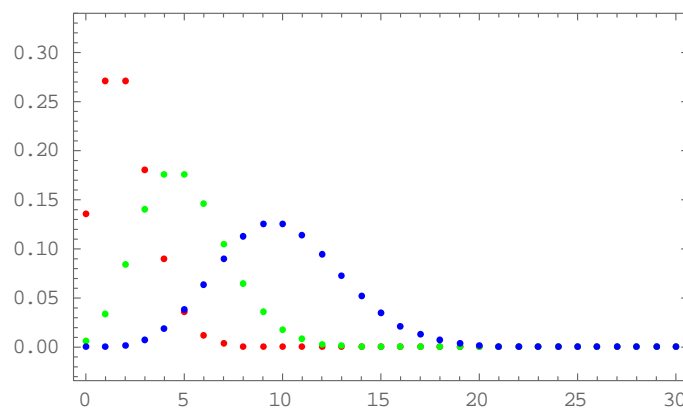


Figura 6.2.1. Funzione di probabilità per la legge di Poisson e $\lambda = 2, 5, 10$ (rispettivamente in rosso, verde e blu).

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. $X : \Omega \rightarrow S$, 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, affinché si verifichino k “successi”, si devono effettuare almeno k esperimenti). Si assuma inoltre che $k = 1, 2, \dots$. 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)

$$E[X_j] = p \sum_{x=0}^{\infty} x q^x = \frac{q}{(1-q)^2} = \frac{q}{p}.$$

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

$$E[X_j(X_j - 1)] = p \sum_{x=0}^{\infty} x(x-1)q^x = \frac{2pq^2}{(1-q)^3} = \frac{2q^2}{p^2}.$$

e quindi

$$\text{Var}[X_j] = \frac{q}{p^2}.$$

Dunque, per le proprietà relative alla media e varianza di somme di v.a. indipendenti si ha

$$E[X] = \sum_{j=1}^k E[X_j] = \frac{kq}{p}$$

e

$$\text{Var}[X] = \sum_{j=1}^k \text{Var}[X_j] = \frac{kq}{p^2}.$$

Questi risultati saranno anche verificati con un differente metodo nell'Esempio 7.2.6.

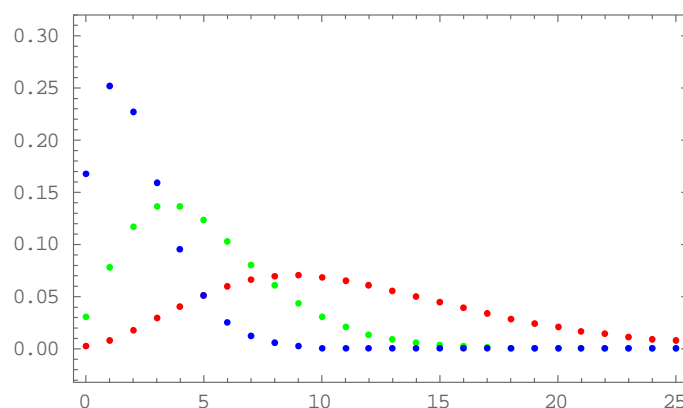


Figura 6.3.1. Funzione di probabilità per la legge Binomiale Negativa e $(k, p) = (5, \frac{3}{10}), (5, \frac{1}{2}), (5, \frac{7}{10})$ (rispettivamente in rosso, verde e blu).

Risulta inoltre interessante ottenere la f.p. limite per $k \rightarrow \infty$, 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 $D \leq N$. Si consideri inoltre l'esperimento aleatorio che consiste nell'estrarre in blocco n palline, dove n è un intero positivo tale che $n \leq N$. 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. $X : \Omega \rightarrow S$, che ad ogni risultato associa il 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 $n \geq N - D$ almeno $(n - N + D)$ palline estratte devono essere rosse, mentre se $n \geq D$ 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 $\mathcal{I}(n, D, N)$.

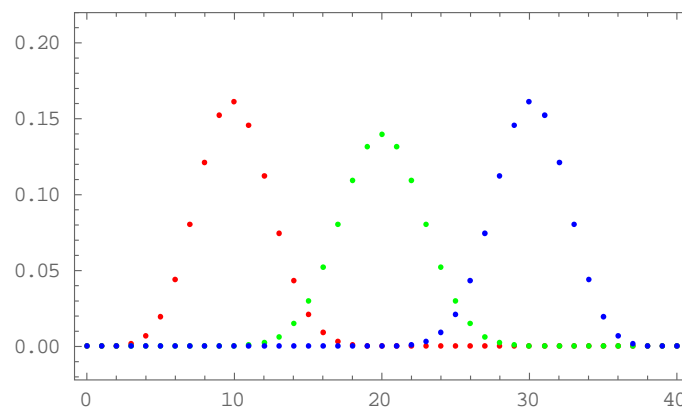


Figura 6.4.1. Funzione di probabilità per la legge Ipergeometrica e $(n, D, N) = (40, 50, 200), (40, 100, 200), (40, 150, 200)$ (rispettivamente in rosso, verde e blu).

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

$$E[X] = \sum_{x=\max(0, n-N+D)}^{\min(n, D)} x \frac{\binom{D}{x} \binom{N-D}{n-x}}{\binom{N}{n}} = \frac{nD}{N} \sum_{x=\max(0, n-N+D-1)}^{\min(n-1, D-1)} \frac{\binom{D-1}{x} \binom{N-D}{n-1-x}}{\binom{N-1}{n-1}} = \frac{nD}{N}$$

e, tenendo presente l'espressione della f.p. relativa alla legge Ipergeometrica $\mathcal{I}(n-2, D-2, N-2)$, si ha inoltre

$$\begin{aligned} E[X(X-1)] &= \sum_{x=\max(0, n-N+D)}^{\min(n, D)} x(x-1) \frac{\binom{D}{x} \binom{N-D}{n-x}}{\binom{N}{n}} \\ &= \frac{nD(n-1)(D-1)}{N(N-1)} \sum_{x=\max(0, n-N+D-2)}^{\min(n-2, D-2)} \frac{\binom{D-2}{x} \binom{N-D}{n-2-x}}{\binom{N-2}{n-2}} = \frac{nD(n-1)(D-1)}{N(N-1)} \end{aligned}$$

da cui

$$\text{Var}[X] = \frac{nD}{N} \frac{N-D}{N} \frac{N-n}{N-1}.$$

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 $N \rightarrow \infty$, 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 $\{0, 1, \dots, n\}$ 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 $\Omega = \Omega_1 \times \dots \times \Omega_n$ con $\Omega_l = \{\omega_1, \dots, \omega_k\}$. 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 $\omega_1, \dots, \omega_k$.

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 $l = 1, \dots, n$ con $j = 1, \dots, k$. 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 $\{(\omega_{j_1}, \dots, \omega_{j_n})\}$ di Ω come

$$P(\{(\omega_{j_1}, \dots, \omega_{j_n})\}) = \prod_{l=1}^n P_l(\{\omega_{j_l}\}),$$

dove $j_l = 1, \dots, k$ e $l = 1, \dots, n$.

Si consideri il v.v.a. $X : \Omega \rightarrow S$, che ad ogni risultato associa il vettore $x = (x_1, \dots, x_k)^T$, dove x_j rappresenta il numero di elementi del tipo ω_j in ogni $(\omega_{j_1}, \dots, \omega_{j_n}) \in \Omega$. Il v.v.a. $X = (X_1, \dots, X_k)^T$ è evidentemente di tipo discreto e $S \subset \mathbb{R}^k$ è tale che

$$S = \{(x_1, \dots, x_k) : x_j \in \{0, 1, \dots, n\}, \sum_{j=1}^k x_j = n\}.$$

Inoltre, la probabilità di ciascun evento elementare di Ω in cui si compare x_1 volte il risultato ω_1, \dots, 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 $p = (p_1, \dots, p_k)^T$, 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

$$\Sigma = n(\text{diag}(p) - pp^T),$$

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 $p = (p_1, p_2)^T$ 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

$$F_X(x) = G\left(\frac{x-a}{b}\right),$$

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

$$F_Z(z) = P(Z \leq z) = P(X \leq a + bz) = F_X(a + bz) = G(z)$$

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

$$f_X(x) = \frac{1}{b} f_Z\left(\frac{x-a}{b}\right),$$

mentre per quanto riguarda la f.r. risulta ovviamente

$$F_X(x) = F_Z\left(\frac{x-a}{b}\right).$$

Si noti infine che il momento di ordine r della v.a. X è esprimibile attraverso i momenti della v.a. standardizzata Z , ovvero

$$\mu_{X,r} = E[X^r] = E[(a + bZ)^r] = E\left[\sum_{k=1}^r \binom{r}{k} a^k (bZ)^{r-k}\right] = \sum_{k=1}^r \binom{r}{k} a^k b^{r-k} \mu_{Z,r-k},$$

da cui $\mu_X = a + b\mu_Z$. Analogamente, il momento centrale di ordine r della v.a. X risulta

$$E[(X - \mu_X)^r] = E[(a + bZ - a - b\mu_Z)^r] = b^r E[(Z - \mu_Z)^r],$$

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

$$f_X(x) = \frac{1}{b} e^{-\frac{x-a}{b}} e^{-e^{-\frac{x-a}{b}}}.$$

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 $\sigma_X^2 = \frac{b^2 \pi^2}{6}$. \square

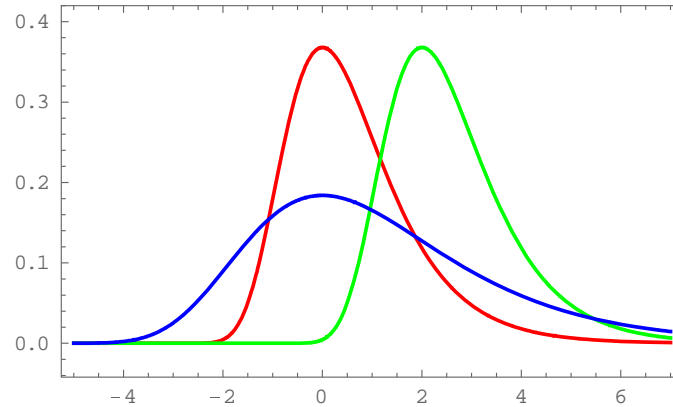


Figura 6.6.1. Densità di probabilità per la legge di Gumbel e $(a, b) = (0, 1), (2, 1), (0, 2)$ (rispettivamente in rosso, verde e blu).

• **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

$$f_Z(z) = kz^{k-1} e^{-z^k} \mathbf{1}_{]0, \infty[}(z).$$

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 $\sigma_X^2 = b^2(\Gamma(1 + 2k^{-1}) - \Gamma(1 + k^{-1})^2)$. \square

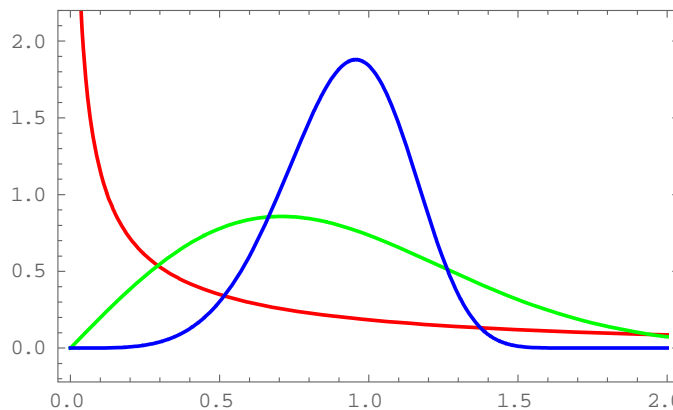


Figura 6.6.2. Densità di probabilità per la legge di Weibull e $(a, b, k) = (0, 1, \frac{1}{2}), (0, 1, 2), (0, 1, 5)$ (rispettivamente in rosso, verde e blu).

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

$$\left(\int_{-\infty}^{\infty} e^{-\frac{1}{2}z^2} dz \right)^2 = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(z^2+y^2)} dz dy = \int_0^{\infty} \int_0^{2\pi} r e^{-\frac{1}{2}r^2} dr d\theta = 2\pi,$$

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

$$F_X(x) = \Phi\left(\frac{x-\mu}{\sigma}\right).$$

Dal momento che la v.a. Z è simmetrica rispetto all'origine si ha $E[Z] = 0$, mentre

$$\text{Var}[Z] = E[Z^2] = \int_{-\infty}^{\infty} z^2 e^{-\frac{1}{2}z^2} dz = 1.$$

Si ha dunque $E[X] = \mu$ e $\text{Var}[X] = \sigma^2$, che rende più chiaro il motivo per cui si adotta la notazione $\mathcal{N}(\mu, \sigma^2)$ per indicare la legge Normale.

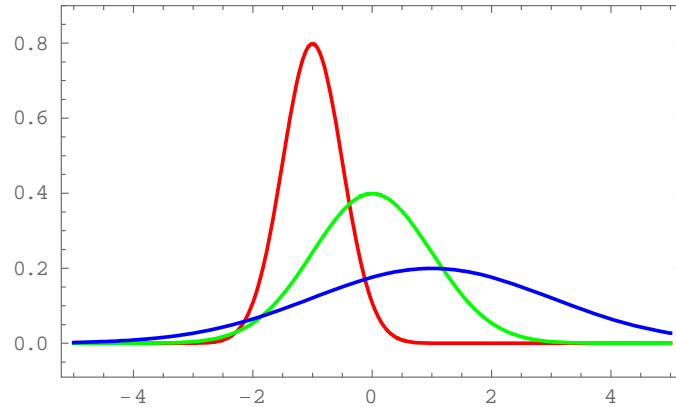


Figura 6.7.2. Densità di probabilità per la legge Normale e $(\mu, \sigma) = (-1, \frac{1}{2}), (0, 1), (1, 2)$ (rispettivamente in rosso, verde e blu).

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

$$F_X(x) = \frac{1}{\Gamma(k)} \gamma\left(k, \frac{x-a}{b}\right) \mathbf{1}_{]a, \infty[}(x).$$

Dal momento che risulta $\Gamma(k+1) = k\Gamma(k)$, si ha

$$E[Z] = \frac{1}{\Gamma(k)} \int_0^{\infty} z^k e^{-z} dz = \frac{\Gamma(k+1)}{\Gamma(k)} = k,$$

e

$$E[Z^2] = \frac{1}{\Gamma(k)} \int_0^{\infty} z^{k+1} e^{-z} dz = \frac{\Gamma(k+2)}{\Gamma(k)} = k(k+1),$$

da cui si ha anche $\text{Var}[X] = k$. Di conseguenza, si ha anche $E[X] = a + bk$ e $\text{Var}[X] = b^2k$.

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

$$F_X(x) = (1 - e^{-\frac{x-a}{b}}) \mathbf{1}_{]a, \infty[}(x).$$

Inoltre, per $a = 0$ e $b = 2$ e se $k = \frac{n}{2}$ dove $n = 1, 2, \dots$, si ottiene la *legge Chi-quadrato* con n gradi di libertà, che viene indicata di solito con χ_n^2 . In questo caso la v.a. X ammette d.p. data da

$$f_X(x) = \frac{1}{2\Gamma(\frac{n}{2})} \left(\frac{x}{2}\right)^{\frac{1}{2}n-1} e^{-\frac{1}{2}x} \mathbf{1}_{]0, \infty[}(x).$$

Si ha inoltre $E[X] = n$ e $\text{Var}[X] = 2n$. 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_1^2 . Inoltre, la v.a. X con legge χ_n^2 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).

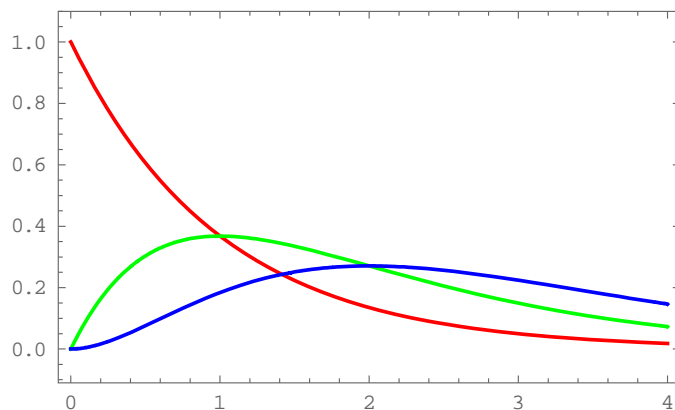


Figura 6.8.1. Densità di probabilità per la legge Gamma e $(a, b, k) = (0, 1, 1), (0, 1, 2), (0, 1, 3)$ (rispettivamente in rosso, verde e blu).

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

$$E[Z] = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \int_0^1 z^\alpha(1-z)^{\beta-1} dz = \frac{\alpha}{\alpha + \beta},$$

e

$$E[Z^2] = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \int_0^1 z^{\alpha+1}(1-z)^{\beta-1} dz = \frac{\alpha(\alpha + 1)}{(\alpha + \beta)(\alpha + \beta + 1)},$$

da cui si ha anche

$$\text{Var}[Z] = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}.$$

La media e la varianza della v.a. X sono quindi date da

$$E[X] = a + b \frac{\alpha}{\alpha + \beta},$$

mentre

$$\text{Var}[X] = b^2 \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta - 1)}.$$

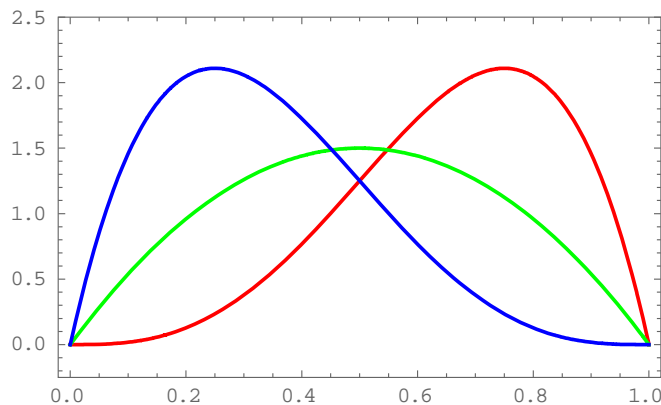


Figura 6.9.1. Densità di probabilità per la legge Beta e $(a, b, \alpha, \beta) = (0, 1, 4, 2), (0, 1, 2, 2), (0, 1, 2, 4)$ (rispettivamente in rosso, verde e blu).

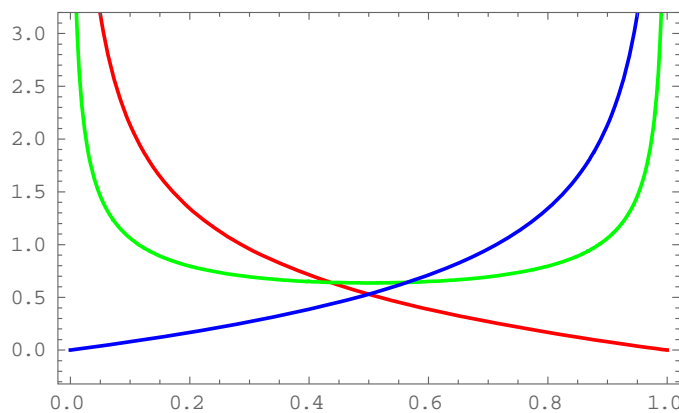


Figura 6.9.2. Densità di probabilità per la legge Beta e $(a, b, \alpha, \beta) = (0, 1, \frac{1}{2}, 2), (0, 1, \frac{1}{2}, \frac{1}{2}), (0, 1, 2, \frac{1}{2})$ (rispettivamente in rosso, verde e blu).

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 trasformato $(Y_1, Y_2)^T = g(Z_1, Z_2)$, dove $g(z_1, z_2) = \left(\frac{z_1}{z_1 + z_2}, z_2\right)^T$. Dal momento che risulta $g^{-1}(y_1, y_2) = \left(\frac{y_1 y_2}{1 - y_1}, y_2\right)^T$, allora $|J(g^{-1}(y_1, y_2))| = \frac{|y_2|}{(1 - y_1)^2}$, e quindi

$$f_{(Y_1, Y_2)}(y) = f_{(Z_1, Z_2)}\left(\frac{y_1 y_2}{1 - y_1}, y_2\right) \frac{|y_2|}{(1 - y_1)^2}.$$

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 $\chi_{n_1}^2$ e $\chi_{n_2}^2$ 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

$$\begin{aligned} f_Y(y) &= \int_{-\infty}^{\infty} f_{X_1}(x) f_{X_2}(xy) |x| dx \\ &= \frac{1}{2^{\frac{1}{2}(n_1+n_2)} \Gamma(\frac{n_1}{2}) \Gamma(\frac{n_2}{2})} y^{\frac{1}{2}n_1-1} \mathbf{1}_{]0,\infty[}(y) \int_0^{\infty} x^{\frac{1}{2}(n_1+n_2)-1} e^{-\frac{1}{2}(1+y)x} dx \\ &= \frac{\Gamma(\frac{n_1+n_2}{2})}{\Gamma(\frac{n_1}{2}) \Gamma(\frac{n_2}{2})} y^{\frac{1}{2}n_1-1} (1 + y)^{-\frac{1}{2}(n_1+n_2)} \mathbf{1}_{]0,\infty[}(y). \end{aligned}$$

Dunque, la v.a. $X = \frac{n_2}{n_1} Y$ ammette d.p. data da

$$f_X(x) = \frac{\Gamma(\frac{n_1+n_2}{2})}{\Gamma(\frac{n_1}{2}) \Gamma(\frac{n_2}{2})} \left(\frac{n_1}{n_2}\right)^{\frac{1}{2}n_1} x^{\frac{1}{2}n_1-1} \left(1 + \frac{n_1}{n_2} x\right)^{-\frac{1}{2}(n_1+n_2)} \mathbf{1}_{]0,\infty[}(x).$$



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 $n_2 = 3, 4, \dots$, si ha

$$E[X_2^{-1}] = \frac{1}{4\Gamma(\frac{n_2}{2})} \int_0^\infty \left(\frac{x}{2}\right)^{\frac{1}{2}n_2-2} e^{-\frac{x}{2}} dx = \frac{\Gamma(\frac{n_2}{2} - 1)}{2\Gamma(\frac{n_2}{2})} = \frac{1}{n_2 - 2},$$

mentre, se $n_2 = 4, 5, \dots$, risulta

$$E[X_2^{-2}] = \frac{1}{8\Gamma(\frac{n_2}{2})} \int_0^\infty \left(\frac{x}{2}\right)^{\frac{1}{2}n_2-3} e^{-\frac{x}{2}} dx = \frac{\Gamma(\frac{n_2}{2} - 2)}{4\Gamma(\frac{n_2}{2})} = \frac{1}{(n_2 - 2)(n_2 - 4)}.$$

Dunque, sulla base della definizione della v.a. X e dal momento che X_1 e X_2 sono indipendenti, si ottiene

$$E[X] = \frac{n_2}{n_1} E[X_1]E[X_2^{-1}] = \frac{n_2}{n_2 - 2},$$

se $n_2 = 3, 4, \dots$, e

$$\text{Var}[X] = \frac{n_2^2}{n_1^2} E[X_1^2]E[X_2^{-2}] - E[X]^2 = \frac{2n_2^2(n_1 + n_2 - 2)}{n_1(n_2 - 2)^2(n_2 - 4)},$$

se $n_2 = 4, 5, \dots$. Dunque, se $n_2 = 1, 2$ la media non è finita, mentre se $n_2 = 1, 2, 3, 4$ la varianza non è finita.

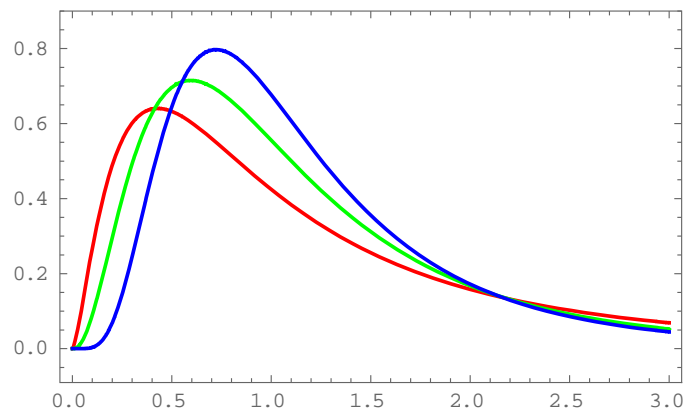


Figura 6.10.1. Densità di probabilità per la legge F di Snedecor e $(n_1, n_2) = (5, 5), (7, 10), (15, 10)$ (rispettivamente in rosso, verde e blu).

In secondo luogo, si consideri le v.a. indipendenti X_1 e X_2 rispettivamente con legge $\mathcal{N}(0, 1)$ e legge χ_n^2 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_1^2 . 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 $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$. Inoltre, tenendo presente che $X_1 \stackrel{\mathcal{L}}{=} -X_1$ e che $X = \text{sgn}(X_1)\sqrt{Y}$, 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).

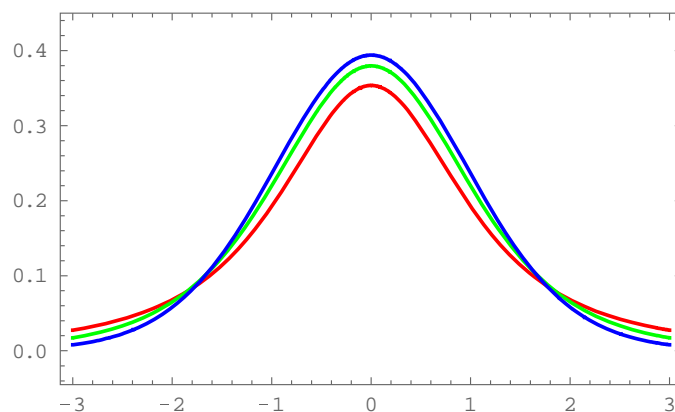


Figura 6.10.2. Densità di probabilità per la legge t di Student e $n = 2, 5, 20$ (rispettivamente in rosso, verde e blu).

Tenendo presente la definizione della v.a. X , dal momento che X_1 e X_2 sono indipendenti e che $E[X_1] = 0$ e $E[X_2^{-\frac{1}{2}}]$ è finito se $n = 2, 3, \dots$, la media della v.a. X è data da

$$E[X] = \sqrt{n} E[X_1] E[X_2^{-\frac{1}{2}}] = 0,$$

se $n = 2, 3, \dots$. Inoltre, dal momento che si ha $E[X_1^2] = 1$ e per quanto visto in precedenza in questa Sezione si ha $E[X_2^{-1}] = \frac{1}{n-2}$ se $n = 3, 4, \dots$, allora

$$\text{Var}[X] = E[X^2] = n E[X_1^2] E[X_2^{-1}] = \frac{n}{n-2},$$

se $n = 3, 4, \dots$. 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 $X = (X_1, \dots, X_n)^\top$ che ammette d.p.c. data da

$$f_X(x) = \det(2\pi\Sigma)^{-\frac{1}{2}} e^{-\frac{1}{2}(x-\mu)^\top \Sigma^{-1}(x-\mu)},$$

dove $x = (x_1, \dots, x_n)^T$, 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 $y = (y_1, \dots, y_n)^T$ e con Jacobiano dato da $\det(\Sigma)^{\frac{1}{2}}$, si ha

$$\begin{aligned} \int_{\mathbb{R}^n} e^{-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)} dx &= \det(\Sigma)^{\frac{1}{2}} \int_{\mathbb{R}^n} e^{-\frac{1}{2}y^T y} dy = \det(\Sigma)^{\frac{1}{2}} \int_{\mathbb{R}} \dots \int_{\mathbb{R}} \left(\prod_{k=1}^n e^{-\frac{1}{2}y_k^2} \right) dy_1 \dots dy_n \\ &= \det(\Sigma)^{\frac{1}{2}} \prod_{k=1}^n \int_{\mathbb{R}} e^{-\frac{1}{2}y_k^2} dy_k = (2\pi)^{\frac{n}{2}} \det(\Sigma)^{\frac{1}{2}} = \det(2\pi\Sigma)^{\frac{1}{2}}. \end{aligned}$$

Inoltre, la d.p.c. risulta costante per tutti i vettori x che appartengono al luogo geometrico $(x - \mu)^T \Sigma^{-1}(x - \mu) = c$, 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. $(X_{j_1}, \dots, X_{j_k})^T$ 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 $\mathcal{N}(\mu_{X_k}, \sigma_{X_k}^2)$. Risulta inoltre valido per una legge Normale Multivariata la seguente Proposizione.

Proposizione 6.11.1. *Il v.v.a. $X = (X_1, \dots, X_n)^T$ 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 $\det(2\pi\Sigma) = \prod_{k=1}^n 2\pi\sigma_{X_k}^2$ e $\Sigma^{-1} = \text{diag}(\sigma_{X_1}^{-2}, \dots, \sigma_{X_n}^{-2})$, da cui

$$(x - \mu)^T \Sigma^{-1}(x - \mu) = \sum_{k=1}^n \frac{(x_k - \mu_{X_k})^2}{\sigma_{X_k}^2}.$$

Sostituendo queste espressioni nella d.p.c. del v.v.a. X si ottiene infine

$$\begin{aligned} f_X(x) &= \frac{1}{\prod_{k=1}^n \sqrt{2\pi\sigma_{X_k}}} \exp\left(-\sum_{k=1}^n \frac{(x_k - \mu_{X_k})^2}{2\sigma_{X_k}^2}\right) \\ &= \prod_{k=1}^n \frac{1}{\sqrt{2\pi\sigma_{X_k}}} \exp\left(-\frac{(x_k - \mu_{X_k})^2}{2\sigma_{X_k}^2}\right) = \prod_{k=1}^n f_{X_k}(x_k), \end{aligned}$$

che è quanto si voleva dimostrare. \square

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 $Z = (Z_1, \dots, Z_n)^T$ che ammette d.p.c. data da

$$f_Z(z) = f_Z(z_1, \dots, z_n) = \frac{\Gamma(\alpha_0)}{\prod_{k=1}^n \Gamma(\alpha_k)} \prod_{k=1}^n z_k^{\alpha_k-1} \mathbf{1}_S(z_1, \dots, z_n),$$

dove $z = (z_1, \dots, z_n)^T$, mentre $\alpha = (\alpha_1, \dots, \alpha_n)^T$ è un vettore di parametri tale che $\alpha_k \in]0, \infty[$ per $k = 1, \dots, n$ e $\alpha_0 = \sum_{k=1}^n \alpha_k$, e

$$S = \{(z_1, \dots, z_n) : z_k \in]0, 1[, \sum_{k=1}^n z_k = 1\}.$$

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

$$\frac{\prod_{k=1}^n \Gamma(\alpha_k)}{\Gamma(\alpha_0)} = \int_S \prod_{k=1}^n z_k^{\alpha_k-1} dz_1 \dots dz_n$$

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 $X = (X_1, \dots, X_n)^T$ 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 $Y = (Y_1, \dots, Y_n)^T = g(X)$, dove

$$g(x) = \left(\frac{x_1}{\sum_{k=1}^n x_k}, \dots, \frac{x_{n-1}}{\sum_{k=1}^n x_k}, \sum_{k=1}^n x_k \right)^T,$$

per cui risulta $g^{-1}(y) = (y_1 y_n, \dots, y_{n-1} y_n, (1 - \sum_{k=1}^{n-1} y_k) y_n)^T$ e dunque $|J(g^{-1}(y))| = y_n^{n-1}$. Dal momento che

$$f_X(x_1, \dots, x_n) = \prod_{k=1}^n \frac{1}{\Gamma(\alpha_k)} x_k^{\alpha_k-1} e^{-x_k} \mathbf{1}_{]0, \infty[}(x_k),$$

allora si ha

$$f_Y(y_1, \dots, y_n) = \frac{1}{\prod_{k=1}^n \Gamma(\alpha_k)} \left(\prod_{k=1}^{n-1} y_k^{\alpha_k-1} \right) \left(1 - \sum_{k=1}^{n-1} y_k \right)^{\alpha_n-1} y_n^{\alpha_0-1} e^{-y_n} \mathbf{1}_T(y_1, \dots, y_{n-1}) \mathbf{1}_{]0, \infty[}(y_n),$$

dove

$$T = \{(y_1, \dots, y_{n-1}) : y_k \in]0, 1[, \sum_{k=1}^{n-1} y_k \leq 1\}.$$

Dunque, posto $V = (Y_1, \dots, Y_{n-1})^T$ si ha

$$\begin{aligned}
 f_V(y_1, \dots, y_{n-1}) &= \frac{1}{\prod_{k=1}^n \Gamma(\alpha_k)} \left(\prod_{k=1}^{n-1} y_k^{\alpha_k-1} \right) \left(1 - \sum_{k=1}^{n-1} y_k \right)^{\alpha_n-1} \mathbf{1}_T(y_1, \dots, y_{n-1}) \int_0^\infty y_n^{\alpha_0-1} e^{-y_n} dy_n \\
 &= \frac{\Gamma(\alpha_0)}{\prod_{k=1}^n \Gamma(\alpha_k)} \left(\prod_{k=1}^{n-1} y_k^{\alpha_k-1} \right) \left(1 - \sum_{k=1}^{n-1} y_k \right)^{\alpha_n-1} \mathbf{1}_T(y_1, \dots, y_{n-1}).
 \end{aligned}$$

Il v.v.a. linearmente degenere $(Z_1, \dots, Z_n)^\top = (Y_1, \dots, Y_{n-1}, 1 - \sum_{k=1}^{n-1} Y_k)^\top$ 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, \dots, j_k) di k v.a. $(Z_{j_1}, \dots, Z_{j_k})^\top$ con $k = 2, \dots, n-1$, 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

$$\Sigma = \frac{1}{\alpha_0^2(\alpha_0 + 1)} (\alpha_0 \text{diag}(\alpha) - \alpha \alpha^\top),$$

con varianza generalizzata che deve necessariamente risultare $\det(\Sigma) = 0$.

Si osservi infine che per $n = 2$ e $\alpha = (\alpha_1, \alpha_2)^\top$ 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 $(Z_1, Z_2) \stackrel{\mathcal{L}}{=} (Y, 1 - Y)$.

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 (Ω, \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

$$\begin{aligned} P(X_1 = X_2) &= \sum_{x=0}^n P(X_1 = x)P(X_2 = x) = 2^{-2n} \sum_{x=0}^n \binom{n}{x}^2 \\ &= 2^{-2n} \sum_{x=0}^n \binom{n}{x} \binom{n}{n-x} = \binom{2n}{n} 2^{-2n}. \end{aligned}$$

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 $n \rightarrow \infty$. \square

• **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 $S = \{n+1, \dots, 2n+1\}$. 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

$$p_X(x) = 2 \binom{x-1}{n} \left(\frac{1}{2}\right)^{x-1} \frac{1}{2} \mathbf{1}_{\{n+1, \dots, 2n+1\}}(x) = \binom{x-1}{n} 2^{-x+1} \mathbf{1}_{\{n+1, \dots, 2n+1\}}(x).$$

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

$$\begin{aligned} E[X] &= \sum_{x=n+1}^{2n+1} x \binom{x-1}{n} 2^{-x+1} = (n+1) \sum_{x=n+1}^{2n+1} \binom{x}{n+1} 2^{-x+1} \\ &= 2(n+1) \sum_{x=n+2}^{2n+2} \binom{x-1}{n+1} 2^{-x+1}, \end{aligned}$$

ovvero, aggiungendo e sottraendo in modo opportuno e tenendo di nuovo presente l'Esercizio 3.3.4, si ottiene

$$\begin{aligned} E[X] &= 2(n+1) \left(\sum_{x=n+2}^{2n+3} \binom{x-1}{n+1} 2^{-x+1} - \binom{2n+2}{n+1} 2^{-2n-2} \right) \\ &= 2(n+1) \left(1 - \frac{(2n+2)(2n+1)}{(n+1)^2} \binom{2n}{n} 2^{-2n-2} \right) = 2(n+1) - (2n+1) \binom{2n}{n} 2^{-2n}. \end{aligned}$$

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

$$E[Y] = 2n + 1 - E[X] = (2n + 1) \binom{2n}{n} 2^{-2n} - 1.$$

Tenendo presente l'Esercizio 6.1.1, si ha infine

$$E[Y] \sim \frac{2n + 1}{\sqrt{\pi n}} - 1.$$

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. \square

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

$$\begin{aligned} p_{X_1|X_1+X_2=n}(x_1) &= P(X_1 = x_1 | X_1 + X_2 = n) = \frac{P(X_1 + X_2 = n | X_1 = x_1)P(X_1 = x_1)}{P(X_1 + X_2 = n)} \\ &= \frac{P(X_2 = n - x_1)P(X_1 = x_1)}{P(X_1 + X_2 = n)} \mathbf{1}_{\{0,1,\dots,n\}}(x_1). \end{aligned}$$

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

$$\begin{aligned} p_{X_1|X_1+X_2=n}(x_1) &= \frac{e^{-\lambda_2} \frac{\lambda_2^{n-x_1}}{(n-x_1)!} e^{-\lambda_1} \frac{\lambda_1^{x_1}}{x_1!}}{e^{-(\lambda_1+\lambda_2)} \frac{(\lambda_1+\lambda_2)^n}{n!}} \mathbf{1}_{\{0,1,\dots,n\}}(x_1) \\ &= \binom{n}{x_1} \left(\frac{\lambda_1}{\lambda_1 + \lambda_2} \right)^{x_1} \left(1 - \frac{\lambda_1}{\lambda_1 + \lambda_2} \right)^{n-x_1} \mathbf{1}_{\{0,1,\dots,n\}}(x_1), \end{aligned}$$

ovvero la v.a. X_1 condizionata all'evento $\{X_1 + X_2 = n\}$ si distribuisce con legge Binomiale $\mathcal{B}(n, p)$ dove $p = \frac{\lambda_1}{\lambda_1 + \lambda_2}$. \square

• **Esercizio 6.2.2.** (Problema della segretaria distratta, terza parte) Si ottenga il limite della successione $(g_n(x))_{x \geq 0}$, dove

$$g_n(x) = \begin{cases} \sum_{k=x}^n \binom{k}{x} \frac{(-1)^{k-x}}{k!} & x \in \{0, 1, \dots, n\} \\ 0 & x \in \mathbb{N} \setminus \{0, 1, \dots, n\}, \end{cases}$$

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 $p_X(m)$. \square

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). \square

• **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 $E[X]$ e $\text{Var}[X]$.

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

$$E[X] = \sum_{k=1}^n E[Y_k + 1] = \sum_{k=1}^n \frac{n}{n - k + 1} = n \sum_{k=1}^n \frac{1}{k} = nH_n,$$

dove H_n rappresenta l' n -esimo numero armonico. Per le proprietà dei numeri armonici si ottiene inoltre che

$$E[X] = n \log(n) + \gamma n + \frac{1}{2} + O(n^{-1}),$$

dove $\gamma \simeq 0.5772$ rappresenta la costante di Eulero-Mascheroni. Risulta anche

$$\begin{aligned} \text{Var}[X] &= \sum_{k=1}^n \text{Var}[Y_k] = \sum_{k=1}^n \frac{n(k-1)}{(n-k+1)^2} = n \sum_{k=1}^n \frac{n-k}{k^2} \\ &= n^2 \sum_{k=1}^n \frac{1}{k^2} - n \sum_{k=1}^n \frac{1}{k} = n^2 H_{n,2} - nH_n, \end{aligned}$$

dove $H_{n,2}$ rappresenta l' n -esimo numero armonico del secondo ordine. Si noti che

$$\text{Var}[X] < n^2 H_{n,2} < n^2 \zeta(2) = \frac{\pi^2 n^2}{6}$$

e dunque dalla disuguaglianza di Chebyshev si ottiene anche

$$P(|X - nH_n| \geq cn) \leq \frac{\pi^2}{6c^2}. \quad \square$$

• **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

$$P(X > x + y | X > x) = \frac{P(\{X > x + y\} \cap \{X > x\})}{P(X > x)} = \frac{P(X > x + y)}{P(X > x)},$$

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

$$P(Z > x + y | Z \geq x) = P(Z > y). \quad \square$$

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 $\mathcal{I}(k, n_1, n_1 + n_2)$. \square

Sezione 6.5

• **Esercizio 6.5.1.** Si consideri il v.v.a. $X = (X_1, \dots, X_k)^\top$ 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 $x = (x_1, \dots, x_k)^\top$ e $S = \{(x_1, \dots, x_k) : x_j \in \{0, 1, \dots, n\}, \sum_{j=1}^k x_j = n\}$. 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

$$p_{X|\sum_{j=1}^k X_j}(x) = \frac{\prod_{j=1}^k e^{-\lambda_j} \frac{\lambda_j^{x_j}}{x_j!}}{e^{-\sum_{j=1}^k \lambda_j} \frac{(\sum_{j=1}^k \lambda_j)^n}{n!}} \mathbf{1}_S(x) = \binom{n}{x_1 \dots x_k} \prod_{j=1}^k \left(\frac{\lambda_j}{\sum_{j=1}^k \lambda_j} \right)^{x_j} \mathbf{1}_S(x).$$

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 $p = (p_1, \dots, p_k)^\top$ con $p_j = \frac{\lambda_j}{\sum_{j=1}^k \lambda_j}$. \square

• **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. $X = (X_1, \dots, X_6)^\top$, 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 $(\frac{1}{6}, \dots, \frac{1}{6})^\top$. Quindi, se p_X rappresenta la f.p.c. del v.v.a. X , la probabilità richiesta è data da

$$p_X(2, \dots, 2) = \binom{12}{2 \dots 2} \left(\frac{1}{6} \right)^{12} = \frac{1925}{559872} \simeq 0.003.$$

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). \square

Sezione 6.7

• **Esercizio 6.7.1.** Si consideri la v.a. X con legge Normale $\mathcal{N}(0, \sigma^2)$ e si determini $E[e^{aX}]$ dove $a \in \mathbb{R}$.

Soluzione. Mediante il completamento di quadrato, si ha

$$E[e^{aX}] = \int_{-\infty}^{\infty} e^{ax} \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{x^2}{2\sigma^2}} dx = e^{\frac{1}{2}a^2\sigma^2} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{1}{2\sigma^2}(x-a\sigma^2)^2} dx.$$

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

$$E[e^{aX}] = e^{\frac{1}{2}a^2\sigma^2}. \quad \square$$

• **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 $\alpha = 0$. \square

• **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 $E[X^r] = E[Y^r]$ 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 $r = 0, 1, \dots$ 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

$$\begin{aligned} \int_0^\infty y^r \sin(2\pi \log(y)) f_X(y) dy &= e^{\frac{1}{2}r^2} \int_{-\infty}^\infty \sin(2\pi(z - r)) \phi(z - r) dz \\ &= e^{\frac{1}{2}r^2} \int_{-\infty}^\infty \sin(2\pi z) \phi(z) dz = e^{\frac{1}{2}r^2} E[\sin(2\pi Z)] = 0, \end{aligned}$$

dove si è tenuto presente che $\sin(2\pi z) = \sin(2\pi(z - r))$. Quindi, segue che $E[X^r] = E[Y^r]$ 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. \square

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. \square

• **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. \square

• **Esercizio 6.8.3.** Si consideri il v.v.a. $X = (X_1, \dots, X_k)^T$ 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

$$E[X_j^r] = \frac{1}{\Gamma(\alpha + \frac{j-1}{k})} \int_0^\infty x^{\alpha+r+\frac{j-1}{k}-1} e^{-x} dx = \frac{\Gamma(\alpha + r + \frac{j-1}{k})}{\Gamma(\alpha + \frac{j-1}{k})}.$$

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

$$E[Y^{kr}] = E[k^{kr} \prod_{j=1}^k X_j^r] = k^{kr} \prod_{j=1}^k E[X_j^r] = k^{kr} \prod_{j=1}^k \frac{\Gamma(\alpha + r + \frac{j-1}{k})}{\Gamma(\alpha + \frac{j-1}{k})} = \frac{\Gamma(k\alpha + kr)}{\Gamma(k\alpha)}.$$

Dunque, dal momento che $k \in \mathbb{Z}^+$ anche $s = kr \in \mathbb{Z}^+$ e si ha

$$E[Y^s] = \frac{\Gamma(k\alpha + s)}{\Gamma(k\alpha)},$$

che in effetti rappresenta il momento di ordine s di una v.a. distribuita con legge Gamma ridotta $\mathcal{G}(0, 1, k\alpha)$. \square

• **Esercizio 6.8.4.** Si assuma che $\lfloor x \rfloor = \max\{n \in \mathbb{Z} : n \leq x\}$ per $x \in \mathbb{R}$, ovvero $\lfloor x \rfloor$ è la funzione parte intera. Quindi, $x - \lfloor x \rfloor$ è la funzione parte frazionaria. Se X è una v.a. che si distribuisce con legge Esponenziale ridotta $\mathcal{G}(0, 1, 1)$, si verifichi che le v.a. $Y = \lfloor X \rfloor$ e $Z = X - \lfloor X \rfloor$ sono indipendenti.

Soluzione. Dalla definizione, si ha $P(Y \in \mathbb{N}) = 1$ e $P(Z \in [0, 1]) = 1$. Dunque, la v.a. Y è discreta, mentre la v.a. Z è assolutamente continua. Assumendo che $y \in \mathbb{N}$ e $z \in [0, 1]$, si osservi che

$$P(Y = y, Z \leq z) = P(y \leq X \leq y + z) = \int_y^{y+z} e^{-x} dx = e^{-y}(1 - e^{-z}).$$

Inoltre, la f.p. relativa alla v.a. Y è data da

$$p_Y(y) = P(Y = y) = P(y \leq X < y + 1) = \mathbf{1}_{\mathbb{N}}(y) \int_y^{y+1} e^{-x} dx = e^{-y}(1 - e^{-1}) \mathbf{1}_{\mathbb{N}}(y)$$

e dunque la v.a. Y si distribuisce con legge Geometrica di parametro e^{-1} . Per $y \in \mathbb{N}$ e $z \in [0, 1]$ si ha

$$P(Z \leq z | Y = y) = \frac{P(Z \leq z, Y = y)}{P(Y = y)} = \frac{1 - e^{-z}}{1 - e^{-1}}.$$

Dal momento che la precedente probabilità condizionata non dipende da y , risulta che Y e Z sono v.a. indipendenti in quanto $P(Z \leq z | Y = y) = P(Z \leq z)$. Inoltre, la f.r. della v.a. Z è data da

$$F_Z(z) = P(Z \leq z) = \frac{1 - e^{-z}}{1 - e^{-1}} \mathbf{1}_{[0,1]}(y) + \mathbf{1}_{[1,\infty]}(y),$$

mentre la corrispondente d.p. risulta

$$f_Z(z) = \frac{e^{-z}}{1 - e^{-1}} \mathbf{1}_{[0,1]}(y),$$

ovvero la v.a. Z si distribuisce con legge Esponenziale troncata. \square

• **Esercizio 6.8.5.** (Una identità binomiale notevole) Dato un v.v.a. $X = (X_1, X_2)^T$, mediante il binomio di Newton risulta immediato osservare che

$$E[(X_1 + X_2)^n] = \sum_{k=0}^n \binom{n}{k} E[X_1^k X_2^{n-k}].$$

Sulla base della precedente relazione, se $n \in \mathbb{N}$ si verifichi l'identità binomiale

$$\sum_{k=0}^n \binom{2k}{k} \binom{2n-2k}{n-k} = 4^n.$$

Soluzione. Si osservi che il momento di ordine r di una v.a. Z distribuita con legge Chi-quadrato χ_n^2 con n gradi di libertà è dato da

$$E[Z^r] = \int_0^\infty x^r \frac{1}{2\Gamma(\frac{n}{2})} \left(\frac{x}{2}\right)^{\frac{1}{2}n-1} e^{-\frac{1}{2}x} dx = \frac{2^r}{\Gamma(\frac{n}{2})} \int_0^\infty x^{\frac{1}{2}n+r-1} e^{-x} dx = \frac{2^r \Gamma(\frac{n}{2} + r)}{\Gamma(\frac{n}{2})}.$$

Inoltre, mediante la formula di duplicazione della funzione Gamma di Eulero, ovvero

$$\Gamma(2k) = \frac{2^{2k-1}}{\Gamma(\frac{1}{2})} \Gamma(k) \Gamma\left(\frac{1}{2} + k\right),$$

per $n = 1$ l'espressione di $E[Z^r]$ si riduce a

$$E[Z^r] = \frac{2^r \Gamma(\frac{1}{2} + r)}{\Gamma(\frac{1}{2})} = \frac{2^{1-r} \Gamma(2r)}{\Gamma(r)} = \frac{2^{1-r} (2r-1)!}{(r-1)!} = \frac{(2r)!}{2^r r!}.$$

Si assuma che X_1 e X_2 siano v.a. indipendenti distribuite con la legge χ_1^2 . Dunque, si ha

$$\begin{aligned} \sum_{k=0}^n \binom{n}{k} E[X_1^k X_2^{n-k}] &= \sum_{k=0}^n \binom{n}{k} E[X_1^k] E[X_2^{n-k}] = \sum_{k=0}^n \binom{n}{k} \frac{(2k)!}{2^k k!} \frac{(2n-2k)!}{2^{n-k} (n-k)!} \\ &= \frac{n!}{2^n} \sum_{k=0}^n \binom{2k}{k} \binom{2n-2k}{n-k}. \end{aligned}$$

Inoltre, la d.p. della v.a. $(X_1 + X_2)$ è data da

$$\begin{aligned} f_{X_1+X_2}(x) &= \int_{-\infty}^{\infty} f_{X_1}(z) f_{X_2}(x-z) dz = \frac{1}{2\pi} e^{-\frac{1}{2}x} \mathbf{1}_{[0,\infty[}(x) \int_0^x \frac{1}{\sqrt{z(x-z)}} dz \\ &= \frac{1}{2\pi} e^{-\frac{1}{2}x} \mathbf{1}_{[0,\infty[}(x) \int_0^1 \frac{1}{\sqrt{z(1-z)}} dz = \frac{1}{2} e^{-\frac{1}{2}x} \mathbf{1}_{[0,\infty[}(x), \end{aligned}$$

ovvero $(X_1 + X_2)$ si distribuisce con legge χ_2^2 e quindi

$$E[(X_1 + X_2)^n] = 2^n n!.$$

Sostituendo le espressioni ottenute nella relazione suggerita, si verifica immediatamente l'identità binomiale. \square

Sezione 6.9

• **Esercizio 6.9.1.** Si consideri il v.v.a. $X = (X_1, X_2)^T$ 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

$$E[X_1^r] = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \int_0^1 x^{\alpha+r-1} (1-x)^{\beta-1} dx = \frac{\Gamma(\alpha + \beta)\Gamma(\alpha + r)}{\Gamma(\alpha)\Gamma(\alpha + \beta + r)}.$$

In modo simile, si ha

$$E[X_2^r] = \frac{\Gamma(\alpha + \beta + \gamma)\Gamma(\alpha + \beta + r)}{\Gamma(\alpha + \beta)\Gamma(\alpha + \beta + \gamma + r)}.$$

Inoltre, si ha

$$E[Y^r] = E[X_1^r X_2^r] = E[X_1^r] E[X_2^r] = \frac{\Gamma(\alpha + \beta + \gamma)\Gamma(\alpha + r)}{\Gamma(\alpha)\Gamma(\alpha + \beta + \gamma + r)},$$

che è in effetti il momento di ordine r di una v.a. con legge Beta ridotta $\mathcal{BE}(0, 1, \alpha, \beta + \gamma)$. \square

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

$$f_X(x) = \frac{\Gamma(n)}{\Gamma(\frac{n}{2})^2} x^{\frac{1}{2}n-1} (1+x)^{-n} \mathbf{1}_{]0, \infty[}(x).$$

Si consideri inizialmente la trasformata $U = \sqrt{X}$, per cui è immediato ottenere che

$$f_U(u) = \frac{2\Gamma(n)}{\Gamma(\frac{n}{2})^2} u^{n-1} (1+u^2)^{-n} \mathbf{1}_{]0, \infty[}(u).$$

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

$$f_V(v) = \frac{\Gamma(n)}{2^{n-1} \Gamma(\frac{n}{2})^2} (1+v^2)^{-\frac{1}{2}(n+1)}.$$

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

$$f_Y(y) = \frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi} \Gamma(\frac{n}{2})} \left(1 + \frac{y^2}{n}\right)^{-\frac{1}{2}(n+1)},$$

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. $X = (X_1, X_2)^T$ distribuito con legge Normale Multivariata $\mathcal{N}_2(\mu, \Sigma)$, dove

$$\mu = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix},$$

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

$$f_Y(y) = f_Z(\tan(y)) \frac{1}{\cos^2(y)} \mathbf{1}_{[-\frac{\pi}{2}, \frac{\pi}{2}]}(y) = \frac{\sqrt{1-\rho^2}}{\pi(1-\rho \sin(2y))} \mathbf{1}_{[-\frac{\pi}{2}, \frac{\pi}{2}]}(y). \quad \square$$

• **Esercizio 6.11.2.** Si consideri il v.v.a. $X = (X_1, X_2)^T$ 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 $E[X_2 | X_1 = x_1]$.

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 $E[X_2 | X_1 = x_1] = \rho x_1$ e dunque il valore atteso condizionato è una funzione lineare. Infine, si noti che $\text{Var}[X_2 | X_1 = x_1] = 1-\rho^2$ è una costante. \square

• **Esercizio 6.11.3.** Si consideri il v.v.a. assolutamente continuo $X = (X_1, X_2)^T$ che ammette d.p.c.

$$f_X(x_1, x_2) = f_{X_1}(x_1)f_{X_2}(x_2)(1 - \sin(x_1)\sin(x_2)),$$

dove $f_{X_1} = f_{X_2} = \phi$. Si verifichi che f_X è una d.p.c.

Soluzione. Tenendo presente che ϕ è una funzione pari, mentre la funzione seno è dispari, si ha $\int_{-\infty}^{\infty} f_{X_2}(x_2)\sin(x_2)dx_2 = 0$, da cui

$$\int_{-\infty}^{\infty} f_{X_2}(x_2)(1 - \sin(x_1)\sin(x_2)) dx_2 = 1$$

e

$$\int_{-\infty}^{\infty} f_X(x_1, x_2) dx_2 = f_{X_1}(x_1) = \phi(x_1),$$

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. Si osservi che si ha inoltre

$$\int_{-\infty}^{\infty} f_X(x_1, x_2) dx_1 = f_{X_2}(x_2) = \phi(x_2),$$

ovvero le v.a. X_1 e X_2 si distribuiscono con legge Normale $\mathcal{N}(0, 1)$. Tuttavia, il v.v.a. X non si distribuisce con legge Normale multivariata. Si deve anche notare che f_X è in generale una d.p.c. assumendo che X_1 e X_2 siano v.a. assolutamente continue simmetriche, ovvero che le d.p. f_{X_1} e f_{X_2} siano funzioni pari. \square

• **Esercizio 6.11.4.** (Una identità binomiale notevole) Dato un v.v.a. $X = (X_1, X_2)^T$, mediante il binomio di Newton risulta immediato osservare che

$$E[(X_2 - X_1)^n] = \sum_{k=0}^n \binom{n}{k} (-1)^k E[X_1^k X_2^{n-k}].$$

Sulla base della precedente relazione, se $n \in \mathbb{N}$ si verifichi l'identità binomiale

$$\sum_{k=0}^n (-1)^k \binom{2k}{k} \binom{2n-2k}{n-k} = \begin{cases} 2^n \binom{n}{\frac{n}{2}} & n \text{ pari} \\ 0 & n \text{ dispari} \end{cases}.$$

Soluzione. Si noti che il momento di ordine r di una v.a. Z distribuita con legge Normale ridotta $\mathcal{N}(0, 1)$ è dato da

$$E[Z^r] = \int_{-\infty}^{\infty} x^r \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx = \frac{2^{\frac{1}{2}r}}{\sqrt{\pi}} \int_0^{\infty} x^{\frac{1}{2}(r-1)} e^{-x} dx = \frac{2^{\frac{1}{2}r} \Gamma(\frac{r+1}{2})}{\Gamma(\frac{1}{2})}.$$

per r pari, mentre è immediato verificare che $E[Z^r] = 0$ per r dispari. Inoltre, tenendo presente la formula di duplicazione della funzione Gamma di Eulero, per r pari si ha

$$E[Z^r] = \frac{2^{\frac{1}{2}r} \Gamma(\frac{r+1}{2})}{\Gamma(\frac{1}{2})} = \frac{\Gamma(r)}{2^{\frac{1}{2}r-1} \Gamma(\frac{r}{2})} = \frac{(r-1)!}{2^{\frac{1}{2}r-1} (\frac{r}{2}-1)!} = \frac{r!}{2^{\frac{1}{2}r} (\frac{r}{2})!}.$$

Si assuma che X_1 e X_2 siano v.a. indipendenti distribuite con la legge χ_1^2 . Tenendo presente l'Esercizio 6.8.5, si ha

$$\begin{aligned} \sum_{k=0}^n \binom{n}{k} (-1)^k E[X_1^k X_2^{n-k}] &= \sum_{k=0}^n \binom{n}{k} E[X_1^k] E[X_2^{n-k}] = \sum_{k=0}^n \binom{n}{k} \frac{(2k)!}{2^k k!} \frac{(2n-2k)!}{2^{n-k} (n-k)!} \\ &= \frac{n!}{2^n} \sum_{k=0}^n (-1)^k \binom{2k}{k} \binom{2n-2k}{n-k}. \end{aligned}$$

Inoltre, risulta che $X_1 \stackrel{\mathcal{L}}{=} Z_1^2$ e $X_2 \stackrel{\mathcal{L}}{=} Z_2^2$, dove Z_1 e Z_2 sono distribuite con legge Normale ridotta $\mathcal{N}(0, 1)$. Si osservi che $Y_1 = \frac{\sqrt{2}}{2}(Z_1 + Z_2)$ e $Y_2 = \frac{\sqrt{2}}{2}(Z_2 - Z_1)$ sono v.a. indipendenti a loro volta distribuite con legge Normale ridotta. In effetti, Y_1 e Y_2 sono combinazioni lineari di due v.a. con legge Normale, tali che $E[Y_1] = E[Y_2] = 0$, $\text{Var}[Y_1] = \text{Var}[Y_2] = 1$, mentre $\text{Cov}[Y_1, Y_2] = 0$ che implica l'indipendenza nel caso di una legge Normale bivariata. Dunque, si ha

$$E[(X_2 - X_1)^n] = E[(Z_2^2 - Z_1^2)^n] = 2^n E[Y_1^n] E[Y_2^n] = \begin{cases} \frac{n!n!}{(\frac{n}{2})!(\frac{n}{2})!} & n \text{ pari} \\ 0 & n \text{ dispari} \end{cases}.$$

Sostituendo le espressioni ottenute nella relazione suggerita, si verifica immediatamente l'identità binomiale. \square

Sezione 6.12

• **Esercizio 6.12.1.** (Statistica ordinata) Dato il v.v.a. assolutamente continuo $Z = (Z_1, \dots, Z_{n+1})^T$ distribuito con legge di Dirichlet $\mathcal{D}(\alpha)$ dove $\alpha = (1, \dots, 1)^T$, si determini la d.p.c. della trasformata $Y = (Y_1, \dots, Y_{n+1})^T = g(Z)$, dove

$$g(z) = (z_1, z_1 + z_2, \dots, z_1 + \dots + z_{n+1})^T.$$

Si determini inoltre le leggi delle componenti marginali del v.v.a. Z .

Soluzione. Dal momento che risulta $g^{-1}(y) = (y_1, y_2 - y_1, \dots, y_{n+1} - y_n)^T$, si ha $|J(g^{-1}(y))| = 1$. Dunque, risulta

$$f_Y(y_1, \dots, y_{n+1}) = n! \mathbf{1}_S(y_1, \dots, y_{n+1}),$$

dove $S = \{(y_1, \dots, y_{n+1}) : 0 \leq y_1 \leq y_2 \leq \dots \leq y_{n+1} = 1\}$. 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 $(X_1, \dots, X_{n+1})^T$ 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 $k = 1, \dots, n$, 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. \square

Pagina intenzionalmente vuota