

Politecnico di Milano
Appunti di calcolo delle probabilità¹

Ilenia Epifani
Lucia Ladelli
Gustavo Posta

Anno accademico 2019/20

¹© I diritti d'autore sono riservati. Ogni sfruttamento commerciale non autorizzato sarà perseguito.

Indice

1	Probabilità	1
1.1	Introduzione	1
1.2	Spazi di probabilità	3
1.2.1	Spazio campionario	3
1.2.2	Eventi	4
1.2.3	Spazio di probabilità	7
1.3	Proprietà della probabilità	8
1.4	Spazi finiti o numerabili	11
1.5	Probabilità condizionata ed indipendenza	16
1.5.1	Alcune formule importanti	17
1.5.2	Indipendenza	21
1.5.3	Prove di Bernoulli	24
2	Variabili aleatorie	27
2.1	Variabili aleatorie	27
2.1.1	Funzione di ripartizione	29
2.2	Variabili aleatorie discrete	32
2.3	Esempi di densità discrete notevoli	36
2.3.1	Densità binomiale e bernoulliana	37
2.3.2	Densità Geometrica	39
2.3.3	Densità di Poisson come limite di densità binomiale	40
2.3.4	Densità ipergeometrica	42
2.4	Variabili aleatorie assolutamente continue	45
2.5	Esempi di densità continue notevoli	48
2.5.1	Densità uniforme continua	48
2.5.2	Densità esponenziale	49
2.5.3	Densità gaussiana standard	50
2.6	Funzioni di variabili aleatorie	53
2.6.1	*Cenno alla simulazione di variabili aleatorie	58
3	Media varianza e momenti	59
3.1	Valore atteso (o media)	59
3.1.1	Valore atteso di funzioni di variabili aleatorie	62

3.1.2	Proprietà del valore atteso	63
3.2	Varianza	64
3.2.1	Proprietà della varianza	65
3.3	Disuguaglianza di Chebychev	68
3.4	Standardizzazione di una variabile aleatoria	69
3.5	Densità gaussiana $\mathcal{N}(\mu, \sigma^2)$	69
3.6	Approssimazione gaussiana della funzione di ripartizione binomiale	71
3.7	*Momenti e funzione generatrice dei momenti	73
4	Vettori Aleatori	77
4.1	Variabili aleatorie indipendenti	77
4.2	Vettori aleatori	80
4.3	Vettori aleatori discreti	82
4.4	Vettori aleatori assolutamente continui	84
4.5	Funzioni di vettori aleatori	87
4.5.1	Funzioni di vettori aleatori discreti	88
4.5.2	Funzioni di vettori aleatori assolutamente continui	89
4.6	*Vettori aleatori indipendenti	92
4.7	Valore atteso di funzioni di vettori aleatori	93
4.8	Covarianza, Coefficiente di correlazione	95
4.8.1	Matrice di covarianza	99
4.9	*Funzione generatrice dei momenti	100
4.10	Vettori gaussiani	102
4.11	Teoremi limite per somme di variabili aleatorie	106
4.11.1	Legge dei grandi numeri	106
4.11.2	Teorema centrale del limite	108
A	Richiami di analisi matematica	A-1
A.1	Richiami di teoria degli insiemi	A-1
A.2	Alcuni limiti notevoli	A-2
A.3	Calcolo integrale	A-2
A.3.1	Proprietà dell'integrale	A-2
A.3.2	Regole di integrazione	A-2
A.3.3	Alcuni integrali immediati	A-3
A.4	Successioni e serie	A-4
B	Calcolo combinatorio	B-7
B.1	Introduzione	B-7
B.2	Disposizioni e permutazioni	B-7
B.3	Combinazioni	B-9
B.4	Esercizi	B-10

Capitolo 1

Probabilità

1.1 Introduzione

Lo scopo di questi appunti è quello di introdurre il lettore ai concetti base della teoria e del calcolo delle probabilità. Il calcolo delle probabilità si occupa dello studio e della formalizzazione matematica di fenomeni “casuali”, cioè di fenomeni per i quali non possiamo predire a priori l’esito. I motivi per i quali può accadere che per un certo fenomeno non sia possibile dare una descrizione deterministica sono molteplici: può accadere che le informazioni riguardanti il fenomeno sul quale vogliamo fare previsioni siano incomplete, può accadere che non esista una teoria che permetta di arrivare a dedurre delle conseguenze per il fenomeno in osservazione, o che magari la teoria esista ma risulti di difficile applicazione, oppure può accadere semplicemente che il fenomeno sia veramente “casuale”. Come esempio pensiamo al lancio di una moneta. Il moto di un corpo rigido nello spazio, come è la moneta, è ben descritto dalle equazioni della meccanica newtoniana, quindi in linea di principio, se riusciamo a tenere conto della velocità iniziale con la quale viene lanciata la moneta, dell’attrito effettuato dall’aria e degli urti anelastici che la moneta subisce quando ricade a terra, potremmo calcolare se alla fine la moneta esibirà sulla faccia superiore testa o croce. Tuttavia un conto reale di questo genere risulta infattibile, sia perché non è possibile in generale misurare sperimentalmente le grandezze fisiche coinvolte, sia perché il sistema in esame esibisce una dipendenza sensibile dalle condizioni iniziali: una piccola (infinitesima) variazione delle condizioni iniziali (ad esempio la forza applicata nel lancio o posizione dalla quale si lancia) porta ad un effetto macroscopico notevole (esce testa piuttosto che croce). Risulta invece chiaro che se la moneta è sufficientemente simmetrica ci attendiamo che la “possibilità” che dopo un lancio si presenti testa sia la stessa che si presenti croce. Da qui l’esigenza di modellizzare questo fenomeno attraverso una teoria diversa dalla meccanica newtoniana.

Dall’esempio precedente può sembrare che mentre una teoria deterministica come la meccanica newtoniana ci potrebbe dire, almeno in linea di principio, se alla fine osserveremo una testa o una croce, una descrizione probabilistica del fenomeno si limita a constatare che se lanciamo una moneta la “possibilità” di ottenere testa è la stessa di quella di ot-

tenere croce, non aiutandoci affatto nel fare previsioni quantitative. Questo, per quanto riguarda l'esempio precedente è almeno parzialmente vero. Per capire quali siano i punti di forza della teoria della probabilità bisogna fare un esempio più complesso. Supponiamo di rovesciare un sacchetto contenente 1000 monete da 1¢ su un tavolo e supponiamo di voler sapere quante sono le monete che esibiscono una testa sulla parte superiore. Questo è un problema totalmente intrattabile dal punto di vista della meccanica classica (lo sarebbe anche nel caso potessimo supporre le monete perfettamente identiche e gli urti perfettamente elastici). Da un punto di vista intuitivo possiamo aspettarci che circa la metà delle monete esibirà una testa mentre l'altra metà esibirà una croce. Tuttavia non sarebbe corretto affermare che osserveremo esattamente 500 teste e 500 croci. La teoria della probabilità ci fornirà invece gli strumenti per dare un significato quantitativo a frasi del tipo “circa la metà delle monete esibirà una testa mentre l'altra metà esibirà una croce”. Ad esempio vedremo che la probabilità di osservare un numero compreso tra 440 e 560 teste vale approssimativamente

$$\frac{1}{\sqrt{2\pi}} \int_{-3.82636}^{3.82636} e^{-\frac{x^2}{2}} dx \simeq 0.99987$$

che indicherà che quasi sicuramente il numero di teste che osserveremo sarà un numero compreso tra 440 e 560.

Come abbiamo detto la nostra sarà solamente una introduzione alle tecniche del calcolo delle probabilità, per questo le applicazioni che vedremo saranno sempre molto semplici e avranno scopo essenzialmente didattico. Non vedremo praticamente mai un'applicazione che risolve un vero problema tecnico-ingegneristico. Piuttosto svilupperemo le tecniche matematiche che potranno poi essere utilizzate per veri problemi applicativi in corsi più avanzati. Il taglio di questo corso sarà quindi di carattere modellistico-matematico, nel senso che il corso svilupperà delle tecniche matematiche, ma terremo sempre d'occhio cosa queste tecniche significhino da un punto di vista pratico-applicativo. Per poter apprendere le tecniche base del calcolo delle probabilità è necessaria una certa familiarità con alcuni concetti matematici elementari, come il calcolo combinatorio e il calcolo differenziale ed integrale di più variabili.

Nel testo sono contenuti anche degli esercizi. Gli esercizi sono tutti molto semplici e vanno svolti tutti, esclusi quelli segnalati da un asterisco “ * ” che sono di carattere più matematico-teorico. Cercare di studiare il testo senza tentare di confrontarsi con gli esercizi è quasi totalmente inutile: lo scopo dell'esercizio è forzare lo studente a pensare in modo non superficiale a quanto ha letto e pensa di aver capito.

Il materiale è organizzato nel modo seguente.

Nel primo capitolo vengono introdotte le nozioni base della teoria delle probabilità quali spazio campionario, eventi e spazio di probabilità; viene poi sviluppato il concetto basilare di indipendenza. Questo capitolo non contiene materiale particolarmente avanzato da un punto di vista tecnico, tuttavia contiene alcuni concetti (come quello di spazio degli eventi elementari e di famiglia di eventi) che vanno letti con attenzione.

Nel secondo capitolo vengono introdotte le variabili aleatorie monodimensionali e le caratteristiche deterministiche ad esse associate. Per comprendere questo capitolo è ne-

cessario avere una certa familiarità con il calcolo differenziale e integrale unidimensionale. Inoltre anche qui alcuni concetti elementari ma profondi come quello di preimmagine richiedono una certa attenzione.

Nel capitolo terzo vengono trattate le variabili aleatorie multidimensionali. Per poter leggere questo capitolo è necessario che il lettore conosca il calcolo integrale e differenziale a più variabili.

Nel capitolo quarto vengono discusse le leggi limite del calcolo delle probabilità; una certa conoscenza del concetto di successione di funzioni è utile anche se non necessaria.

1.2 Spazi di probabilità

In questo paragrafo introdurremo gli oggetti matematici che sono alla base del modello probabilistico assiomatico. Come in tutte le teorie assiomatiche alcune delle definizioni di base possono sembrare inizialmente astratte e prive di contenuto. Per ridurre al minimo questo inconveniente cercheremo sempre di accompagnare le definizioni con semplici esempi applicativi.

1.2.1 Spazio campionario

Supponiamo di condurre un esperimento aleatorio, cioè un esperimento di cui non possiamo prevedere a priori il risultato, e supponiamo che ogni possibile risultato dell'esperimento possa essere identificato con un elemento ω di un certo insieme Ω . L'insieme Ω viene detto *spazio campionario* o *spazio dei campioni* o *spazio degli eventi elementari* relativo all'esperimento, gli elementi (o punti) di Ω si chiamano *eventi elementari*.

Esempio 1.2.1 Si consideri l'esperimento aleatorio: “Giuseppe lancia un dado ed osserva il numero che compare sulla faccia superiore”. I possibili risultati di questo esperimento sono sei: “Giuseppe osserva un uno”, “Giuseppe osserva un due”, ..., “Giuseppe osserva un sei”; sembra allora corretto considerare uno spazio campionario $\Omega = \{\omega_1, \omega_2, \dots, \omega_6\}$ costituito da 6 punti, dove ω_1 è associato all'evento “Giuseppe osserva un uno”, ω_2 è associato all'evento “Giuseppe osserva un due” etc. Ovviamente i punti ω_k possono essere scelti in modo arbitrario, ad esempio si può porre $\omega_1 := a$, $\omega_2 := b$, ..., $\omega_6 := f$. Però risulta più chiaro porre $\Omega := \{1, 2, \dots, 6\}$.

Esempio 1.2.2 Si consideri l'esperimento aleatorio che consiste nell'osservare lo stato di un interruttore in un circuito elettrico. Questo esperimento ha solo due possibili risultati: il circuito è aperto oppure è chiuso. Uno spazio campionario ragionevole può essere $\Omega := \{0, 1\}$ dove 0 significa circuito aperto mentre 1 significa circuito chiuso.

Esempio 1.2.3 Si consideri l'esperimento aleatorio consistente nel lanciare una moneta equilibrata fino a quando non si presenta testa. Il risultato dell'esperimento casuale può essere un qualunque numero naturale $1, 2, \dots$ che indica il lancio in cui compare per la

prima volta testa. Ma testa potrebbe non presentarsi mai; quindi per spazio campionario si può scegliere $\Omega = \mathbb{N} \cup \{\infty\}$.

Esempio 1.2.4 Si consideri l'esperimento aleatorio che consiste nell'osservare il tempo in secondi che intercorre tra l'inizio del funzionamento di un componente di un circuito ed il suo primo guasto (*tempo di vita del componente*). Il risultato dell'esperimento casuale può essere un qualsiasi numero reale non negativo; pertanto, per spazio campionario si può scegliere $\Omega := [0, +\infty) =: \mathbb{R}_+$.

Esempio 1.2.5 Si consideri l'esperimento aleatorio: “Giuseppe lancia due dadi, uno rosso l'altro blu, ed osserva i numeri che compaiono sulle facce superiori”. In questo caso i risultati possibili sono tutte le coppie ordinate di numeri interi tra uno e sei. Uno spazio degli eventi elementari è

$$\begin{aligned}\Omega &:= \{(1, 1), (1, 2), \dots, (1, 6), (2, 1), (2, 2), \dots, (2, 6), \dots, (6, 1), (6, 2), \dots, (6, 6)\} = \\ &= \{(i, j) : i = 1, 2, \dots, 6; j = 1, 2, \dots, 6\}\end{aligned}$$

dove il generico evento elementare (i, j) in Ω rappresenta il risultato “è uscito i sul dado rosso e j sul dado blu”.

Esercizio 1.2.6 Tre palline sono estratte una dopo l'altra senza reimbussolamento da un'urna che ne contiene dieci numerate da 1 a 10 e per il resto identiche. Trovare lo spazio campionario.

1.2.2 Eventi

Abbiamo detto che lo spazio campionario Ω è un insieme che rappresenta tutti i possibili esiti di un dato esperimento aleatorio. Torniamo ora all'Esempio 1.2.1 dove $\Omega = \{1, 2, \dots, 6\}$; ciascun punto di Ω rappresenta il numero che Giuseppe osserva sulla faccia superiore del dado che ha lanciato. Ci piacerebbe ora poter rappresentare eventi del tipo “Giuseppe osserva un numero pari”, oppure “Giuseppe osserva un numero più grande di 4” etc. Questi sono sempre eventi relativi all'esperimento aleatorio ma non sono più elementari, nel senso che, ad esempio, l'evento “Giuseppe osserva un numero pari” può essere descritto in termini di eventi elementari nel modo seguente: “Giuseppe osserva un 2” oppure “Giuseppe osserva un 4” oppure “Giuseppe osserva un 6”. La scelta che si opera nel calcolo delle probabilità è quella di rappresentare gli eventi relativi ad un esperimento aleatorio mediante *sottoinsiemi* dello spazio campionario Ω . In questo modo ad esempio l'evento “Giuseppe osserva un numero pari” è rappresentato dal sottoinsieme $\{2, 4, 6\} \subset \Omega$ mentre l'evento “Giuseppe osserva un numero più grande di 4” è rappresentato dal sottoinsieme $\{5, 6\} \subset \Omega$. Segue che gli eventi elementari vengono rappresentati da insiemi contenenti un solo elemento: l'evento “Giuseppe osserva un 2” è rappresentato dall'insieme $\{2\} \subset \Omega$.

Esercizio 1.2.7 Relativamente all'Esempio 1.2.4 rappresentare come sottoinsiemi di $\Omega = \mathbb{R}_+$ i seguenti eventi

1. il componente si rompe esattamente dopo 2 secondi;
2. il componente dura più di 2 secondi;
3. il componente non si rompe mai.

Esercizio 1.2.8 Relativamente all'Esempio 1.2.5 rappresentare come sottoinsiemi di $\Omega := \{(i, j) : i = 1, 2, \dots, 6; j = 1, 2, \dots, 6\}$ i seguenti eventi

1. i due dadi presentano lo stesso valore;
2. il dado rosso presenta un valore più grande del dado blu;
3. la somma dei due dadi è 7.

Si può osservare che agli operatori logici “o”, “e” e “non”, attraverso la corrispondenza tra eventi ed insiemi, corrispondono operazioni sugli insiemi. Ad esempio, precedentemente, abbiamo descritto l'evento “Giuseppe osserva un numero pari” in termini di eventi elementari come: “Giuseppe osserva un 2” oppure “Giuseppe osserva un 4” oppure “Giuseppe osserva un 6”; questa decomposizione corrisponde alla seguente ovvia relazione insiemistica $\{2, 4, 6\} = \{2\} \cup \{4\} \cup \{6\}$, cioè l'operatore logico “o” corrisponde all'unione insiemistica “ \cup ”¹. Analogamente l'operatore logico “e” corrisponde all'intersezione insiemistica “ \cap ”: “Giuseppe osserva un numero pari e più grande di 4” corrisponde al sottoinsieme $\{6\} = \{2, 4, 6\} \cap \{5, 6\}$. L'operatore logico “non” corrisponde al complementare insiemistico: “Giuseppe non osserva un numero pari” corrisponde al sottoinsieme $\{1, 3, 5\} = \Omega \setminus \{2, 4, 6\} = \{2, 4, 6\}^c$. Abbiamo quindi che gli eventi relativi ad un esperimento aleatorio possono essere rappresentati da sottoinsiemi dello spazio campionario e quindi costituiscono una *famiglia* o *collezione* di sottoinsiemi di Ω che indicheremo con \mathcal{F} (questo significa che se $E \in \mathcal{F}$ allora $E \subset \Omega$). Inoltre, diremo che si è verificato un evento E , se il risultato dell'esperimento aleatorio è $\omega \in E$.

* Proprietà di chiusura della famiglia di eventi \mathcal{F}

Sia Ω lo spazio campionario relativo ad un esperimento aleatorio e \mathcal{F} una collezione di eventi relativi ad esso. Ci domandiamo: da quali sottoinsiemi deve essere costituita \mathcal{F} ? Sembra piuttosto ragionevole richiedere, ad esempio, che se reputiamo E un evento, quindi se siamo in grado di dire se E si è verificato, siamo anche in grado di dire se E non si è verificato, cioè se si è verificato E^c . Pertanto sembra ragionevole supporre che se $E \in \mathcal{F}$ allora $E^c \in \mathcal{F}$. Analogamente, se E ed F sono eventi, se cioè sappiamo dire se E ed F si sono verificati, sappiamo anche dire se l'evento “ E o F ” si è verificato. Ne segue che se $E, F \in \mathcal{F}$ allora $E \cup F \in \mathcal{F}$. È inoltre ragionevole che l'evento certo Ω , cioè l'evento che si verifica sicuramente, appartenga a \mathcal{F} .

Una famiglia di insiemi che soddisfa alle precedenti proprietà viene chiamata *algebra di sottoinsiemi*:

Definizione 1.2.9 Sia Ω un insieme ed \mathcal{F} una famiglia di sottoinsiemi di Ω . \mathcal{F} è un'algebra di sottoinsiemi di Ω se soddisfa alle seguenti proprietà:

¹In generale useremo l'operatore logico “o” in modo *inclusivo*, cioè l'evento “ A oppure B ” si verifica se si verifica A ma non B oppure si verifica B ma non A oppure si verificano sia A che B

1. $\Omega \in \mathcal{F}$;
2. $E \in \mathcal{F} \Rightarrow E^c := \Omega \setminus E \in \mathcal{F}$;
3. $E, F \in \mathcal{F} \Rightarrow E \cup F \in \mathcal{F}$.

Esercizio 1.2.10 Sia Ω un insieme qualsiasi, verificare che l'algebra banale $\mathcal{F}_1 := \{\emptyset, \Omega\}$ e l'insieme delle parti $\mathcal{F}_2 := \mathcal{P}(\Omega) = \{\text{tutti i sottoinsiemi di } \Omega\}$ sono algebre di sottoinsiemi di Ω .

Esercizio 1.2.11 Verificare che se \mathcal{F} è un'algebra di sottoinsiemi di Ω allora:

1. $\emptyset \in \mathcal{F}$;
2. $E, F \in \mathcal{F} \Rightarrow E \cap F \in \mathcal{F}$;
3. $E_1, E_2, \dots, E_n \in \mathcal{F} \Rightarrow \bigcup_{k=1}^n E_k \in \mathcal{F}$;
4. $E_1, E_2, \dots, E_n \in \mathcal{F} \Rightarrow \bigcap_{k=1}^n E_k \in \mathcal{F}$.

Se Ω è finito gli assiomi della Definizione 1.2.9, e in particolare l'assioma 3. (e la sua conseguenza naturale data da 3. dell'Esercizio 1.2.11) sono adeguati. Tuttavia, se Ω non è finito, essi non bastano per la teoria che vogliamo costruire. Si consideri, a tal fine, l'esperimento descritto nell'Esempio 1.2.3 e supponiamo di aver costruito la nostra algebra di eventi \mathcal{F} . Sia E_k l'evento “esce testa al k -esimo lancio” e supponiamo che $E_k \in \mathcal{F}$ per ogni $k = 1, 2, \dots$. Sembrerebbe naturale supporre che l'evento E “prima o poi esce testa” sia in \mathcal{F} . Notiamo che E può essere descritto come “esce testa al primo lancio, oppure al secondo, oppure al terzo, ...”. Questo significa che $E = \bigcup_{k=1}^{+\infty} E_k$. Ma se \mathcal{F} è semplicemente un'algebra, il fatto che $E_k \in \mathcal{F}$ per ogni $k = 1, 2, \dots$ non implica che $E = \bigcup_{k=1}^{+\infty} E_k \in \mathcal{F}$. Quindi E non è un evento che viene considerato dal nostro modello, il che sembra piuttosto deludente. Per ovviare a questa situazione si introduce una nozione più restrittiva di quella di algebra di insiemi:

Definizione 1.2.12 Sia Ω un insieme ed \mathcal{F} una famiglia di sottoinsiemi di Ω . \mathcal{F} è una σ -algebra² di sottoinsiemi di Ω se soddisfa alle seguenti proprietà:

1. $\Omega \in \mathcal{F}$;
2. $E \in \mathcal{F} \Rightarrow E^c := \Omega \setminus E \in \mathcal{F}$;
3. $E_1, E_2, \dots \in \mathcal{F} \Rightarrow \bigcup_{k=1}^{+\infty} E_k \in \mathcal{F}$.

Esercizio 1.2.13 Verificare che una σ -algebra di sottoinsiemi è anche un'algebra di insiemi di Ω .

Esercizio 1.2.14 Risolvere l'Esercizio 1.2.10 sostituendo alla parola “algebra” la parola “ σ -algebra”.

Esercizio 1.2.15 Verificare che se \mathcal{F} è una σ -algebra di sottoinsiemi di Ω allora:

1. $\emptyset \in \mathcal{F}$;
2. $E, F \in \mathcal{F} \Rightarrow E \cap F \in \mathcal{F}$;
3. $E_1, E_2, \dots \in \mathcal{F} \Rightarrow \bigcap_{k=1}^{+\infty} E_k \in \mathcal{F}$.

²Si legge “sigma algebra”

1.2.3 Spazio di probabilità

Abbiamo visto che a un esperimento aleatorio è associata una coppia (Ω, \mathcal{F}) in cui Ω è lo spazio campionario ed \mathcal{F} è una famiglia (σ -algebra) di sottoinsiemi di Ω rappresentanti i possibili eventi relativi all'esperimento. Questa coppia viene talvolta chiamata *spazio probabilizzabile*. Ora, l'unica cosa che manca alla nostra teoria è l'ingrediente fondamentale, cioè la probabilità. Quello che vogliamo è poter dire che la probabilità di un evento è uguale ad un numero. Quindi per noi la probabilità sarà una funzione che ad ogni evento $E \in \mathcal{F}$ associa un numero $P(E)$. Diamo ora la definizione di *probabilità* e di *spazio di probabilità*.

Definizione 1.2.16 *Sia (Ω, \mathcal{F}) uno spazio probabilizzabile. Una probabilità su (Ω, \mathcal{F}) è una funzione su \mathcal{F} tale che:*

1. $P(E) \geq 0$ per ogni $E \in \mathcal{F}$;
2. $P(\Omega) = 1$;
3. se $E_1, E_2, \dots \in \mathcal{F}$ sono eventi a due a due disgiunti, cioè $E_h \cap E_k = \emptyset$ se $h \neq k$, allora $P\left(\bigcup_{k=1}^{+\infty} E_k\right) = \sum_{k=1}^{+\infty} P(E_k)$.

La terna (Ω, \mathcal{F}, P) viene detta spazio di probabilità.³

Gli assiomi che definiscono la probabilità sono assolutamente naturali. L'assioma 1. ci dice che la probabilità associa ad ogni evento un numero non negativo che interpretiamo come la sua probabilità di accadere. Scopriremo che l'assioma 2. ci dice semplicemente che attribuiamo all'evento certo Ω (cioè l'evento che si verifica sicuramente) il valore massimo che può assumere la probabilità. Infine, l'assioma 3. esprime il fatto che data una successione di eventi E_1, E_2, \dots incompatibili, o "mutuamente escludentesi" (ossia, gli eventi E_1, E_2, \dots non possono verificarsi simultaneamente), allora la probabilità dell'evento "almeno uno degli eventi E_1, E_2, \dots si verifica" è dato dalla somma delle singole probabilità degli eventi E_1, E_2, \dots . Questo assioma prende il nome di σ -additività o *additività completa*.

Una immediata conseguenza degli assiomi sono le seguenti proprietà della probabilità.

Proposizione 1.2.17 *Sia (Ω, \mathcal{F}, P) uno spazio di probabilità. Allora:*

1. $P(\emptyset) = 0$ (probabilità dell'evento impossibile);
2. se $E_1, E_2, \dots, E_n \in \mathcal{F}$, $E_h \cap E_k = \emptyset$ se $h \neq k$, allora $P\left(\bigcup_{k=1}^n E_k\right) = \sum_{k=1}^n P(E_k)$ (additività finita).

Dimostrazione

³Questa formulazione matematica è detta impostazione assiomatica della probabilità ed è dovuta al matematico sovietico A.N. Kolmogorov (1933)

1. Se $E_k := \emptyset$ per $k = 1, 2, \dots$, allora E_1, E_2, \dots è una successione di eventi disgiunti a coppie e $\bigcup_{k=1}^{+\infty} E_k = \emptyset$. Per l'assioma 3. della Definizione 1.2.16 si ha:

$$P(\emptyset) = P\left(\bigcup_{k=1}^{+\infty} E_k\right) = \sum_{k=1}^{+\infty} P(\emptyset)$$

che è verificata solo se $P(\emptyset) = 0$.

2. Se $E_k := \emptyset$ per $k = n+1, n+2, \dots$, allora E_1, E_2, \dots è una successione di eventi disgiunti a coppie (verificare!) e $\bigcup_{k=1}^{+\infty} E_k = \bigcup_{k=1}^n E_k$. Per l'assioma 3. della Definizione 1.2.16 si ha:

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

poiché $P(E_k) = P(\emptyset) = 0$ per $k = n+1, n+2, \dots$ ■

L'assioma 3 della Definizione 1.2.16 è equivalente al punto 2. della Proposizione 1.2.17 se lo spazio Ω è finito.

Esercizio 1.2.18 Perché?

Esempio 1.2.19 Se lanciamo tre monete distinguibili e non truccate, lo spazio campionario è

$$\Omega := \{TTT, TTC, TCT, TCC, CTT, CTC, CCT, CCC\}$$

e come famiglia di eventi possiamo scegliere $\mathcal{F} := \mathcal{P}(\Omega)$. Infine, scelta la funzione $P(E) := |E|/|\Omega|$, dove $|E|$ indica la cardinalità di E , si può verificare direttamente che con questa definizione (Ω, \mathcal{F}, P) costituisce uno spazio di probabilità.

La maggior generalità dell'assioma 3 è necessaria nel caso di spazi campionari infiniti.

Esempio 1.2.20 Consideriamo l'esperimento descritto nell'Esempio 1.2.3 e sia E_k l'evento “esce testa per la prima volta al k -esimo lancio”. Gli E_k , $k = 1, 2, \dots$, sono a due a due incompatibili (cioè hanno intersezione vuota) e l'evento E “prima o poi esce testa” è quindi l'unione disgiunta $E = \bigcup_{k=1}^{+\infty} E_k$. Segue dall'assioma 3. che $P(E) = \sum_{k=1}^{+\infty} P(E_k)$. Vedremo in seguito che, se la moneta non è truccata, si assume $P(E_k) = 1/2^k$ e quindi $P(E) = 1$.

1.3 Proprietà della probabilità

Vediamo altre proprietà che seguono direttamente dagli assiomi della Definizione 1.2.16.

Proposizione 1.3.1 Sia (Ω, \mathcal{F}, P) uno spazio di probabilità. Allora:

1. se $E \in \mathcal{F}$ allora $P(E^c) = 1 - P(E)$ (probabilità del complementare);
2. se $E \in \mathcal{F}$ allora $P(E) \leq 1$;

3. se $E, F \in \mathcal{F}$ e $F \subset E$ allora $P(E \setminus F) = P(E) - P(F)$;
4. se $E, F \in \mathcal{F}$ e $F \subset E$ allora $P(F) \leq P(E)$ (monotonia);
5. se $E, F \in \mathcal{F}$ allora $P(E \cup F) = P(E) + P(F) - P(E \cap F)$ (probabilità dell'unione).

Dimostrazione

1. Notiamo che $\Omega = E \cup E^c$ e $E \cap E^c = \emptyset$; quindi per l'assioma 2. della Definizione 1.2.16 e il punto 2. della Proposizione 1.2.17 vale $1 = P(\Omega) = P(E) + P(E^c)$ che implica $P(E^c) = 1 - P(E)$.

2. Per il punto precedente $P(E) = 1 - P(E^c)$, ma $P(E^c) \geq 0$ per l'assioma 1. della Definizione 1.2.16; segue che necessariamente $P(E) \leq 1$.

3. Se $F \subset E$ allora $E = (E \setminus F) \cup F$ e l'unione è disgiunta; applicando il punto 2. della Proposizione 1.2.17: $P(E) = P(E \setminus F) + P(F)$ e quindi $P(E \setminus F) = P(E) - P(F)$.

4. Per il punto precedente $P(E) - P(F) = P(E \setminus F)$ che è non negativo per l'assioma 1. della Definizione 1.2.16.

5. Possiamo scrivere $E \cup F = (E \cap F^c) \cup (E \cap F) \cup (E^c \cap F)$ e l'unione è disgiunta; sempre il punto 2. della Proposizione 1.2.17 implica $P(E \cup F) = P(E \cap F^c) + P(E \cap F) + P(E^c \cap F)$; quindi $P(E \cup F) + P(E \cap F) = P(E \cap F^c) + P(E \cap F) + P(E \cap F) + P(E^c \cap F)$. Ma $P(E \cap F^c) + P(E \cap F) = P(E)$ e $P(E \cap F) + P(E^c \cap F) = P(F)$ (verificare!) e quindi $P(E \cup F) + P(E \cap F) = P(E) + P(F)$. ■

Applicando due volte la proprietà 5. della Proposizione 1.3.1, possiamo calcolare la probabilità dell'unione di tre eventi $E, F, G \in \mathcal{F}$:

$$\begin{aligned} P(E \cup F \cup G) &= P((E \cup F) \cup G) \\ &= [P(E) + P(F) + P(G)] - [P(E \cap F) + P(E \cap G) + P(F \cap G)] + P(E \cap F \cap G) \end{aligned}$$

Una generalizzazione della precedente formula è la seguente proposizione.

Proposizione* 1.3.2 (Principio di inclusione-esclusione di Poincaré) Sia (Ω, \mathcal{F}, P) uno spazio di probabilità ed $E_1, E_2, \dots, E_n \in \mathcal{F}$ eventi. Allora

$$\begin{aligned} P\left(\bigcup_{k=1}^n E_k\right) &= \sum_{r=1}^n (-1)^{r+1} \sum_{\substack{k_1, k_2, \dots, k_r=1 \\ k_1 < k_2 < \dots < k_r}}^n P(E_{k_1} \cap E_{k_2} \cap \dots \cap E_{k_r}) = \\ &= \sum_{r=1}^n (-1)^{r+1} \sum_{\{k_1, k_2, \dots, k_r\} \subset \{1, 2, \dots, n\}} P(E_{k_1} \cap E_{k_2} \cap \dots \cap E_{k_r}) \quad (1.3.1) \end{aligned}$$

Dimostrazione La dimostrazione è per induzione. La (1.3.1) è vera per $n = 2$ per il punto 5. della Proposizione 1.3.1. Supponiamo ora che (1.3.1) sia verificata per tutti gli interi $\leq n$ e per ogni famiglia di

n eventi in \mathcal{F} e proviamola per $n + 1$. Dall'ipotesi induttiva deriva:

$$\begin{aligned}
 P\left(\bigcup_{k=1}^{n+1} E_k\right) &= P\left(\left(\bigcup_{k=1}^n E_k\right) \cup E_{n+1}\right) = P\left(\bigcup_{k=1}^n E_k\right) + P(E_{n+1}) - P\left(\bigcup_{k=1}^n (E_k \cap E_{n+1})\right) = \\
 &= \sum_{r=1}^n (-1)^{r+1} \sum_{\{k_1, k_2, \dots, k_r\} \subset \{1, 2, \dots, n\}} P(E_{k_1} \cap E_{k_2} \cap \dots \cap E_{k_r}) + \\
 &+ P(E_{n+1}) - \sum_{r=1}^n (-1)^{r+1} \sum_{\{k_1, k_2, \dots, k_r\} \subset \{1, 2, \dots, n\}} P(E_{k_1} \cap E_{k_2} \cap \dots \cap E_{k_r} \cap E_{n+1}) = \\
 &= \sum_{r=1}^{n+1} (-1)^{r+1} \sum_{\substack{\{k_1, k_2, \dots, k_r\} \subset \{1, 2, \dots, n+1\} \\ \{k_1, k_2, \dots, k_r\} \not\subset \{1, 2, \dots, n\}}} P(E_{k_1} \cap E_{k_2} \cap \dots \cap E_{k_r}) + \\
 &+ \sum_{r=1}^{n+1} (-1)^{r+1} \sum_{\substack{\{k_1, k_2, \dots, k_r\} \subset \{1, 2, \dots, n+1\} \\ \{k_1, k_2, \dots, k_r\} \ni (n+1)}} P(E_{k_1} \cap E_{k_2} \cap \dots \cap E_{k_r}) = \\
 &= \sum_{r=1}^{n+1} (-1)^{r+1} \sum_{\{k_1, k_2, \dots, k_r\} \subset \{1, 2, \dots, n+1\}} P(E_{k_1} \cap E_{k_2} \cap \dots \cap E_{k_r}). \blacksquare
 \end{aligned}$$

Esercizio 1.3.3 Relativamente alla prima sessione d'esame del primo anno del corso di laurea XXX è noto che la probabilità che uno studente superi:

- l'esame A è 0.4,
- l'esame B è 0.5,
- l'esame C è 0.3,
- l'esame A e l'esame B è 0.35,
- l'esame A e l'esame C è 0.2,
- l'esame B e l'esame C è 0.25,
- tutti e tre esami è 0.15.

Determinare la probabilità che nella prima sessione uno studente scelto a caso

1. non superi l'esame A;
2. superi A ma non superi B;
3. superi almeno un esame;
4. non superi alcun esame.

Soluzione Indichiamo con A l'evento "lo studente supera l'esame A", con B l'evento "lo studente supera l'esame B" e con C l'evento "lo studente supera l'esame C". Allora le probabilità richieste sono:

1. $P(A^c) = 1 - P(A) = 0.6$;
2. $P(A \cap B^c) = P(A \setminus (A \cap B)) = P(A) - P(A \cap B) = 0.4 - 0.35 = 0.05$;
3. $P(A \cup B \cup C) = P(A) + P(B) + P(C) - [P(A \cap B) + P(A \cap C) + P(B \cap C)] + P(A \cap B \cap C) = 0.4 + 0.5 + 0.3 - 0.35 - 0.2 - 0.25 + 0.15 = 0.55$;

$$4. P(A^c \cap B^c \cap C^c) = P((A \cup B \cup C)^c) = 1 - 0.55 = 0.45. \blacksquare$$

1.4 Spazi finiti o numerabili

In questo paragrafo vedremo come probabilizzare uno spazio campionario finito o numerabile, cioè come costruire modelli probabilistici per esperimenti aleatori che hanno al più una infinità numerabile di esiti possibili.

Fissiamo inizialmente l'attenzione sul caso Ω numerabile e sia $\{\omega_1, \omega_2, \dots\}$ una numerazione dei punti di Ω . In generale, in questo caso, si sceglie come σ -algebra \mathcal{F} l'insieme di tutti i sottoinsiemi di Ω , $\mathcal{P}(\Omega)$. Si definisce una probabilità su (Ω, \mathcal{F}) assegnando una successione p_1, p_2, \dots tale che

$$p_k \geq 0 \text{ per ogni } k = 1, 2, \dots$$

e

$$\sum_{k=1}^{\infty} p_k = 1 \tag{1.4.1}$$

Infatti se attribuiamo agli eventi elementari le probabilità $P(\{\omega_1\}) = p_1, P(\{\omega_2\}) = p_2, \dots$, allora la probabilità di ogni evento $E \in \mathcal{F}$ risulta automaticamente individuata come segue. Per ogni evento $E \in \mathcal{F}$ possiamo scrivere $E = \bigcup_{k: \omega_k \in E} \{\omega_k\}$ e l'unione è disgiunta, quindi per la proprietà di σ -additività di cui deve godere una probabilità definiamo (necessariamente)

$$P(E) = \sum_{k: \omega_k \in E} P(\{\omega_k\}) = \sum_{k: \omega_k \in E} p_k \tag{1.4.2}$$

È immediato verificare che la P così definita è una probabilità su $\mathcal{P}(\Omega)$. Infatti $P(\emptyset) = 0$ e $P(\Omega) = \sum_{k=1}^{+\infty} p_k = 1$. Inoltre la proprietà di σ -additività segue dalla Definizione 1.4.2 e dal fatto che, poiché $\sum_{k=1}^{+\infty} p_k$ è una serie a termini positivi convergente, allora si possono sommare somme parziali disgiunte ed ottenere sempre il medesimo risultato come somma totale.

Viceversa, una qualunque misura di probabilità su $\mathcal{P}(\Omega)$ soddisfa $P(\{\omega_k\}) \geq 0$ per $k = 1, 2, \dots$ e $\sum_{k=1}^{+\infty} P(\{\omega_k\}) = P(\Omega) = 1$. Abbiamo dimostrato la seguente proposizione.

Proposizione 1.4.1 *Sia Ω un insieme numerabile e sia $\{\omega_1, \omega_2, \dots\}$ una numerazione dei punti di Ω . Sia $\mathcal{F} = \mathcal{P}(\Omega)$.*

1. *Ogni probabilità su (Ω, \mathcal{F}) individua una successione di numeri reali p_1, p_2, \dots che soddisfano (1.4.1) ponendo $P(\{\omega_k\}) = p_k$ per ogni k .*
2. *Data una successione p_1, p_2, \dots che soddisfa (1.4.1), esiste un'unica misura di probabilità su (Ω, \mathcal{F}) tale che $P(\{\omega_k\}) = p_k$ per ogni k . Tale probabilità è data da*

$$P(E) = \sum_{k: \omega_k \in E} p_k \quad \forall E \subset \Omega$$

Notiamo che quanto detto sopra per spazi numerabile può essere ripetuto per Ω finito.

Esercizio 1.4.2 Enunciare e dimostrare la proposizione precedente nel caso di spazi campionari finiti.

Esempio 1.4.3 Ogni successione [sequenza] di termini positivi per la quale la somma dei termini è uno fornisce un esempio di modello probabilistico su uno spazio numerabile [finito]. Tuttavia alcune di queste si impongono come modelli naturali per certi tipi di fenomeni aleatori. Ricordiamo qui i principali modelli utili nelle applicazioni. Una trattazione più approfondita viene rimandata al capitolo dedicato alle variabili aleatorie.

1. *Modello di Poisson.* In questo modello la probabilità, dipendente da un parametro positivo λ , è definita su $\Omega = \{0, 1, 2, \dots\}$ dalla successione

$$p_k = \frac{e^{-\lambda} \lambda^k}{k!} \quad k = 0, 1, \dots$$

2. *Modello geometrico.* In questo modello la probabilità, dipendente da un parametro p con $0 < p < 1$, è definita su $\Omega = \{1, 2, \dots\}$ dalla successione

$$p_k = p(1-p)^{k-1} \quad k = 1, 2, \dots$$

3. *Modello binomiale.* In questo modello la probabilità, dipendente da due parametri n intero positivo e p con $0 < p < 1$, è definita su $\Omega = \{0, 1, \dots, n\}$ dalla sequenza

$$p_k = \binom{n}{k} p^k (1-p)^{n-k} \quad k = 0, 1, \dots, n$$

Esercizio 1.4.4 Verificare che i p_k assegnati nei punti 1., 2. e 3. dell'Esempio 1.4.3 verificano (1.4.1) e quindi definiscono una probabilità.

Consideriamo ora un esperimento aleatorio che ammette solo un numero finito n di risultati possibili, sia $\Omega = \{\omega_1, \omega_2, \dots, \omega_n\}$ lo spazio campionario associato e $\mathcal{F} = \mathcal{P}(\Omega)$. Supponiamo che la natura dell'esperimento aleatorio ci suggerisca di assumere $p_1 = p_2 = \dots = p_n = p$, cioè di assegnare la stessa probabilità ad ogni evento elementare. In questo caso si parla di *spazio di probabilità uniforme* oppure *spazio equiprobabile finito*. Dall'assioma 2. della Definizione 1.2.16 e dalla proprietà di additività finita (cfr. Proposizione 1.2.17) segue che

$$1 = P(\Omega) = \sum_{k=1}^n P(\{\omega_k\}) = \sum_{k=1}^n p = np = |\Omega|p \implies p = \frac{1}{|\Omega|}$$

e la probabilità di ogni evento $E \in \mathcal{F}$ è data da

$$P(E) = \sum_{k: \omega_k \in E} P(\{\omega_k\}) = \sum_{k: \omega_k \in E} \frac{1}{|\Omega|} = \frac{|E|}{|\Omega|}$$

Esempio 1.4.5 (segue Esempio 1.2.5) Consideriamo ancora l'esempio del lancio di due dadi. In questo caso lo spazio degli eventi elementari è $\Omega = \{(i, j) : i, j = 1, 2, \dots, 6\}$ e come famiglia (σ -algebra) degli eventi possiamo scegliere $\mathcal{F} := \mathcal{P}(\Omega)$. Per quanto riguarda l'assegnazione di una probabilità P su (Ω, \mathcal{F}) osserviamo che se assumiamo che i due dadi non siano truccati e vogliamo che il nostro spazio di probabilità (Ω, \mathcal{F}, P) modellizzi questo fatto fisico, dobbiamo ammettere che tutti gli eventi elementari di Ω abbiano la stessa probabilità $p = 1/|\Omega| = 1/36$. Sia E_k l'evento "la somma dei due dadi è k " per $k = 2, 3, \dots, 12$. Allora,

$$\begin{aligned} E_2 &= \{(1, 1)\} \\ E_3 &= \{(1, 2), (2, 1)\} \\ E_4 &= \{(1, 3), (2, 2), (3, 1)\} \\ E_5 &= \{(1, 4), (2, 3), (3, 2), (4, 1)\} \\ E_6 &= \{(1, 5), (2, 4), (3, 3), (4, 2), (5, 1)\} \\ E_7 &= \{(1, 6), (2, 5), (3, 4), (4, 3), (5, 2), (6, 1)\} \\ E_8 &= \{(2, 6), (3, 5), (4, 4), (5, 3), (6, 2)\} \\ E_9 &= \{(3, 6), (4, 5), (5, 4), (6, 3)\} \\ E_{10} &= \{(4, 6), (5, 5), (6, 4)\} \\ E_{11} &= \{(5, 6), (6, 5)\} \\ E_{12} &= \{(6, 6)\} \end{aligned}$$

Applicando la formula $P(E) = |E|/|\Omega|$ otteniamo: $P(E_2) = P(E_{12}) = 1/36$, $P(E_3) = P(E_{11}) = 1/18$, $P(E_4) = P(E_{10}) = 1/12$, $P(E_5) = P(E_9) = 1/9$, $P(E_6) = P(E_8) = 5/36$, $P(E_7) = 1/6$.

Esempio 1.4.6 Consideriamo l'esempio del lancio di due dadi, ma assumiamo di essere interessati solamente alla somma dei risultati dei due dadi. In questo caso lo spazio degli eventi elementari è dato da $\Omega = \{2, 3, \dots, 12\}$ e come famiglia degli eventi possiamo scegliere $\mathcal{F} := \mathcal{P}(\Omega)$. Per quanto riguarda l'assegnazione di una probabilità P su (Ω, \mathcal{F}) osserviamo che se assumiamo che i due dadi non siano truccati, per l'esempio precedente, dobbiamo porre $P(\{2\}) = P(\{12\}) = 1/36$, $P(\{3\}) = P(\{11\}) = 1/18$, $P(\{4\}) = P(\{10\}) = 1/12$, $P(\{5\}) = P(\{9\}) = 1/9$, $P(\{6\}) = P(\{8\}) = 5/36$, $P(\{7\}) = 1/6$. Se invece assumiamo che i possibili risultati della somma dei due dadi siano equiprobabili, dobbiamo porre $P(\{k\}) = 1/11$ per ogni $k = 2, 3, \dots, 12$: lo spazio di probabilità così costruito è matematicamente corretto, ma non ha nulla a che vedere con la realtà fisica e sperimentale.

Campionamento da urne

Esempi classici di probabilità uniforme sono quelli associati agli esperimenti aleatori di *campionamento da un'urna* contenente M palline numerate da 1 a M e per il resto indistinguibili. L'esperimento consiste nell'estrarre un numero n di palline. A seconda

delle modalità secondo cui vengono effettuate le estrazioni si ottengono differenti spazi campionari.

Campionamento senza reimmissione Estraiamo una dopo l'altra $n \leq M$ palline dall'urna eliminando di volta in volta la pallina estratta (*campionamento senza reimmissione o senza rimpiazzo*). Possiamo scegliere come spazio campionario

$$\Omega_1 := \{(a_1, \dots, a_n) : a_i = 1, \dots, M \text{ e } a_i \neq a_j \forall i \neq j\}$$

dove la i -esima componente del caso elementare (a_1, \dots, a_n) rappresenta il numero della i -esima pallina estratta. Se non vi è reimmissione, la prima coordinata a_1 può essere scelta in M modi e per ciascuno di questi abbiamo $M - 1$ possibilità per scegliere a_2 ... e $M - n + 1$ per l' n -esima. Detto diversamente, lo spazio campionario è l'insieme di tutte le disposizioni senza ripetizione di ordine n delle M palline. La cardinalità di Ω_1 è $|\Omega_1| = (M)_n = M(M - 1) \cdots (M - n + 1)$.

Se $n = M$ allora $|\Omega_1| = M! =$ numero delle permutazioni (senza ripetizione) di M oggetti.

Esempio 1.4.7 Un'associazione è formata da 25 iscritti. Tra questi devono essere scelti un presidente ed un segretario. Quanti sono i modi possibili per ricoprire le due cariche?

Considerando che la prima carica può essere ricoperta da 25 persone diverse e che per ciascuna di queste si hanno 24 scelte possibili della seconda carica, allora

$$|\Omega_1| = |\{(a_1, a_2) : a_1, a_2 = 1, \dots, 25 \text{ e } a_1 \neq a_2\}| = 25 \times 24 = 600$$

Se gli individui vengono scelti a caso per ricoprire le cariche, qual è la probabilità che un assegnato membro dell'associazione ne ricopra una?

Sia A : "Un assegnato membro dell'associazione ricopre una carica". Per fissare le idee, e senza perdere in generalità, il membro in questione sia il numero 1. Allora, $A = \{(a_1, a_2) \in \Omega_1 : a_1 = 1 \text{ o } a_2 = 1\}$ e $|A| = |\{(a_1, a_2) \in \Omega_1 : a_1 = 1\}| + |\{(a_1, a_2) \in \Omega_1 : a_2 = 1\}| = 24 + 24$, da cui

$$P(A) = \frac{|A|}{|\Omega_1|} = \frac{48}{25 \times 24} = \frac{2}{25} = 0.08$$

Se *non interessa l'ordine* con cui le palline sono estratte, si può scegliere come spazio campionario⁴

$$\Omega_2 := \{E : E \subset \{1, \dots, M\}, |E| = n\} = \{\{a_1, \dots, a_n\} : a_i = 1, \dots, M, a_i \neq a_j \forall i \neq j\}$$

La cardinalità di Ω_2 è $|\Omega_2| = \binom{M}{n}$.

⁴ Ω_2 è l'insieme delle *combinazioni di classe n* di $\{1, \dots, M\}$, cfr. Appendice B.

Esempio 1.4.8 Se una persona gioca a poker con un mazzo di 32 carte, in quanti modi può essere servito?

Le 32 carte del mazzo sono così ripartite: quattro semi \heartsuit , \diamondsuit , \clubsuit e \spadesuit , per ognuno dei quali si hanno le 8 carte distinte: $A, K, Q, J, 10, 9, 8, 7$. Ogni mano è un insieme di 5 carte scelte dal mazzo. Allora: $\Omega_2 = \{E: E \subset \{1, \dots, 32\}, |E| = 5\}$ e il numero di mani possibili è $|\Omega_2| = \binom{32}{5} = 201376$.

Qual è la probabilità che il giocatore abbia un tris “servito”?

Sia A l’evento: “il giocatore ha un tris servito (e non un gioco migliore)”. Allora $P(A) = |A|/|\Omega_2|$. Per calcolare $|A|$ scegliamo il valore del tris (Es. tris di K) tra gli 8 disponibili, per ciascuna scelta abbiamo $\binom{4}{3}$ modi di scegliere i semi delle carte che compongono il tris

(Es. \heartsuit , \diamondsuit e \clubsuit): in totale abbiamo $8 \times \binom{4}{3}$ modi di scegliere il tris. Ora dobbiamo prendere le rimanenti 2 carte. I valori di queste carte devono necessariamente essere differenti tra di loro (altrimenti avremmo un “full”) e differenti dal valore precedentemente scelto per il tris (altrimenti avremmo un “poker”), abbiamo quindi $\binom{7}{2}$ modi di scegliere i valori delle rimanenti 2 carte⁵. Rimangono da decidere i semi delle 2 carte: per ciascuna carta abbiamo 4 semi possibili. In definitiva $|A| = 8 \times \binom{4}{3} \times \binom{7}{2} \times 4 \times 4$ e la probabilità del tris servito è

$$\frac{8 \times \binom{4}{3} \times \binom{7}{2} \times 4 \times 4}{\binom{32}{5}} = \frac{48}{899} \simeq 0.0534 \simeq 5.3\%$$

Campionamento con reimmissione e ordinato Estraiamo ora una pallina dalla solita urna, registriamo il numero della pallina e prima di procedere alla successiva estrazione rimettiamo la pallina nell’urna. Quindi ripetiamo n volte le estrazioni secondo questo schema (*campionamento con reimmissione o con rimpiazzo*). In questo caso n può essere un numero naturale qualunque. Possiamo scegliere il seguente spazio campionario:

$$\Omega_3 := \{(a_1, \dots, a_n): a_i = 1, \dots, M\}$$

Cioè lo spazio campionario è l’insieme di tutte le disposizioni con ripetizione di M elementi di ordine n e $|\Omega_3| = M^n$. Infine, assegniamo a ogni ω uguale probabilità: $P(\{\omega\}) = 1/M^n$.

Esempio 1.4.9 Quanto vale la probabilità che ciascuna delle n palline estratte sia diversa dalle altre. Detto A tale evento, è evidente che se $n > M$ allora $P(A) = 0$. Invece, se $n \leq M$ vale quanto segue:

$$P(A) = \frac{|A|}{M^n} = \frac{M(M-1) \cdots (M-n+1)}{M^n} = \frac{M!}{(M-n)!M^n}$$

⁵7 sono i valori disponibili e ne scegliamo 2 senza ripetizione e senza tenere conto dell’ordine

1.5 Probabilità condizionata ed indipendenza

In questa sezione vengono introdotti e discussi i concetti di indipendenza e probabilità condizionata. Questi sono concetti fondamentali per la teoria della probabilità, sia da un punto di vista teorico sia da un punto di vista applicativo, rivestiranno un ruolo centrale in tutto ciò che segue e traducono in termini matematici il concetto di aggiornamento della probabilità sulla base di nuove conoscenze in possesso dello sperimentatore.

Esempio 1.5.1 (segue Esempio 1.4.6) Supponiamo vengano lanciati due dadi e supponiamo che ci venga chiesto di calcolare la probabilità che la somma dei due dadi sia 12. Per l'Esempio 1.4.6 risponderemmo $1/36$. Rispondiamo ora alla stessa domanda ma sapendo che sul primo dado è uscito un 6. Questa ulteriore informazione cambia radicalmente le nostre valutazioni. Infatti, se sappiamo che sul primo dado è uscito un 6, la probabilità che la somma dei due dadi faccia 12 è uguale alla probabilità che sia uscito un 6 anche sull'altro dado, cioè $1/6$.

Questo esempio mostra la necessità di dare una definizione per situazioni in cui si vuole calcolare le probabilità di un evento E sapendo che si è verificato un altro evento F . La definizione che segue va in questa direzione.

Definizione 1.5.2 (Probabilità condizionata) Sia (Ω, \mathcal{F}, P) uno spazio di probabilità e sia $F \in \mathcal{F}$ un evento tale che $P(F) > 0$. Dato un qualsiasi evento $E \in \mathcal{F}$ si chiama probabilità condizionata di E dato F il numero

$$P(E|F) := \frac{P(E \cap F)}{P(F)}$$

Nota 1.5.3 Come abbiamo detto $P(E|F)$ va interpretata come la probabilità che si verifichi E sapendo che si è verificato F . Un errore tipico è confondere la probabilità condizionata con la probabilità dell'intersezione, cioè con la “probabilità che si verifichino sia E che F ”. Tornando all'Esempio 1.5.1 se E è l'evento “la somma dei due dadi è 12” ed F è l'evento “sul primo dado esce 6”, allora $E \cap F = E$ è l'evento “la somma dei due dadi è 12”, quindi $P(E|F) = 1/6 \neq 1/36 = P(E \cap F)$.

Esercizio 1.5.4 Quanto vale la probabilità che la somma delle facce di due dadi regolari sia 12, se si è verificato che su uno dei due dadi è uscito 6?

Soluzione Siano E =“la somma dei due dadi è 12” e G =“su uno dei due dadi esce 6”. Se calcoliamo la probabilità condizionata che si verifichi E sapendo che si è verificato G usando la nozione intuitiva di probabilità condizionata, sbagliamo. Infatti, la nozione intuitiva di probabilità condizionata ci porta a ripetere erroneamente un ragionamento analogo a prima (se sappiamo che su un dado è uscito un 6, la probabilità che la somma dei due dadi faccia 12 è uguale alla probabilità che sia uscito un 6 anche sull'altro dado) così ottenendo per $P(E|G)$ il valore $1/6$. Ma questo ragionamento è falso: applicando la formula per il calcolo

della probabilità condizionata otteniamo

$$P(E|G) = \frac{P(E \cap G)}{P(G)} = \frac{P(\{(6, 6)\})}{P(\{(1, 6), (2, 6), \dots, (6, 6), (6, 5), \dots, (6, 1)\})} = \frac{1/36}{11/36} = \frac{1}{11} < \frac{1}{6} \blacksquare$$

Esercizio 1.5.5 (Tratto da [8]) Un lotto è costituito da 25 transistor accettabili, 10 parzialmente difettosi (cioè che si rompono dopo qualche ora d'uso) e 5 difettosi (cioè che si guastano immediatamente). Un transistor viene preso a caso dal lotto. Se non si rompe subito qual è la probabilità che sia accettabile?

Soluzione In questo caso abbiamo tre eventi A “il transistor è accettabile”, B “il transistor è parzialmente difettoso”, C “il transistor è difettoso”. Ci viene chiesto di calcolare $P(A|C^c)$. Abbiamo che:

$$P(A|C^c) = \frac{P(A \cap C^c)}{P(C^c)} = \frac{P(A)}{1 - P(C)} = \frac{25/40}{35/40} = \frac{5}{7} \blacksquare$$

Esercizio 1.5.6 Sia (Ω, \mathcal{F}, P) uno spazio di probabilità e sia $F \in \mathcal{F}$ un evento tale che $P(F) > 0$. Poniamo $P_F(E) := P(E|F)$ per ogni $E \in \mathcal{F}$.

1. Verificare che $(\Omega, \mathcal{F}, P_F)$ è uno spazio di probabilità;
2. verificare che $P_F(F) = 1$;
3. verificare che se $E \in \mathcal{F}$ è P -impossibile, cioè $P(E) = 0$, allora E è P_F -impossibile, cioè $P_F(E) = 0$.

Nota 1.5.7 Dal punto 1. dell'esercizio precedente segue che $P_F = P(\cdot|F)$ gode di tutte le proprietà generali di cui godono le probabilità. Ad esempio: se $E_1, E_2, \dots, E_n \in \mathcal{F}$, con $E_h \cap E_k = \emptyset \ \forall h \neq k$, allora $P(\bigcup_{k=1}^n E_k|F) = \sum_{k=1}^n P(E_k|F)$, oppure: se $E \in \mathcal{F}$ allora $P(E^c|F) = 1 - P(E|F)$.

1.5.1 Alcune formule importanti

Riuniamo in questo paragrafo alcune formule utili nelle applicazioni che coinvolgono il concetto di probabilità condizionata.

Formula delle probabilità totali

Spesso nelle applicazioni si ha a che fare con esperimenti aleatori in cui le condizioni di preparazione dell'esperimento aleatorio sono a loro volta casuali: la *formula delle probabilità totali* è utile per calcolare probabilità di eventi relativamente a questi casi.

Esempio 1.5.8 Ci sono due urne dette “urna A” ed “urna B”. La prima contiene 1000 biglie bianche ed 1 nera mentre la seconda ne contiene 2 nere. Si lancia una moneta equa e se viene testa si pesca una biglia dall’urna A mentre se viene croce si pesca una biglia dall’urna B. Qual è la probabilità che la biglia pescata sia nera?

Un errore tipico in queste situazioni è di pensare che la probabilità di pescare una biglia nera, seguendo la procedura sopra descritta, sia la stessa che pescare una biglia nera da un’urna C in cui siano stati spostati i contenuti delle urne A e B, cioè che contiene 1000 biglie bianche e 3 biglie nere. Questo è evidentemente un errore grossolano, infatti la probabilità di pescare una biglia nera dall’urna C è di $3/1003$ cioè prossima a 0, mentre la probabilità di pescare una biglia nera seguendo la procedura di cui sopra è maggiore di $1/2$, in quanto è maggiore della probabilità di ottenere croce su una moneta equa (se si ottiene croce allora si sceglie l’urna B e quindi necessariamente si estrae una biglia nera).

La formula delle probabilità totali fornisce la risposta su come gestire situazioni di questo genere.

Proposizione 1.5.9 (Formula delle probabilità totali) Sia (Ω, \mathcal{F}, P) uno spazio di probabilità e $F_1, F_2, \dots, F_n \in \mathcal{F}$ una partizione finita di Ω , $\bigcup_{k=1}^n F_k = \Omega$ e $F_h \cap F_k = \emptyset$ se $h \neq k$, tale che $P(F_k) > 0$ per $k = 1, 2, \dots, n$. Allora per ogni evento $E \in \mathcal{F}$ si ha

$$P(E) = \sum_{k=1}^n P(E|F_k)P(F_k) \quad (1.5.1)$$

Dimostrazione Sia $E \in \mathcal{F}$, poiché $\Omega = \bigcup_{k=1}^n F_k$ ed $E \subset \Omega$, segue che $E = E \cap \Omega = \bigcup_{k=1}^n (E \cap F_k)$; inoltre, poiché $F_h \cap F_k = \emptyset$ se $h \neq k$, allora $\bigcup_{k=1}^n (E \cap F_k)$ è un’unione disgiunta e dall’additività otteniamo

$$P(E) = \sum_{k=1}^n P(E \cap F_k) = \sum_{k=1}^n P(E|F_k)P(F_k)$$

(l’ultima uguaglianza segue direttamente dalla definizione di probabilità condizionata). ■

Esempio 1.5.10 Riprendiamo l’Esempio 1.5.8. In questo caso poniamo F_1 “esce testa”, $F_2 := F_1^c$ “esce croce”, E “viene pescata una biglia nera”. F_1 ed F_2 costituiscono ovviamente una partizione di Ω . Inoltre si ha $P(F_1) = P(F_2) = 1/2$, $P(E|F_1) = 1/1001$ mentre $P(E|F_2) = 1$. Dalla formula delle probabilità totali deriva che

$$P(E) = P(E|F_1)P(F_1) + P(E|F_2)P(F_2) = \frac{1}{1001} \times \frac{1}{2} + 1 \times \frac{1}{2} \simeq 0.5$$

Nota 1.5.11 Si noti che nell’esempio precedente non abbiamo detto nulla sullo spazio di probabilità (Ω, \mathcal{F}, P) in cui tutto avviene, abbiamo solamente assunto che tale spazio esista. Inoltre per calcolare le probabilità condizionate non abbiamo utilizzato la Definizione 1.5.2, che si sarebbe rivelata inutile senza una conoscenza esplicita di (Ω, \mathcal{F}, P) , ma

abbiamo utilizzato il significato euristico di probabilità condizionata, cioè la “probabilità che venga presa una biglia nera sapendo da quale urna si sta pescando”. Questo modo di procedere, tralasciando i dettagli formali e utilizzando nozioni intuitive, è tipico del calcolo delle probabilità e verrà utilizzato ancora in seguito. Lasciamo al lettore più pignolo il compito di verificare che effettivamente esiste uno spazio (Ω, \mathcal{F}, P) in cui è possibile immergere rigorosamente la nostra discussione.

Esercizio 1.5.12 Dimostrare la formula delle probabilità totali per una partizione *numerabile* F_1, F_2, \dots di eventi.

Formula di Bayes

Torniamo ancora all'Esempio 1.5.8. Supponiamo che qualcuno, non visto da noi, abbia lanciato la moneta, abbia di conseguenza scelto l'urna ed ora ci mostri una biglia nera. Se ci viene chiesto di scommettere se sia uscito testa o croce sulla moneta, dopo qualche ragionamento quasi tutti scommetterebbero su croce. Infatti è assai improbabile che la biglia che è stata pescata provenga dall'urna A , costituita quasi interamente da biglie bianche. La *formula di Bayes* è utile in situazioni di questo tipo, in cui cioè ci viene data un'informazione a posteriori su un evento aleatorio e ci viene chiesto in che modo si sia realizzato tale evento.

Proposizione 1.5.13 (Formula di Bayes) Sia (Ω, \mathcal{F}, P) uno spazio di probabilità e $F_1, F_2, \dots, F_n \in \mathcal{F}$ una partizione finita di Ω tale che $P(F_k) > 0$ per $k = 1, 2, \dots, n$. Se $E \in \mathcal{F}$ è tale che $P(E) > 0$ allora si ha

$$P(F_h|E) = \frac{P(E|F_h)P(F_h)}{\sum_{k=1}^n P(E|F_k)P(F_k)} \quad h = 1, 2, \dots, n \quad (1.5.2)$$

Dimostrazione Dalla definizione di probabilità condizionata si ha

$$P(F_h|E) = \frac{P(F_h \cap E)}{P(E)} = \frac{P(E|F_h)P(F_h)}{P(E)}$$

così che la (1.5.2) si ottiene applicando la formula delle probabilità totali (1.5.1) al denominatore di questa uguaglianza. ■

Esempio 1.5.14 (Test clinici) ⁶ In un test clinico un individuo di una certa popolazione viene sottoposto ad un'analisi di laboratorio (*test*) per sapere se ha o meno una certa malattia. Il risultato del test può essere *negativo*, ad indicare che l'individuo è *sano* [rispetto a quella malattia], oppure *positivo*, ad indicare che l'individuo è *malato*. Tuttavia tutti i test utilizzati in pratica non sono completamente affidabili, nel senso che può accadere che

a sottoponendo un individuo sano al test, il test fornisce un risultato positivo (*falso positivo*)

⁶Si veda [3]

b sottoponendo un individuo malato al test, il test dà un risultato negativo (*falso negativo*).

Ovviamente un test è “buono” se rende minime le probabilità di osservare falsi positivi o negativi. Così, per valutare la bontà di un test, prima di applicarlo su larga scala, lo si verifica su individui di cui si conosce lo stato di salute. Supponiamo di sottoporre ad un test clinico un individuo, e siano M l’evento “l’individuo è malato”, S l’evento “l’individuo è sano”, I l’evento “il test è positivo” e O l’evento “il test è negativo”. Le grandezze $P(I|M)$ e $P(O|S)$ sono note nella letteratura epidemiologica rispettivamente come *sensibilità* e *specificità* del test e possono essere calcolate, o meglio stimate, utilizzando il test su individui dei quali si conosce lo stato di salute. In un buon test queste grandezze devono essere quanto più possibile prossime ad 1. Se il test viene utilizzato per capire se un individuo è malato o meno la grandezza che interessa è $P(M|I)$ detta *valore predittivo* del test. Per la formula di Bayes si ha che:

$$P(M|I) = \frac{P(I|M)P(M)}{P(I|M)P(M) + P(I|S)P(S)} = \frac{P(I|M)P(M)}{P(I|M)P(M) + [1 - P(O|S)][1 - P(M)]}$$

quindi per conoscere il valore predittivo del test non basta conoscere la specificità e la sensibilità del test ma bisogna conoscere anche $P(M)$. In definitiva bisogna avere informazioni *a priori* sulla frequenza relativa della malattia nella popolazione. Nella letteratura epidemiologica, $P(M)$ è nota con il nome di *prevalenza della malattia nella popolazione*. Si noti inoltre che se $P(M) \rightarrow 0$, anche $P(M|I)$ è piccolo, cosicché il test usato su una popolazione sana dà *quasi sempre* falsi positivi. Tanto per fare un esempio pratico consideriamo la metodica “ELISA” per la rilevazione degli anticorpi relativi al retrovirus HIV. Nel ’95 si stimava che gli individui che avevano sviluppato anticorpi relativi all’HIV in Italia fossero lo 0.0025% della popolazione totale. La sensibilità del test è 0.993 mentre la sua specificità è 0.9999. Ne segue che il valore predittivo del test è dato da:

$$P(M|I) = \frac{0.993 \times 0.000025}{0.993 \times 0.000025 + (1 - 0.9999) \times (1 - 0.000025)} \simeq 0.2 = 20\%$$

questo significa che se si effettuasse il test ELISA per l’HIV “a tappeto” su tutta la popolazione italiana l’80% circa dei positivi sarebbero falsi positivi! Per ovviare a questo inconveniente nella pratica si restringe la popolazione da esaminare alla cosiddetta “popolazione a rischio”, elevando in questo modo $P(M)$, e si consiglia a chi è risultato positivo alla metodica ELISA di sottoporsi ad un altro test, più costoso, ma anche più accurato.

Esercizio 1.5.15 (Test di collaudo; tratto da [7]) Un’impresa industriale ha installato un sistema automatico per il controllo di qualità, che garantisce che, se un pezzo è difettoso, esso viene eliminato con probabilità 0.995. Tuttavia, c’è una probabilità (piccola) pari a 0.001 che un pezzo non difettoso sia eliminato. Inoltre, si sa anche che la probabilità che un pezzo sia difettoso è 0.2. Si calcoli la probabilità che un pezzo non eliminato dopo il controllo di qualità sia difettoso.

Esercizio 1.5.16 Dimostrare la formula di Bayes per una partizione *numerabile* F_1, F_2, \dots di eventi.

Regola di moltiplicazione

Consideriamo ora l'esperimento di estrarre in sequenza e senza rimpiazzo delle biglie da un'urna che inizialmente ne contiene r rosse e b bianche. Per calcolare la probabilità che la prima biglia estratta sia rossa e la seconda bianca possiamo procedere come segue. Siano B_k l'evento "la k -esima biglia estratta è bianca" ed R_k l'evento "la k -esima biglia estratta è rossa". La probabilità richiesta è

$$P(R_1 \cap B_2) = P(B_2|R_1)P(R_1) = \frac{r}{r+b} \cdot \frac{b}{r+b-1}$$

Vogliamo ora calcolare la probabilità che la prima biglia estratta sia rossa, la seconda bianca, la terza rossa e la quarta ancora bianca, cioè $P(R_1 \cap B_2 \cap R_3 \cap B_4)$. Come possiamo estendere a questo caso il ragionamento precedente? In casi come questo risulta utile la seguente formula.

Proposizione 1.5.17 (Formula di moltiplicazione) *Sia (Ω, \mathcal{F}, P) uno spazio di probabilità ed $E_1, E_2, \dots, E_n \in \mathcal{F}$ eventi tali che $P(E_1 \cap E_2 \cap \dots \cap E_{n-1}) > 0$. Allora*

$$P(E_1 \cap E_2 \cap \dots \cap E_n) = P(E_1)P(E_2|E_1)P(E_3|E_2 \cap E_1) \dots P(E_n|E_1 \cap E_2 \cap \dots \cap E_{n-1})$$

Dimostrazione Poiché $E_1 \cap E_2 \cap \dots \cap E_{n-1} \subset E_1 \cap E_2 \cap \dots \cap E_{n-2} \subset \dots \subset E_1$, per la proprietà di monotonia si ha

$$0 < P(E_1 \cap E_2 \cap \dots \cap E_{n-1}) \leq P(E_1 \cap E_2 \cap \dots \cap E_{n-2}) \leq \dots \leq P(E_1)$$

quindi possiamo scrivere

$$\begin{aligned} P(E_1 \cap E_2 \cap \dots \cap E_n) &= \\ &= P(E_1) \cdot \frac{P(E_1 \cap E_2)}{P(E_1)} \cdot \frac{P(E_1 \cap E_2 \cap E_3)}{P(E_1 \cap E_2)} \dots \frac{P(E_1 \cap E_2 \cap \dots \cap E_n)}{P(E_1 \cap E_2 \cap \dots \cap E_{n-1})} = \\ &= P(E_1)P(E_2|E_1)P(E_3|E_1 \cap E_2) \dots P(E_n|E_1 \cap E_2 \cap \dots \cap E_{n-1}) \blacksquare \end{aligned}$$

Ritornando all'esempio dell'inizio del paragrafo

$$\begin{aligned} P(R_1 \cap B_2 \cap R_3 \cap B_4) &= P(R_1)P(B_2|R_1)P(R_3|R_1 \cap B_2)P(B_4|R_1 \cap B_2 \cap R_3) = \\ &= \frac{r}{r+b} \cdot \frac{b}{r+b-1} \cdot \frac{r-1}{r+b-2} \cdot \frac{b-1}{r+b-3} \end{aligned}$$

1.5.2 Indipendenza

L'indipendenza di eventi gioca un ruolo fondamentale nel calcolo delle probabilità. Intuitivamente due eventi sono indipendenti se il realizzarsi di uno dei due non influenza il verificarsi dell'altro. Analogamente un numero finito e qualunque di eventi sono indipendenti se il realizzarsi di un numero finito di essi non influenza il verificarsi dei rimanenti. Diamo ora le definizioni rigorose che si usano per formalizzare questi concetti.

Definizione 1.5.18 Sia (Ω, \mathcal{F}, P) uno spazio di probabilità. Gli eventi $E, F \in \mathcal{F}$ sono indipendenti se

$$P(E \cap F) = P(E)P(F)$$

Si noti che se E ed F sono eventi indipendenti tali che $P(E), P(F) > 0$ allora $P(E|F) = P(E)$ e $P(F|E) = P(F)$, in accordo con l'idea intuitiva di indipendenza e probabilità condizionata.

Definizione 1.5.19 Sia (Ω, \mathcal{F}, P) uno spazio di probabilità. Gli eventi E_1, E_2, \dots, E_n sono indipendenti se comunque preso un sottoinsieme $\{h_1, h_2, \dots, h_k\} \subset \{1, 2, \dots, n\}$ con $k \geq 2$ si ha

$$P(E_{h_1} \cap E_{h_2} \cap \dots \cap E_{h_k}) = P(E_{h_1})P(E_{h_2}) \cdot \dots \cdot P(E_{h_k}) \quad (1.5.3)$$

Esempio 1.5.20 Tre eventi A, B e C sono indipendenti se e solo se valgono tutte le seguenti relazioni: $P(A \cap B) = P(A)P(B)$, $P(A \cap C) = P(A)P(C)$, $P(B \cap C) = P(B)P(C)$ e $P(A \cap B \cap C) = P(A)P(B)P(C)$.

Esercizio 1.5.21 Analogamente all'esempio qui sopra, indicare le $2^4 - 4 - 1 = 11$ relazioni necessarie e sufficienti per l'indipendenza di 4 eventi A, B, C e D .

Esercizio* 1.5.22 Verificare che sono $2^n - n - 1$ le relazioni del tipo (1.5.3) necessarie e sufficienti per l'indipendenza di n eventi E_1, E_2, \dots, E_n .

Nota 1.5.23 Si noti che la Definizione 1.5.19 cattura il senso intuitivo di indipendenza secondo quanto detto all'inizio della sezione. Infatti se E_1, E_2, \dots, E_n sono eventi indipendenti si ha ad esempio

$$P(E_1 | E_{h_1} \cap E_{h_2} \cap \dots \cap E_{h_k}) = \frac{P(E_1 \cap E_{h_1} \cap E_{h_2} \cap \dots \cap E_{h_k})}{P(E_{h_1} \cap E_{h_2} \cap \dots \cap E_{h_k})} = P(E_1)$$

per ogni sottoinsieme $\{h_1, h_2, \dots, h_k\} \subset \{2, \dots, n\}$ e $P(E_{h_j}) > 0$ per ogni $j = 1, \dots, k$: cioè il realizzarsi di qualsivoglia scelta di eventi tra E_2, \dots, E_n non influenza il realizzarsi di E_1 . Un discorso analogo si può fare sostituendo E_2 ad E_1 etc.

Esercizio* 1.5.24 Siano E_1, E_2, \dots, E_n , con $n \geq 2$, eventi in uno spazio di probabilità (Ω, \mathcal{F}, P) tali che $P(\bigcap_{j=1}^n E_j) > 0$. Provare che E_1, E_2, \dots, E_n sono indipendenti se e solo se per ogni $k \geq 1$

$$P(E_1 | E_{h_1^1} \cap E_{h_2^1} \cap \dots \cap E_{h_k^1}) = P(E_1) \quad \text{per ogni} \quad \{h_1^1, h_2^1, \dots, h_k^1\} \subset \{1, 2, \dots, n\} \setminus \{1\}$$

$$P(E_2 | E_{h_1^2} \cap E_{h_2^2} \cap \dots \cap E_{h_k^2}) = P(E_2) \quad \text{per ogni} \quad \{h_1^2, h_2^2, \dots, h_k^2\} \subset \{1, 2, \dots, n\} \setminus \{2\}$$

...

$$P(E_n | E_{h_1^n} \cap E_{h_2^n} \cap \dots \cap E_{h_k^n}) = P(E_n) \quad \text{per ogni} \quad \{h_1^n, h_2^n, \dots, h_k^n\} \subset \{1, 2, \dots, n\} \setminus \{n\}$$

Nota 1.5.25 La Definizione 1.5.19 va letta e compresa con attenzione. Un errore tipico consiste nel non capirne il significato, tentando quindi di ricostruirla mnemonicamente a partire dal suo caso particolare e più facile da ricordare dato nella Definizione 1.5.18. In questo modo si arriva spesso al seguente errore: “gli eventi E_1, E_2, \dots, E_n sono indipendenti se $P(E_1 \cap E_2 \cap \dots \cap E_n) = P(E_1)P(E_2) \dots P(E_n)$ ” oppure “gli eventi E_1, E_2, \dots, E_n sono indipendenti se $P(E_h \cap E_k) = P(E_h)P(E_k)$ per ogni $h \neq k$ ”. Un altro errore tipico, in un certo senso più grave dei precedenti, è il seguente: “due eventi E ed F sono indipendenti se $E \cap F = \emptyset$ ”.

Esercizio 1.5.26 Provare che se E ed F sono due eventi *non impossibili*, cioè tali che $P(E) > 0$ e $P(F) > 0$, e se $E \cap F = \emptyset$, allora E ed F *non sono* indipendenti.

La nozione di indipendenza si estende naturalmente a successioni di eventi nel modo seguente:

Definizione 1.5.27 Sia (Ω, \mathcal{F}, P) uno spazio di probabilità. Si dice che gli eventi E_1, E_2, \dots sono indipendenti se preso comunque un sottoinsieme finito di eventi della successione esso è costituito da eventi indipendenti.

Cioè una successione di eventi è costituita da eventi indipendenti se preso comunque un sottoinsieme finito di eventi della successione esso è costituito da eventi indipendenti.

Esercizio 1.5.28 Sia (Ω, \mathcal{F}, P) uno spazio di probabilità, mostrare che gli eventi \emptyset, Ω sono indipendenti da qualsiasi evento o famiglia o successione di eventi in \mathcal{F} . Qual è il significato euristico di questa proprietà?

Esercizio 1.5.29 Sia (Ω, \mathcal{F}, P) uno spazio di probabilità, mostrare che se $E, F \in \mathcal{F}$ sono eventi indipendenti, allora lo sono anche E ed F^c , E^c ed F , E^c ed F^c . Quale è il significato euristico di questa proprietà?

Abbiamo già messo in evidenza che, se $F \in \mathcal{F}$ con $P(F) > 0$, allora la funzione $P_F(\cdot) = P(\cdot | F)$ è una probabilità su (Ω, \mathcal{F}) . Possiamo quindi considerare la nozione di indipendenza rispetto a questa probabilità.

Definizione 1.5.30 Sia (Ω, \mathcal{F}, P) uno spazio di probabilità e siano A_1, \dots, A_n e F eventi con $P(F) > 0$. Allora A_1, \dots, A_n si dicono condizionatamente indipendenti, dato F se essi sono indipendenti rispetto alla probabilità P_F .

Nota 1.5.31 *Attenzione!* L'indipendenza di due eventi *non* implica la loro indipendenza condizionatamente ad un terzo evento come mostra il seguente semplice esempio.

Esempio 1.5.32 Si lanciano due dadi regolari. Sia A l'evento: “il punteggio dei due dadi è uguale”, B l'evento: “il punteggio del secondo dado è 2” e C l'evento: “il punteggio del primo dado è pari”. Mostriamo che gli eventi A e B sono indipendenti ma non condizionatamente indipendenti, dato C . Lo spazio campionario relativo all'esperimento

“lancio di due dadi” è quello introdotto nell'Esempio 1.2.5 e gli eventi A , B e C corrispondono ai sottoinsiemi di Ω , $A = \{(i, i): i = 1, \dots, 6\}$, $B = \{(i, 2): i = 1, \dots, 6\}$, $C = \{(2i, j): i = 1, \dots, 3, j = 1, \dots, 6\}$ e $A \cap B = \{(2, 2)\}$. Quindi $P(A) = |A|/|\Omega| = 1/6$, $P(B) = |B|/|\Omega| = 1/6$ e $P(A \cap B) = |A \cap B|/|\Omega| = 1/36$. Poichè $P(A)P(B) = 1/36 = P(A \cap B)$, A e B sono indipendenti. Se invece calcoliamo le probabilità degli stessi eventi, ma condizionatamente all'evento C , otteniamo $P(A|C) = |A \cap C|/|C| = 3/18 = 1/6$, $P(B|C) = |B \cap C|/|C| = 3/18 = 1/6$ e $P(A \cap B|C) = |A \cap B \cap C|/|C| = 1/18 \neq 1/36 = P(A|C)P(B|C)$.

Esercizio 1.5.33 Mostrare con un controesempio che l'indipendenza condizionale non implica l'indipendenza.

Per comprendere meglio il significato della nozione di indipendenza condizionata proponiamo al lettore il seguente esercizio.

Esercizio 1.5.34 Un tribunale sta investigando sulla possibilità che sia accaduto un evento E molto raro e a tal fine interroga due testimoni, Arturo e Bianca. L'affidabilità dei due testimoni è nota alla corte: Arturo dice la verità con probabilità α e Bianca con probabilità β , e i loro comportamenti sono indipendenti. Siano A e B gli eventi Arturo e Bianca rispettivamente affermano che E è accaduto, e sia $p = P(E)$. Qual è la probabilità che E sia accaduto sapendo che Arturo e Bianca hanno dichiarato che E è accaduto? Assumendo $\alpha = \beta = 0.9$ e $p = 10^{-3}$, quale conclusione ne trae?

1.5.3 Prove di Bernoulli

Supponiamo di voler studiare un esperimento aleatorio, che chiameremo “prova”, in cui è possibile ottenere solo *due possibili risultati*: “successo” o “fallimento”. Supponiamo di poter *ripetere in condizioni identiche* questo esperimento un certo numero $n \in \mathbf{N}$ di volte in modo tale che *ogni prova non influenzi le altre*. L'esempio tipico è il lancio di una moneta. Indichiamo con successo l'uscita sulla moneta di una testa e con fallimento l'uscita di una croce e lanciamo la moneta un certo numero di volte. Vogliamo rispondere a domande del tipo, “qual è la probabilità di osservare 2 teste in 4 lanci?”. Poiché questo schema è relativamente generale conviene sviluppare un modello generale. Costruiamo quindi uno spazio di probabilità (Ω, \mathcal{F}, P) partendo dalle caratteristiche, sopra specificate in corsivo, dell'esperimento aleatorio.

Sia n il numero delle prove. Ogni possibile risultato delle n prove può essere rappresentato da una stringa binaria o n -upla, (a_1, a_2, \dots, a_n) dove $a_k = 1$ se la k -esima prova è un successo mentre $a_k = 0$ se la k -esima prova è un fallimento. Per esempio, se lanciamo una moneta $n = 4$ volte, la stringa $(1, 0, 0, 0)$ indica che al primo lancio si è ottenuta una testa, mentre ai rimanenti si sono ottenute croci. Ne segue che un buon candidato come spazio degli eventi elementari è l'insieme

$$\Omega = \{(a_1, a_2, \dots, a_n): a_k \in \{0, 1\}, k = 1, 2, \dots, n\}.$$

Essendo Ω un insieme finito di cardinalità 2^n (verificare!), possiamo prendere (cfr. Sezione 1.4) $\mathcal{F} := \mathcal{P}(\Omega)$ e per individuare P è sufficiente determinare $P(\{\omega\})$ per ogni $\omega \in \Omega$. A tal fine osserviamo che il fatto che *le varie prove non si influenzino a vicenda* si traduce nell'*indipendenza degli eventi*

$$\begin{aligned} E_1 &:= \{\text{la prima prova è un successo}\} \\ E_2 &:= \{\text{la seconda prova è un successo}\} \\ &\vdots \\ E_n &:= \{\text{l'n-esima prova è un successo}\} \end{aligned}$$

mentre, il fatto che ripetiamo l'esperimento in *condizioni identiche* si traduce nell'ipotesi di *uguale probabilità di successo ad ogni prova*: $P(E_1) = P(E_2) = \dots = P(E_n) = p \in (0, 1)$. Considerato che per ogni $\omega = (a_1, a_2, \dots, a_n) \in \Omega$ vale⁷

$$\{\omega\} = \left(\bigcap_{h \text{ tali che } a_h=1} E_h \right) \cap \left(\bigcap_{k \text{ tali che } a_k=0} E_k^c \right)$$

allora

$$P(\{\omega\}) = \prod_{h \text{ tali che } a_h=1} P(E_h) \prod_{k \text{ tali che } a_k=0} P(E_k^c)$$

[per l'indipendenza di E_1, E_2, \dots, E_n]

$$= \prod_{h \text{ tali che } a_h=1} p \prod_{k \text{ tali che } a_k=0} (1-p) = p^{\sum_{i=1}^n a_i} (1-p)^{n-\sum_{i=1}^n a_i},$$

dove per ottenere l'ultima eguaglianza abbiamo utilizzato il fatto che il numero degli h tali che $a_h = 1$ è $\sum_{i=1}^n a_i$, mentre il numero dei k tali che $a_k = 0$ è

$$n - \text{"il numero degli } h \text{ tali che } a_h = 1" = n - \sum_{i=1}^n a_i$$

Quindi per ogni $\omega \in \Omega$, $P(\{\omega\})$ è determinata una volta che sia noto il numero di cifre uguali ad 1 di ω , cioè il numero di successi ottenuti nelle n prove; cioè, $P(\{\omega\}) = p^k (1-p)^{n-k}$ se il numero di successi è k e p è la probabilità di ottenere un successo in una singola prova. Risulta così giustificata la seguente definizione.

Definizione 1.5.35 (Spazio di probabilità di Bernoulli) *Sia $n \in \mathbf{N}$ e $p \in (0, 1)$. Poniamo $\Omega := \{(a_1, a_2, \dots, a_n) : a_k \in \{0, 1\}, k = 1, 2, \dots, n\}$, $\mathcal{F} := \mathcal{P}(\Omega)$ e $P(\{(a_1, a_2, \dots, a_n)\}) = p^{\sum_{k=1}^n a_k} (1-p)^{n-\sum_{k=1}^n a_k}$ per ogni $(a_1, a_2, \dots, a_n) \in \Omega$. La terna (Ω, \mathcal{F}, P) si chiama spazio di probabilità di Bernoulli o spazio di probabilità di n prove di Bernoulli.*

⁷Supponiamo ad esempio $n = 4$ ed $\omega = (1, 0, 0, 1)$, il corrispondente evento è allora: "successo alla prima prova, fallimento alla seconda e terza prova, successo alla quarta prova", che è l'intersezione $E_1 \cap E_2^c \cap E_3^c \cap E_4$.

Supponiamo ora di lanciare una moneta 10 volte (o anche di lanciare 10 monete identiche); sappiamo che questo esperimento aleatorio può essere rappresentato mediante uno spazio di Bernoulli con $n = 10$ e $p = 1/2$ (se la moneta è equa). Ci chiediamo, ad esempio, “qual è la probabilità di osservare 4 teste e 6 croci?” Per rispondere a domande di questo genere è utile la seguente proposizione.

Proposizione 1.5.36 *La probabilità di osservare $k \leq n$ successi in una sequenza di $n \geq 1$ prove di Bernoulli se la probabilità di successo della singola prova è $p \in (0, 1)$ è data da*

$$\binom{n}{k} p^k (1-p)^{n-k}$$

Dimostrazione Sia (Ω, \mathcal{F}, P) lo spazio di probabilità di Bernoulli, e $B_k \in \mathcal{F}$ l'evento “si osservano k successi”, cioè

$$B_k = \left\{ (a_1, a_2, \dots, a_n) \in \Omega : \sum_{h=1}^n a_h = k \right\}$$

allora

$$P(B_k) = \sum_{\omega \in B_k} P(\{\omega\}) = \sum_{\omega \in B_k} p^k (1-p)^{n-k} = |B_k| p^k (1-p)^{n-k}$$

ma $|B_k| = \binom{n}{k}$, infatti per elencare tutte le stringhe lunghe n in cui k cifre sono uguali ad 1 ed $n - k$ sono uguali a 0, basta fissare i k posti degli 1 e questo può essere fatto in $\binom{n}{k}$ modi. ■

Nota 1.5.37 Si noti che gli eventi B_k , $k = 0, 1, \dots, n$, che fissano il numero di successi in n prove di Bernoulli, hanno probabilità che corrispondono ai valori p_k del modello binomiale (vedi Esempio 1.4.3). Quindi uno spazio di probabilità di Bernoulli induce sullo spazio campionario $\tilde{\Omega} = \{0, 1, \dots, n\}$ dell'esperimento che considera il numero dei successi nelle n prove, un modello binomiale di parametri n e p .

Capitolo 2

Variabili aleatorie

2.1 Variabili aleatorie

Abbiamo visto nel capitolo precedente come la teoria assiomatica del calcolo delle probabilità modella gli eventi casuali. In particolare abbiamo fatto la scelta di associare ad un esperimento aleatorio uno spazio di probabilità, cioè una terna (Ω, \mathcal{F}, P) , dove Ω è l'insieme di tutti i possibili risultati dell'esperimento casuale, \mathcal{F} è un insieme costituito da sottoinsiemi di Ω che vanno interpretati come eventi associati all'esperimento casuale e P è una funzione che ad ogni insieme $E \in \mathcal{F}$ associa un numero $P(E) \in [0, 1]$ da interpretare come la probabilità che l'evento (associato ad) E avvenga¹.

Una classe molto importante di eventi casuali sono quelli che hanno a che fare con i “numeri casuali”. Un numero casuale è proprio quello che il linguaggio comune suggerisce. Sia ad esempio T il tempo di vita di un componente elettronico: possiamo pensare a T come ad un numero casuale. Sia X il numero di teste che si presentano se lanciamo 1000 monete da un euro, allora X è un numero casuale. Per ragioni storiche nel calcolo delle probabilità i numeri casuali vengono chiamati *variabili aleatorie*. In questo capitolo introdurremo il concetto di variabile aleatoria da un punto di vista assiomatico e vedremo alcune applicazioni di questo concetto. In realtà nel Capitolo 1 abbiamo già studiato dei fenomeni casuali che nascondevano delle variabili aleatorie; quindi l'introduzione che ne faremo qui non aggiunge nulla da un punto di vista concettuale. Tuttavia parlare di numeri casuali, piuttosto che di eventi casuali, consente di utilizzare tutto l'apparato matematico che è stato sviluppato dall'analisi; ad esempio, potremo parlare di somma di variabili aleatorie, di limiti di successioni di variabili aleatorie etc., ottenendo così degli strumenti matematici piuttosto potenti.

Come abbiamo fatto nel Capitolo 1 per gli eventi casuali, dobbiamo dare una definizione matematicamente soddisfacente del concetto di numero casuale. Per la teoria assiomatica della probabilità le variabili aleatorie sono *funzioni sullo spazio degli eventi elementari* Ω . Per meglio capire questo concetto vediamo un esempio.

¹In quanto segue, se non c'è possibilità di errore, ometteremo frasi del tipo “un insieme E associato ad un certo evento” ma parleremo semplicemente dell'evento E , identificando gli insiemi con gli “eventi”.

Esempio 2.1.1 Viene lanciata tre volte una moneta non truccata e sia X il numero di teste che si presentano. Chiaramente X è un numero casuale che può assumere i valori 0, 1, 2, 3. L'esperimento che stiamo considerando rappresenta tre prove di Bernoulli con probabilità di successo in ogni singola prova pari ad $1/2$. Il modello probabilistico adeguato è quindi lo spazio di Bernoulli (Ω, \mathcal{F}, P) , dove $\Omega = \{(a_1, a_2, a_3) : a_i = 0, 1 \text{ } i = 1, 2, 3\}$ con $a_i = 1$ se all' i -esimo lancio esce testa e 0 altrimenti, $\mathcal{F} = \mathcal{P}(\Omega)$ e $P(E) = |E|/|\Omega|$. Questo spazio, fatta eccezione per la diversa rappresentazione degli eventi elementari, coincide con quello dell'Esempio 1.2.19. Ora possiamo pensare alla variabile aleatoria X come ad una regola che ad ogni $(a_1, a_2, a_3) \in \Omega$ associa il numero di teste che sono uscite se accade l'evento elementare rappresentato da (a_1, a_2, a_3) . Questo numero verrà denotato con $X((a_1, a_2, a_3))$ e vale $X((a_1, a_2, a_3)) = a_1 + a_2 + a_3$. Notiamo come in questo caso possiamo calcolare la probabilità che X assuma un certo valore. Ad esempio

$$\begin{aligned} P(X = 2) &= P(\{(a_1, a_2, a_3) \in \Omega : X((a_1, a_2, a_3)) = 2\}) \\ &= P(\{(1, 1, 0), (1, 0, 1), (0, 1, 1)\}) = \frac{3}{8} \end{aligned}$$

L'esempio appena visto dovrebbe far vedere perché nella teoria assiomatica della probabilità si pensa alle variabili aleatorie come a funzioni definite su Ω .

Un altro fatto importante al quale bisogna pensare, prima di vedere la definizione formale di variabile aleatoria, è il seguente. Sia X una variabile aleatoria definita su uno spazio di probabilità (Ω, \mathcal{F}, P) (cioè X è una funzione da Ω in \mathbb{R}) e chiediamoci: qual è la probabilità che X sia minore di un certo numero fissato x ? Oppure, qual è la probabilità che X sia maggiore di un certo numero fissato x ? Queste sembrano essere domande totalmente legittime e vorremmo che il nostro modello matematico contenesse al suo interno la possibilità di rispondere a domande di questo genere. In realtà, chiedersi ad esempio qual è la probabilità che X sia minore o uguale di un certo numero fissato x equivale a chiedersi qual è la probabilità dell'evento $E = \{\omega \in \Omega : X(\omega) \leq x\}$ e per calcolare questa probabilità è necessario che $E \in \mathcal{F}$, dal momento che $P(E)$ non è definita se $E \notin \mathcal{F}$. Questa questione “tecnica” non si pone se $\mathcal{F} = \mathcal{P}(\Omega)$ (come accade se Ω è finito o numerabile), perché essa è banalmente soddisfatta per ogni x . Tuttavia, la questione è rilevante in quanto si possono fare esempi, che non vedremo in questo corso, di spazi di probabilità (Ω, \mathcal{F}, P) e di funzioni $X : \Omega \rightarrow \mathbb{R}$ per i quali $\{\omega \in \Omega : X(\omega) \leq x\} \notin \mathcal{F}$. Ora possiamo dare la definizione di variabile aleatoria.

Definizione 2.1.2 (Variabile aleatoria) Sia (Ω, \mathcal{F}, P) uno spazio di probabilità. Una variabile aleatoria X è una funzione da Ω in \mathbb{R} tale che per ogni $x \in \mathbb{R}$, l'insieme $\{X \leq x\} := \{\omega \in \Omega : X(\omega) \leq x\} \in \mathcal{F}$.

Esempio 2.1.3 (Segue Esempio 2.1.1) Torniamo all'Esempio 2.1.1. Poiché $\{\omega \in \Omega : X(\omega) \leq 1\} = \{\omega : X(\omega) = 0\} \cup \{\omega : X(\omega) = 1\} = \{(0, 0, 0), (1, 0, 0), (0, 1, 0), (0, 0, 1)\}$, allora $P(X \leq 1) = P\{(0, 0, 0), (1, 0, 0), (0, 1, 0), (0, 0, 1)\} = 4/8 = 1/2$.

Esempio 2.1.4 (Segue Esempio 1.2.4) Torniamo all'Esempio 1.2.4 del tempo di vita di un componente elettronico. Ricordiamo che $\Omega := \mathbb{R}_+$ dove il punto $t \in \mathbb{R}_+$ significa che il componente si è guastato all'istante t . Un esempio di scelta per la probabilità P è $P((s, t]) = e^{-\lambda s} - e^{-\lambda t}$, se $0 \leq s \leq t$, dove $\lambda > 0$ è un parametro che dipende dal modello. Vedremo in seguito che questa scelta modella il guasto accidentale di un componente monitorato nel tempo continuo e non soggetto ad usura. L'istante di guasto T è una funzione $T : \Omega \rightarrow \mathbb{R}$ definita come la *funzione identità* cioè $T(\omega) := \omega$ per ogni $\omega \in \Omega$. Allora, se $0 \leq s < t$ abbiamo che

$$\{\omega \in \Omega : s < T(\omega) \leq t\} = \{\omega \in \Omega : s < \omega \leq t\} = (s, t]$$

da cui

$$P(s < T \leq t) = P((s, t]) = e^{-\lambda s} - e^{-\lambda t} \text{ se } 0 \leq s \leq t$$

Si noti che “fraudolentemente” non abbiamo detto chi è \mathcal{F} in questo caso. Il motivo non è una semplice dimenticanza, il problema è che in questo caso \mathcal{F} è un oggetto piuttosto complicato. Ci accontenteremo di dire che è possibile costruire \mathcal{F} in modo che contenga tutti gli intervalli di \mathbb{R}_+ (compreso lo stesso \mathbb{R}_+), i loro complementari e le loro unioni.

Se (Ω, \mathcal{F}, P) è uno spazio di probabilità ed X una variabile aleatoria su questo spazio, allora, per definizione, $\{X \leq x\} \in \mathcal{F}$ per ogni $x \in \mathbb{R}$. A partire da questa richiesta, si ottiene la seguente proposizione.

Proposizione 2.1.5 *Se X è una variabile aleatoria allora $\{X < x\}, \{X \geq x\}, \{X > x\}, \{x < X < y\}, \{x \leq X < y\}, \{x < X \leq y\}, \{x \leq X \leq y\}, \{X = x\}, \{X \neq x\}$ sono eventi (cioè sottoinsiemi di Ω che appartengono a \mathcal{F}).*

Esercizio* 2.1.6 Si dimostri la Proposizione 2.1.5

Aiuto Si usi nella dimostrazione il fatto che \mathcal{F} è una σ -algebra (quindi valgono le proprietà della Definizione 1.2.12). Per cominciare, si osservi che

$$\{X < x\} = \{\omega \in \Omega : X(\omega) < x\} = \bigcup_{n=1}^{+\infty} \{\omega \in \Omega : X(\omega) \leq x - 1/n\}$$

e $\{\omega \in \Omega : X(\omega) \leq x - 1/n\} \in \mathcal{F} \forall n \geq 1$, quindi...

2.1.1 Funzione di ripartizione

Nella sezione precedente abbiamo visto che il concetto di numero casuale è modellizzato da una funzione definita sullo spazio degli eventi elementari. In questa sezione vedremo come ad una variabile aleatoria X sia possibile associare una funzione reale F_X che ci permetterà di calcolare probabilità di eventi connessi a X .

Sia X una variabile aleatoria definita su uno spazio di probabilità (Ω, \mathcal{F}, P) e sia $x \in \mathbb{R}$. Per il punto 3. della Proposizione 1.2.17:

$$\begin{aligned} P(X > x) &= P(\{\omega \in \Omega : X(\omega) > x\}) = P(\{\omega \in \Omega : X(\omega) \leq x\}^c) = \\ &= 1 - P(\{\omega \in \Omega : X(\omega) \leq x\}) = 1 - P(X \leq x). \end{aligned}$$

Se invece $x, y \in \mathbb{R}$ con $x < y$, dal punto 3. della Proposizione 1.3.1 deriva che

$$\begin{aligned} P(x < X \leq y) &= P(\{\omega \in \Omega: x < X(\omega) \leq y\}) \\ &= P(\{\omega \in \Omega: X(\omega) \leq y\} \setminus \{\omega \in \Omega: X(\omega) \leq x\}) \\ &= P(\{\omega \in \Omega: X(\omega) \leq y\}) - P(\{\omega \in \Omega: X(\omega) \leq x\}) \\ &= P(X \leq y) - P(X \leq x) \end{aligned}$$

Quanto precede mostra che se conosciamo la funzione $F_X(x) := P(X \leq x)$, $x \in \mathbb{R}$, possiamo facilmente calcolare la probabilità di eventi associati a X . Per questa ragione alla funzione F_X si dà un nome particolare.

Definizione 2.1.7 (Funzione di ripartizione) *Sia X una variabile aleatoria definita su uno spazio di probabilità (Ω, \mathcal{F}, P) . Si chiama funzione di ripartizione di X la funzione $F_X: \mathbb{R} \rightarrow [0, 1]$ definita per ogni $x \in \mathbb{R}$ come $F_X(x) := P(X \leq x)$.*

Esempio 2.1.8 (Segue Esempio 2.1.1) Sia X la variabile aleatoria che indica il numero di teste ottenute in un lancio di tre monete non truccate dell'Esempio 2.1.1. Calcoliamo e rappresentiamo graficamente $F_X(x) = P(X \leq x)$. Innanzi tutto notiamo che X assume solo i valori 0, 1, 2 e 3. Quindi se $x < 0$ allora $F_X(x) = P(X \leq x) = 0$. Se $x = 0$ abbiamo che $F_X(0) = P(X = 0) = P(\{(0, 0, 0)\}) = 1/8$, mentre se $0 < x < 1$ abbiamo che $F_X(x) = P(X \leq x) = P(X \leq 0) = 1/8$, perché la variabile aleatoria X è più piccola o uguale ad un numero in $(0, 1)$ se e solo se è più piccola o uguale a 0. Se $x = 1$ abbiamo che $F_X(1) = P(X \leq 1) = P(X = 0) + P(X = 1) = (1/8) + (3/8) = 1/2$, mentre se $1 < x < 2$ abbiamo che $F_X(x) = P(X \leq x) = P(X \leq 1) = 1/2$. Analogamente, otteniamo $F_X(x) = 7/8$ se $2 \leq x < 3$. Infine, se $x \geq 3$ allora $F_X(x) = P(X \leq x) = 1$ semplicemente perché certamente $X \leq 3$. In definitiva:

$$F_X(x) = \begin{cases} 0 & \text{se } x < 0 \\ \frac{1}{8} & \text{se } 0 \leq x < 1 \\ \frac{1}{2} & \text{se } 1 \leq x < 2 \\ \frac{7}{8} & \text{se } 2 \leq x < 3 \\ 1 & \text{se } x \geq 3 \end{cases}$$

Esempio 2.1.9 (Segue Esempio 2.1.4) Sia T la variabile aleatoria che indica il tempo di rottura del componente elettronico che abbiamo visto nell'Esempio 2.1.4. Allora $F_T(t) = P(T \leq t) = 0$ se $t < 0$ mentre $F_T(t) = 1 - e^{-\lambda t}$ se $t \geq 0$. In definitiva

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

Il grafico di F_X è rappresentato in Figura 2.1 (a).

La funzione di ripartizione di una variabile aleatoria X gode di alcune proprietà:

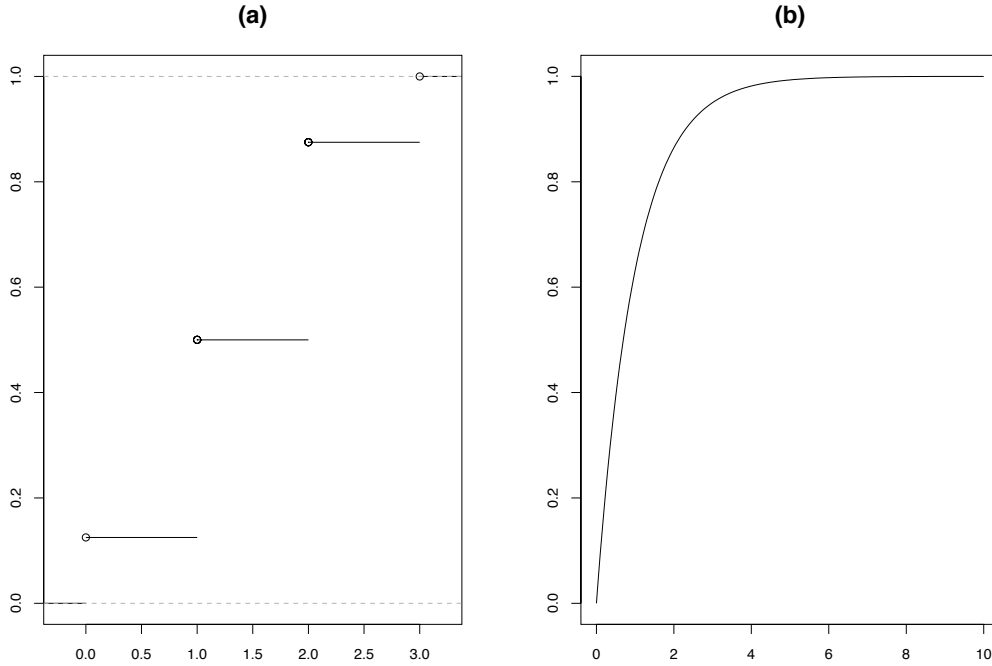


Figura 2.1: (a) Funzione di ripartizione F_X dell'Esempio 2.1.8, (b) funzione di ripartizione F_T dell'Esempio 2.1.9

Proposizione 2.1.10 *Sia X una variabile aleatoria definita su di uno spazio di probabilità (Ω, \mathcal{F}, P) e sia $F_X(x) = P(X \leq x)$ la sua funzione di ripartizione. Allora*

1. F_X è una funzione monotona non decrescente;
2. F_X è continua da destra, cioè $\lim_{x \downarrow x_0} F_X(x) = F_X(x_0)$, $\forall x_0 \in \mathbb{R}$;
3. $\lim_{x \rightarrow -\infty} F_X(x) = 0$ e $\lim_{x \rightarrow +\infty} F_X(x) = 1$.

Dimostrazione Si veda [1] o si provi a dimostrare per esercizio almeno la 1.. ■

Nota* 2.1.11 Le proprietà 1., 2. e 3. della Proposizione 2.1.10 sono importanti perché si può dimostrare (cosa che noi non faremo) che data una funzione F che le soddisfa, è possibile costruire uno spazio di probabilità (Ω, \mathcal{F}, P) e una variabile aleatoria X su (Ω, \mathcal{F}, P) che ha F come funzione di ripartizione. (Vedere Esempio 2.1.9). Potremo quindi parlare di “variabile aleatoria X con funzione di ripartizione F ” senza dover esplicitamente costruire lo spazio di probabilità dove X è definita.

La precedente osservazione giustifica la seguente

Definizione* 2.1.12 *Una funzione $F: \mathbb{R} \rightarrow \mathbb{R}$ è detta funzione di distribuzione su \mathbb{R} se soddisfa le seguenti condizioni*

1. F è funzione monotona non decrescente;

2. $F(x)$ è continua da destra $\forall x \in \mathbb{R}$;

3. $\lim_{x \rightarrow -\infty} F(x) = 0$ e $\lim_{x \rightarrow +\infty} F(x) = 1$.

La funzione di ripartizione di una variabile aleatoria è importante sia da un punto di vista applicativo sia da un punto di vista teorico. Per le applicazioni si può osservare che, se F_X è nota, da essa si possono calcolare facilmente probabilità collegate ad X . Si veda ad esempio l'esercizio seguente:

Esercizio 2.1.13 Sia X una variabile aleatoria definita su di uno spazio di probabilità (Ω, \mathcal{F}, P) e sia F_X la sua funzione di ripartizione. Mostrare che:

1. $P(X > x) = 1 - F_X(x)$ per ogni $x \in \mathbb{R}$;

2. $P(x < X \leq y) = F_X(y) - F_X(x)$ per ogni $x, y \in \mathbb{R}$ tali che $x \leq y$;

3* $P(X < x) = \lim_{y \uparrow x} F_X(y)$;

4. $P(X = x) = F_X(x) - \lim_{y \uparrow x} F_X(y)$.

Per quanto riguarda la teoria, le variabili aleatorie possono essere classificate a seconda di alcune proprietà delle loro funzioni di ripartizione. In generale la classificazione completa è piuttosto complessa e richiede strumenti matematici sofisticati. Noi introdurremo solamente le due classi di variabili aleatorie più importanti per le applicazioni a questo livello elementare, cioè le variabili aleatorie *discrete* e quelle *assolutamente continue*.

2.2 Variabili aleatorie discrete

Definizione 2.2.1 (Variabili aleatorie discrete) La variabile aleatoria X definita su uno spazio di probabilità (Ω, \mathcal{F}, P) è una variabile aleatoria discreta se assume, con probabilità uno, valori in un insieme S al più numerabile ($P(X \in S) = 1$).

Esempi di variabili aleatorie discrete sono: il numero di volte che bisogna lanciare una moneta prima di ottenere testa, il numero di successi in una sequenza di prove di Bernoulli, il numero di teste che si ottengono lanciando tre monete (cfr. Esempio 2.1.1). Per una variabile discreta è possibile definire una *densità discreta* nel modo seguente:

Definizione 2.2.2 Sia X una variabile aleatoria discreta su uno spazio di probabilità (Ω, \mathcal{F}, P) . Allora la funzione $p_X(x) := P(X = x)$ si chiama densità discreta della variabile aleatoria X .

Si noti che, se p_X è la densità di una variabile aleatoria discreta X , allora $p_X(x) = 0$ tranne che per una quantità al più numerabile di $x \in \mathbb{R}$.

Esempio 2.2.3 (Segue Esempio 2.1.1) Sia X il numero di teste che si ottengono lanciando tre volte una moneta equa. Sappiamo quindi che X può assumere solo i valori 0, 1, 2 e 3. Inoltre $P(X = 0) = P(\{(0, 0, 0)\}) = 1/8$, $P(X = 1) = P(\{(1, 0, 0), (0, 1, 0), (0, 0, 1)\}) = 3/8$, $P(X = 2) = P(\{(1, 1, 0), (1, 0, 1), (0, 1, 1)\}) = 3/8$ e $P(X = 3) = P(\{(1, 1, 1)\}) = 1/8$. Quindi

$$p_X(x) = P(X = x) = \begin{cases} \frac{1}{8} & \text{se } x \in \{0, 3\} \\ \frac{3}{8} & \text{se } x \in \{1, 2\} \\ 0 & \text{se } x \notin \{0, 1, 2, 3\} \end{cases}$$

Per rappresentare graficamente l'andamento di questa densità usiamo un diagramma a barre. Un *diagramma a barre* è costruito disegnando in corrispondenza di ogni valore x_k in S una barra perpendicolare all'asse delle ascisse di lunghezza uguale alla densità $p_X(x_k)$, come in Figura 2.2.

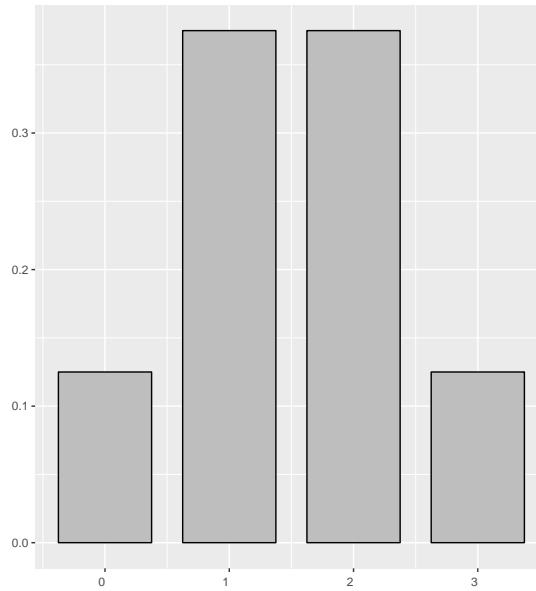


Figura 2.2: Densità di probabilità p_X dell'Esempio 2.2.3

Se p_X è la densità di X allora valgono le seguenti proprietà.

Proposizione 2.2.4 Sia p_X la densità di una variabile aleatoria discreta X che assume, con probabilità uno, valori in $S = \{x_k: k \in I\}$ ($I \subset \mathbb{Z}$). Allora

1. $0 \leq p_X(x) \leq 1$ per ogni $x \in \mathbb{R}$ e $p_X(x) = 0$ per ogni $x \notin S$;
2. $\sum_{k \in I} p_X(x_k) = 1$;

3. se F_X è la funzione di ripartizione di X allora

$$F_X(x) = \sum_{k: x_k \leq x} p_X(x_k) \quad \forall x \in \mathbb{R}$$

4. Se i punti di S possono essere numerati in modo tale che $x_h < x_k$ se $h < k$, allora

$$p_X(x_k) = F_X(x_k) - F_X(x_{k-1}) \quad \forall k \in I;$$

5. se $B \subset \mathbb{R}$ allora

$$P(X \in B) = \sum_{k: x_k \in B} p_X(x_k)$$

Dimostrazione

1. Ovvio, ricordando che $p_X(x) = P(X = x)$.

2. Infatti, per definizione di S abbiamo che $P(X \in S) = 1$

$$1 = P(X \in S) = P\left(\bigcup_{k \in I} \{X = x_k\}\right) = \sum_{k \in I} P(X = x_k) = \sum_{k \in I} p_X(x_k)$$

3. Ricordando che $F_X(x) = P(X \leq x)$ e che $P(X \in S) = 1$, allora

$$\begin{aligned} F_X(x) &= P(X \leq x) = P(X \in (-\infty, x] \cap S) = P\left(\bigcup_{k: x_k \leq x} \{X = x_k\}\right) \\ &= \sum_{k: x_k \leq x} P(X = x_k) = \sum_{k: x_k \leq x} p_X(x_k) \end{aligned}$$

4. Ricordiamo che dal punto 2. dell'Esercizio 2.1.13 segue che $F_X(x_k) - F_X(x_{k-1}) = P(x_{k-1} < X \leq x_k)$. Ma, se i punti di S sono numerati in modo tale che $x_h < x_k$ se $h < k$ allora $P(x_{k-1} < X \leq x_k) = P(X = x_k)$, da cui: $F_X(x_k) - F_X(x_{k-1}) = P(X = x_k) = p_X(x_k)$.

5. Poiché $P(X \in S) = 1$, allora

$$\begin{aligned} P(X \in B) &= P(X \in B \cap S) = P\left(\bigcup_{k: x_k \in B \cap S} \{X = x_k\}\right) = \sum_{k: x_k \in B \cap S} P(X = x_k) = \\ &= \sum_{k: x_k \in B} p(x_k) \quad \blacksquare \end{aligned}$$

I punti 3. e 4. della precedente proposizione mostrano come sia possibile ottenere dalla densità di una variabile aleatoria discreta la sua funzione di ripartizione e viceversa. In particolare ci dicono che se i punti di S possono essere numerati in modo tale che $x_h < x_k$ se $h < k$, allora la funzione di ripartizione di una variabile aleatoria discreta è una funzione “a gradini”, che i gradini sono situati nei punti dell’insieme S e che l’altezza del gradino corrispondente al punto $x_k \in S$ è proprio $p_X(x_k)$.

Esercizio 2.2.5 Capire il significato della Proposizione 2.2.4 verificandola per la variabile aleatoria dell’Esempio 2.1.1.

Il punto 5. della Proposizione 2.2.4 ci fa capire a cosa serve la densità discreta: ci permette di calcolare la probabilità che l’evento $\{X \in B\}$ si verifichi effettuando una semplice operazione algebrica e *senza sapere altro sulla variabile aleatoria X* .

Nota* 2.2.6 In realtà, nel seguito considereremo solamente variabili aleatorie discrete che assumono, con probabilità uno, valori in un insieme S che può essere rappresentato nel seguente modo: $S = \{x_k : k \in I\}$ con $x_h < x_k$ se $h < k$ e $I \subset \mathbb{Z}$. Per esempio, questa rappresentazione di S non è data se S è l’insieme \mathbb{Q} dei numeri razionali, mentre vale se S non ha punti di accumulazione. Se S ammette questa forma, sarà facile rappresentare graficamente la densità (mediante un diagramma a barre) e la funzione di ripartizione.

Nota 2.2.7 Un punto che ci interessa evidenziare è la motivazione euristica della parola “densità” utilizzata nel contesto delle variabili aleatorie discrete. Supponiamo che p_X sia la densità di una variabile aleatoria discreta X : questo significa che p_X attribuisce un numero $p_X(x) \geq 0$ ad ogni $x \in \mathbb{R}$; in particolare questo numero sarà non nullo solo per una quantità al più numerabile di punti $S := \{x_k : k \in I\} \subset \mathbb{R}$ con $I \subset \mathbb{Z}$. Un modo interessante di visualizzare questa situazione è immaginare i punti di S come punti materiali su una retta attribuendo al generico punto x_k la massa $m_k := p_X(x_k)$. In questo modo otteniamo una *distribuzione* di masse discrete sulla retta e p_X è proprio la densità di massa. Questa osservazione sarà particolarmente utile in seguito.

Esempio 2.2.8 Consideriamo i lanci successivi di una moneta equilibrata fino a quando non otteniamo testa. Sia X il numero di volte, inclusa l’ultima, che la moneta viene lanciata. Calcoliamo $P(X = k)$ per $k \in \mathbb{N}$. A tal fine consideriamo per $k = 1, 2, \dots$ gli eventi $E_k =$ “al k -esimo lancio otteniamo una testa” e osserviamo che questi eventi sono indipendenti con $P(E_k) = 1/2$ per $k = 1, 2, \dots$ essendo la moneta lanciata equilibrata. Per calcolare $P(X = 1)$ osserviamo che $X = 1$ se e solo se al primo lancio otteniamo una testa, da cui segue che $P(X = 1) = P(E_1) = 1/2$. Per $P(X = 2)$ osserviamo che $X = 2$ se e solo se al primo lancio otteniamo una croce ed al secondo lancio otteniamo una testa, quindi $\{X = 2\} = E_1^c \cap E_2$, da cui $P(X = 2) = P(E_1^c \cap E_2) = P(E_1^c)P(E_2) = 1/4$. Il ragionamento fatto sopra per $k = 2$ si estende facilmente a ogni $k \geq 2$ nel modo seguente: $X = k$ se e solo se abbiamo lanciato k volte la moneta ottenendo croce nei primi $k - 1$ lanci e testa nel k -esimo lancio. Pertanto, $P(X = k) = P(E_1^c \cap \dots \cap E_{k-1}^c \cap E_k) =$

$P(E_1^c) \cdots P(E_{k-1}^c)P(E_k) = 1/2^k$. Inoltre,

$$P(X \in \mathbf{N}) = \sum_{k=1}^{+\infty} P(X = k) = \sum_{k=1}^{+\infty} \frac{1}{2^k} = 1$$

Concludiamo che X è una variabile aleatoria discreta a valori in \mathbf{N} e la sua densità è

$$p_X(x) = \begin{cases} \frac{1}{2^k} & \text{se } x \in \mathbf{N} \\ 0 & \text{altrove} \end{cases}$$

Se vogliamo ora ad esempio calcolare la probabilità che siano necessari più di 3 lanci per ottenere la prima testa basta utilizzare il punto 5. della Proposizione 2.2.4:

$$P(X > 3) = \sum_{k>3} p_X(k) = \sum_{k=4}^{+\infty} \frac{1}{2^k} = \frac{1}{2^3} \sum_{k=4}^{+\infty} \frac{1}{2^{k-3}} = \frac{1}{2^3} \sum_{k=1}^{+\infty} \frac{1}{2^k} = \frac{1}{2^3}$$

Nota* 2.2.9 Prima di vedere alcuni esempi importanti di densità discrete, torniamo un momento ai punti 1. e 2. della Proposizione 2.2.4. Una domanda naturale è la seguente: una funzione reale $p(\cdot)$, diversa da zero su un insieme al più numerabile $S = \{x_k : k \in I\}$ ($I \subset \mathbb{Z}$), che verifica le proprietà 1. e 2 della Proposizione 2.2.4, può essere sempre vista come densità di una variabile aleatoria discreta? Più precisamente, è sempre possibile costruire uno spazio di probabilità (Ω, \mathcal{F}, P) ed una variabile aleatoria X su di esso che ha $p(\cdot)$ come densità, cioè tale che $p_X(x) = p(x)$? La risposta è affermativa. Infatti basta prendere $\Omega = S$, $\mathcal{F} = \mathcal{P}(S)$ e P l'unica misura di probabilità su S tale che $P(\{x_k\}) = p(x_k)$ con $k \in I$, come mostrato nella Sezione 1.4. È immediato, quindi, verificare che la variabile aleatoria discreta $X(\omega) = \omega$, per ogni $\omega \in \Omega$, ha densità $p(\cdot)$.

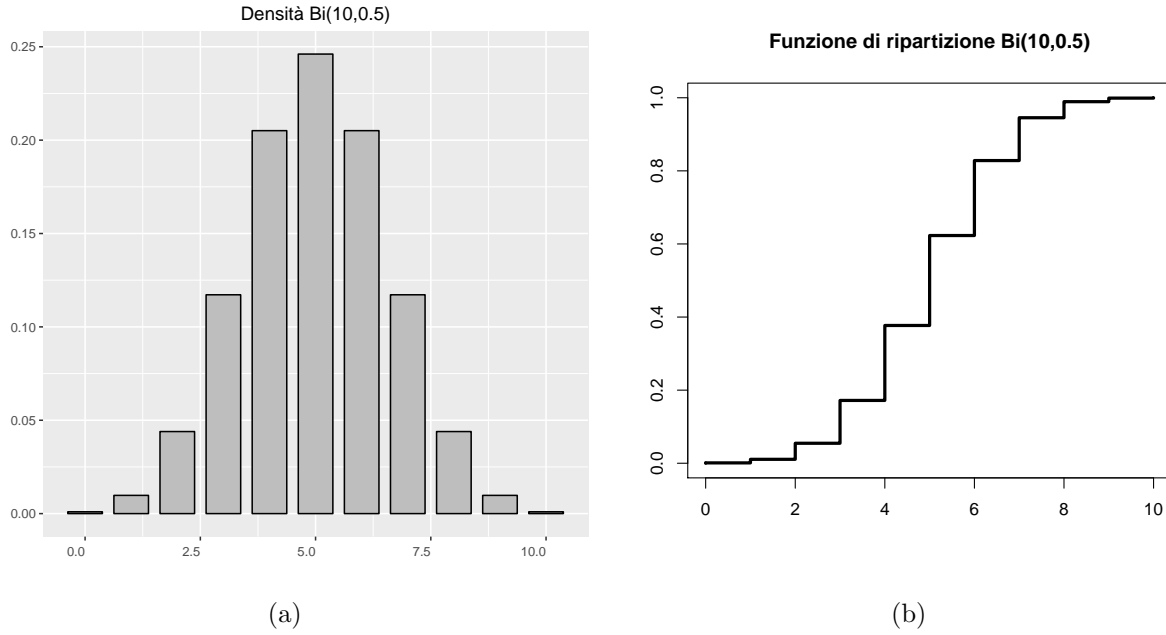
La precedente osservazione ci permetterà di parlare di variabili aleatorie assegnandone la densità, senza costruire esplicitamente lo spazio di probabilità dove X è definita e giustifica la seguente definizione.

Definizione* 2.2.10 Sia $S = \{x_k : k \in I\} \subset \mathbb{R}$ con $I \subset \mathbb{Z}$. Una funzione $p: \mathbb{R} \rightarrow \mathbb{R}$ è una densità discreta su S se

1. $0 \leq p(x) \leq 1$ per ogni $x \in \mathbb{R}$ e $p(x) = 0$ per ogni $x \notin S$;
2. $\sum_{k \in I} p(x_k) = 1$.

2.3 Esempi di densità discrete notevoli

Vediamo ora in dettaglio alcuni esempi di densità discrete che sono importanti per le applicazioni.

Figura 2.3: $\mathbf{Bi}(10, 0.5)$

2.3.1 Densità binomiale e bernoulliana

Consideriamo di nuovo le prove di Bernoulli definite nella Sezione 1.5.3. In quella sezione avevamo visto che se $p \in (0, 1)$ è la probabilità di ottenere il successo in una singola prova di Bernoulli, la probabilità di ottenere k successi in n prove ($k \leq n$) è

$$\binom{n}{k} p^k (1-p)^{n-k}$$

Definiamo ora la variabile aleatoria X come “il numero di successi ottenuti in n prove di Bernoulli”. Si vede subito che X può assumere solo i valori $0, 1, \dots, n$ ed è quindi una variabile aleatoria discreta. Inoltre, per quanto ricordato, la sua densità è

$$p_X(k) = P(X = k) = \begin{cases} \binom{n}{k} p^k (1-p)^{n-k} & \text{se } k \in \{0, 1, \dots, n\} \\ 0 & \text{se } k \notin \{0, 1, \dots, n\} \end{cases}$$

che prende il nome di *densità binomiale di parametri n e p* . Equivalentemente si dice che X è una variabile aleatoria *binomiale di parametri n e p* o ancora $X \sim \mathbf{Bi}(n, p)$. La Figura 2.3 fornisce il grafico a barre della densità ed il grafico della funzione di ripartizione di una variabile aleatoria $X \sim \mathbf{Bi}(10, 0.5)$, mentre la Figura 2.4 mostra l'andamento delle densità $\mathbf{Bi}(10, 0.2)$ e $\mathbf{Bi}(10, 0.8)$, mediante grafici a barre. Sia $X \sim \mathbf{Bi}(n, p)$; se $n = 1$ questa variabile rappresenta il numero di successi in una sola prova con probabilità di successo p , cioè X assume solo i valori 0 e 1, e la densità di X è $p_X(k) = p^k (1-p)^{1-k}$ se $k \in \{0, 1\}$ e

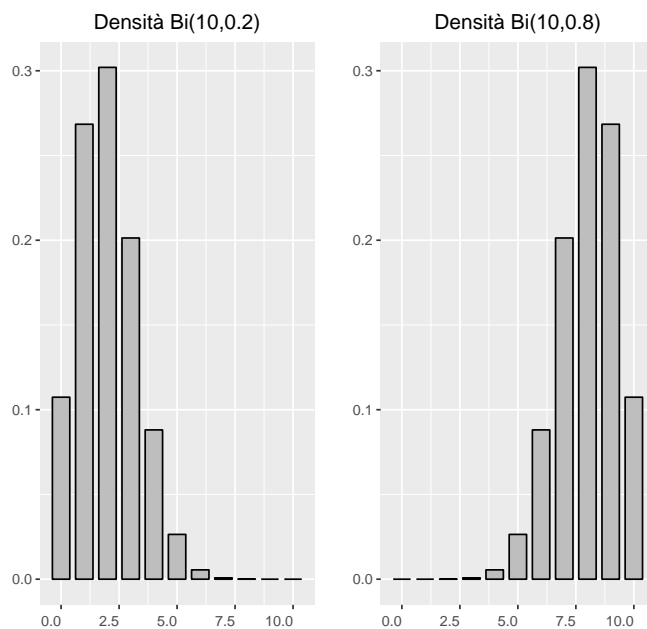


Figura 2.4: Densità di probabilità $\mathbf{Bi}(10, 0.2)$ e $\mathbf{Bi}(10, 0.8)$ a confronto

$p_X(k) = 0$ se $k \notin \{0, 1\}$, cioè

$$p_X(k) = \begin{cases} 1 - p & \text{se } k = 0 \\ p & \text{se } k = 1 \\ 0 & \text{se } k \notin \{0, 1\} \end{cases}$$

Questa densità prende il nome di *densità bernoulliana di parametro p* ; equivalentemente si dice che X è una *variabile aleatoria bernoulliana di parametro p* o ancora $X \sim \mathbf{Be}(p)$.

Per ragioni di comodità si dice che la variabile aleatoria costante $X \equiv 1$, cioè la variabile aleatoria che vale sempre 1, è bernoulliana di parametro 1 e che la variabile aleatoria costante $X \equiv 0$, cioè la variabile aleatoria che vale sempre 0, è bernoulliana di parametro 0.

Esempio 2.3.1 Riempendo a caso una schedina di totocalcio, qual è la probabilità di fare almeno 12?

Su una schedina del totocalcio sono elencate 14 partite e ogni partita può avere tre risultati “1”, “2” o “X”, ad indicare rispettivamente la vittoria della squadra ospitante, della squadra ospite o la parità. La probabilità di azzeccare una singola partita, scrivendo a caso uno dei simboli 1, 2 o X, è -almeno in prima approssimazione- uguale ad $1/3$. Inoltre l’aver azzeccato o meno il risultato di una certa partita non influenza la capacità di azzeccare le altre. Possiamo quindi schematizzare il nostro esperimento aleatorio con una successione di $n = 14$ prove di Bernoulli, con probabilità di successo nella singola prova

$p = 1/3$. Sia Y il numero di partite azzeccate; allora $Y \sim \mathbf{Bi}(14, 1/3)$ e

$$P(Y \geq 12) = \sum_{k=12}^{14} p_Y(k) = \sum_{k=12}^{14} \binom{14}{k} \left(\frac{1}{3}\right)^k \left(\frac{2}{3}\right)^{14-k} = \frac{393}{4782969} \simeq 0.00008$$

Esercizio 2.3.2 Supponiamo che da un'urna contenente r biglie rosse e b biglie bianche estraiamo a caso una biglia, prendiamo nota del suo colore e la reinseriamo nell'urna. Quindi, ripetiamo questa procedura $n \geq 1$ volte e sia X il numero di biglie rosse estratte nelle n estrazioni. Verificare che $X \sim \mathbf{Bi}(n, r/(r+b))$.

2.3.2 Densità Geometrica

Supponiamo di avere un'apparecchiatura *non soggetta ad usura* ed inizialmente funzionante, ma che si può guastare per motivi contingenti. Supponiamo di controllare il funzionamento dell'apparecchiatura agli istanti $1, 2, \dots$. Sia X l'istante in cui l'apparecchiatura si guasta. Vogliamo vedere se è possibile costruire un modello probabilistico per X . A tal fine osserviamo che se controlliamo l'apparecchiatura al tempo $t = k$ e la troviamo funzionante, la probabilità che l'apparecchiatura sia ancora funzionante al tempo $t = k + 1$ è la stessa di quella di trovarla funzionante al tempo $t = 1$; infatti, stiamo semplicemente cercando la probabilità che si guasti in un intervallo di tempo unitario, che è costante per l'ipotizzata assenza di usura. In formule:

$$P(X > k + 1 | X > k) = P(X > 1), \quad k = 1, 2, \dots$$

La precedente identità ci permette di determinare la densità di X se conosciamo $q := P(X > 1)$. Infatti:

$$q = P(X > 1) = \frac{P(X > k + 1, X > k)}{P(X > k)} = \frac{P(X > k + 1)}{P(X > k)}$$

da cui $P(X > k + 1) = qP(X > k)$, $k = 1, 2, \dots$. Quindi

$$P(X > 2) = qP(X > 1) = q^2$$

$$P(X > 3) = qP(X > 2) = q^3$$

$$\vdots$$

$$P(X > k + 1) = qP(X > k) = q^{k+1}$$

Segue che $F_X(k) = 1 - P(X > k) = 1 - q^k$ e per il punto 4. della Proposizione 2.2.4

$$P(X = k) = F_X(k) - F_X(k - 1) = q^{k-1} - q^k = q^{k-1}(1 - q)$$

Se ora chiamiamo $p := 1 - q = P(X \leq 1)$ intensità di guasto, possiamo scrivere

$$P(X = k) = p(1 - p)^{k-1}, \quad k = 1, 2, \dots$$

Notiamo che

$$P(X \in \mathbf{N}) = \sum_{k=1}^{+\infty} P(X = k) = \sum_{k=1}^{+\infty} p(1-p)^{k-1} = p \sum_{k=0}^{+\infty} (1-p)^k = p \frac{1}{1-(1-p)} = 1$$

Quindi X è una variabile aleatoria discreta a valori in \mathbf{N} con densità

$$p_X(k) = \begin{cases} p(1-p)^{k-1} & \text{se } k = 1, 2, \dots \\ 0 & \text{altrove} \end{cases}$$

Questa densità prende il nome di *densità geometrica di parametro p* . Una variabile aleatoria con questa densità è detta *variabile geometrica di parametro p* e si scrive $X \sim \mathbf{Geom}(p)$.

Esempio 2.3.3 Supponiamo di eseguire una successione di prove di Bernoulli, con probabilità di successo nella singola prova pari a $p \in (0, 1)$. Sia X il numero di prove necessarie per osservare il primo successo, inclusa l'ultima. Verificare che X ha densità geometrica di parametro p .

2.3.3 Densità di Poisson come limite di densità binomiale

Consideriamo il centralino di un numero verde. Questo in genere è costituito da un certo numero di linee alle chiamate delle quali rispondono degli operatori. Sia ora X il numero di chiamate che arrivano ad un certo operatore in un'ora. In un modello piuttosto semplificato possiamo pensare ad un grande numero n di utenti ognuno dei quali ha una probabilità molto piccola $p \in (0, 1)$ di chiamare il numero verde in questione per mettersi in contatto con l'operatore. Se assumiamo che i singoli utenti si mettono in contatto con l'operatore indipendentemente uno dall'altro otteniamo che $X \sim \mathbf{Bi}(n, p)$, dove n è un numero molto grande e p un numero molto piccolo. Se il numero verde è organizzato razionalmente il numero delle linee è commisurato al bacino di utenza, in modo tale che vi sia un'alta probabilità di trovare il numero verde libero. Una condizione perché ciò accada è che $\lambda := np$ sia un numero fissato e non eccessivamente grande. In questo caso possiamo scrivere $X \sim \mathbf{Bi}(n, \lambda/n)$, cioè

$$P(X = k) = \binom{n}{k} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k}$$

Per capire cosa succede a $P(X = k)$ se n è grande osserviamo che

$$\begin{aligned} \binom{n}{k} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} &= \frac{n!}{(n-k)!k!} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} \\ &= \frac{n!}{(n-k)!n^k} \left(1 - \frac{\lambda}{n}\right)^{-k} \frac{\lambda^k}{k!} \left(1 - \frac{\lambda}{n}\right)^n \end{aligned}$$

ma

$$\lim_{n \rightarrow +\infty} \frac{n!}{(n-k)!n^k} = 1$$

come rapporto di polinomi di grado k ,

$$\lim_{n \rightarrow +\infty} \left(1 - \frac{\lambda}{n}\right)^{-k} = 1$$

e

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

come ben noto dal corso di analisi. Segue che

$$P(X = k) \simeq \frac{\lambda^k}{k!} e^{-\lambda}, \quad k = 0, 1, \dots, n, \quad \lambda = np \quad (2.3.1)$$

Tenendo conto di quanto detto sopra, per $\lambda > 0$ introduciamo la densità

$$p(k) := \begin{cases} \frac{e^{-\lambda} \lambda^k}{k!} & \text{se } k \in \{0, 1, 2, \dots\} \\ 0 & \text{se } k \notin \{0, 1, 2, \dots\} \end{cases}$$

che prende il nome di *densità di Poisson di parametro λ* . Una variabile aleatoria con questa densità è detta *variabile di Poisson di parametro λ* e si scrive $X \sim \mathcal{P}(\lambda)$.

Esercizio 2.3.4 Verificare che la densità di Poisson di parametro λ è una densità, cioè che verifica la Definizione 2.2.10.

Esempio 2.3.5 Il numero di automobili X che attraversano la porta di un casello autostradale in un minuto è una variabile aleatoria di Poisson di parametro 3.2. La probabilità che in un minuto non passi nessuna automobile è

$$P(X = 0) = e^{-3.2} \simeq 0.0408$$

La probabilità che ne passino più di 2 è

$$\begin{aligned} P(X > 2) &= 1 - P(X \leq 2) = 1 - [p_X(0) + p_X(1) + p_X(2)] \\ &= 1 - e^{-3.2} \times \left(1 + 3.2 + \frac{3.2^2}{2!}\right) \simeq 0.6201 \end{aligned}$$

Nota 2.3.6 La formula (2.3.1) oltre che per introdurre la distribuzione di Poisson può essere utilizzata per calcolare valori approssimati di $P(X = k)$ quando $X \sim \mathbf{Bi}(n, p)$ con n grande e p piccolo in quanto evita il calcolo di coefficienti binomiali.

Esempio 2.3.7 Un computer ha probabilità $p = 10^{-3}$ di ricevere un carattere errato. Sia X il numero di errori in un messaggio di 1000 caratteri. Per calcolare la probabilità che il computer riceva più di un errore in una trasmissione di 1000 caratteri, osserviamo che se gli errori avvengono indipendentemente, allora $X \sim \mathbf{Bi}(1000, 10^{-3})$. Usando l'approssimazione di Poisson con $\lambda = np = 1000 \cdot 10^{-3} = 1$, otteniamo $P(X > 1) = 1 - P(X \leq 1) = 1 - e^{-1} \times 1^0/0! - e^{-1} \times 1^1/1! \simeq 0.2642411$. Effettuando il calcolo esatto abbiamo

$$\begin{aligned} P(X > 1) &= 1 - P(X = 0) - P(X = 1) \\ &= 1 - \binom{1000}{0} (10^{-3})^0 (1 - 10^{-3})^{1000} - \binom{1000}{1} (10^{-3})^1 (1 - 10^{-3})^{999} \\ &\simeq 0.2642410. \end{aligned}$$

2.3.4 Densità ipergeometrica

Siamo ora interessati a contare il numero totale X di biglie rosse ottenute su n estrazioni senza rimpiazzo da un'urna che ne contiene r rosse e b bianche. Ovviamente X è un numero intero e si intuisce subito che X è più piccolo del numero di estrazioni n e anche del numero di biglie rosse contenute nell'urna r ; in definitiva X è più piccolo del minimo $n \wedge r$ tra n ed r . Inoltre X è non negativo, ma se il numero delle biglie bianche b è inferiore a quello delle estrazioni n allora necessariamente verranno estratte $n - b$ biglie rosse e quindi $X \geq n - b$. Abbiamo che X è più grande del massimo $0 \vee (n - b)$ tra 0 ed $(n - b)$. In generale X assume valori in $S := \{0 \vee (n - b), 0 \vee (n - b) + 1, \dots, n \wedge r\}$. Fissato $k \in S$, possiamo calcolare $P(X = k)$ come casi favorevoli su casi possibili. Ci sono $\binom{r+b}{n}$ modi di scegliere n biglie

tra $r + b$. Tra questi ci sono $\binom{r}{k}$ modi di scegliere le k biglie rosse tra le r disponibili e per ciascuna di queste scelte, le rimanenti $n - k$ biglie possono essere scelte fra le b bianche in $\binom{b}{n-k}$ modi. In definitiva:

$$p_X(k) = P(X = k) = \begin{cases} \frac{\binom{r}{k} \binom{b}{n-k}}{\binom{r+b}{n}} & \text{se } k \in \{0 \vee (n - b), \dots, n \wedge r\} \\ 0 & \text{se } k \notin \{0 \vee (n - b), \dots, n \wedge r\} \end{cases}$$

La densità p_X è detta *densità ipergeometrica di parametri $(b + r, r, n)$* e una variabile aleatoria X con questa densità è detta *variabile aleatoria ipergeometrica di parametri $(b + r, r, n)$* e si scrive $X \sim \mathbf{Iperg}(b + r, r, n)$.

Esempio 2.3.8 Il 5% di un lotto di 100 fusibili è soggetto a controllo casuale prima di essere immesso sul mercato. Se un fusibile non brucia ad un determinato amperaggio l'intero lotto viene mandato indietro. Se il lotto contiene 10 fusibili difettosi, qual è la probabilità che il lotto sia rispedito indietro?

Il lotto è rispedito indietro se almeno un fusibile sui 5 (= 5% dei 100) scelti a caso per il controllo non brucia ad un determinato amperaggio. I 5 fusibili da controllare sono estratti

senza rimpiazzo dal lotto di 100 pezzi costituito da 90 fusibili funzionanti e 10 difettosi. Pertanto, la variabile aleatoria X che conta il numero di fusibili difettosi su 5 ha densità ipergeometrica di parametri $(100, 10, 5)$:

$$P(X = k) = \frac{\binom{10}{k} \binom{90}{5-k}}{\binom{100}{5}} \quad k = 0, \dots, 5$$

e

$$\begin{aligned} P(\text{"il lotto è rispedito indietro"}) &= P(X \geq 1) = 1 - P(X = 0) \\ &= 1 - \frac{\binom{10}{0} \binom{90}{5}}{\binom{100}{5}} \simeq 1 - 0.5838 = 0.4162 \end{aligned}$$

Nota* 2.3.9 Supponiamo di estrarre le n biglie dall'urna contenente $r + b$ biglie in sequenza. Sia E_k , $k = 1, \dots, n$ l'evento "estraggo una biglia rossa la k -esima volta". Per calcolare $P(E_k)$ come casi favorevoli su casi possibili questa volta dobbiamo distinguere l'ordine. Ci sono $(r + b)(r + b - 1) \cdots (r + b - n + 1)$ modi di estrarre in sequenza le n biglie tra $r + b$ disponibili. Tra questi quelli in cui la k -esima biglia è rossa sono $r(r + b - 1)(r + b - 2) \cdots (r + b - n + 1)$. Per convincersene basta osservare che posso scegliere la biglia rossa al k -esimo posto tra le r disponibili in r modi, poi posso scegliere le altre $n - 1$ biglie tra le rimanenti $r + b - 1$ in $(r + b - 1)(r + b - 1 - 1) \cdots [r + b - 1 - (n - 1) + 1]$ modi. Quindi

$$P(E_k) = \frac{r(r + b - 1)(r + b - 2) \cdots (r + b - n + 1)}{(r + b)(r + b - 1) \cdots (r + b - n + 1)} = \frac{r}{r + b}$$

Dichiariamo ora di ottenere un "successo" quando viene estratta una biglia rossa e un "fallimento" quando viene estratta una biglia bianca. In questo modo, analogamente a quanto fatto per le prove di Bernoulli, possiamo pensare all'estrazione sequenziale dall'urna come ad una successione di prove, in cui la probabilità di ottenere un successo nella k -esima prova è $p = r/(r + b)$. La differenza sostanziale tra queste prove e quelle di Bernoulli è che questa volta le prove *non sono indipendenti*. Infatti la probabilità di ottenere un successo alla seconda prova se abbiamo ottenuto un successo alla prima è differente dalla probabilità di ottenere un successo alla seconda prova se non abbiamo ottenuto un successo alla prima. Questo perché nel primo caso stiamo estraendo da un'urna contenente $r + b - 1$ biglie di cui $r - 1$ rosse e b bianche, mentre nel secondo caso stiamo estraendo da un'urna contenente $r + b - 1$ biglie di cui r rosse e $b - 1$ bianche. Comunque, la dipendenza tra prove si attenua se il numero delle biglie presenti nell'urna $r + b$ è grande. Infatti ad esempio:

$$\frac{P(E_2|E_1)}{P(E_2)} = \frac{r - 1}{r + b - 1} \cdot \frac{r + b}{r} \rightarrow 1$$

se $r + b$ tende opportunamente a $+\infty$ (per esempio in modo tale che $r/(r + b) \rightarrow \theta \in (0, 1)$). Quindi, se $r + b$ è grande, allora $P(E_2|E_1) \simeq P(E_2)$. In altri termini, se vi sono molte biglie

nell'urna, rimpiazzare o non rimpiazzare le biglie ad ogni successiva estrazione non modifica in modo significativo il risultato. Quanto fin qui detto in parte spiega euristicamente il fatto che per $r + b$ grande, qualche volta, potremo approssimare la densità ipergeometrica $\mathbf{Iperg}(b + r, r, n)$ con la densità binomiale $\mathbf{Bi}(n, r/(r + b))$. Un'esemplificazione di questo fatto è in Figura 2.5² che rappresenta l'andamento della densità ipergeometrica all'aumentare di $r + b$ rispetto alla densità $\mathbf{Bi}(10, r/(r + b))$.

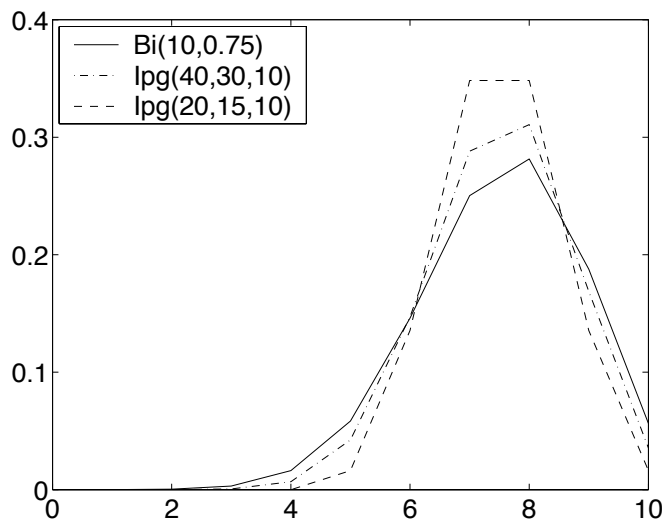


Figura 2.5: Densità ipergeometrica (Ipg) e binomiale a confronto

Esercizio* 2.3.10 C'è qualche legame fra la soluzione dell'Esercizio 2.3.2 e la scelta della densità $\mathbf{Bi}(n, r/(r + b))$ nell'approssimazione della legge ipergeometrica di parametri $(b + r, r, n)$?

Esercizio* 2.3.11 Dimostrare che, fissato n , se $r + b \rightarrow +\infty$ e $r/(r + b) \rightarrow \theta \in (0, 1)$, i valori della densità ipergeometrica di parametri $(b + r, r, n)$ tendono ai corrispondenti valori della densità binomiale di parametri (n, θ) .

Nota 2.3.12 Il lettore avrà già rilevato che le densità delle variabili aleatorie sopra presentate coincidono con alcuni degli esempi di modelli di probabilità su spazi finiti o numerabili presentate nell'Esempio 1.4.3 della Sezione 1.4. Quanto presentato in questa sezione è quindi rivolto anche a mostrare in quali situazioni tali modelli probabilistici vengono adottati.

²in Figura 2.5 gli 11 valori (isolati) in ordinata delle densità sono stati congiunti mediante spezzate.

2.4 Variabili aleatorie assolutamente continue

Un concetto in un certo senso opposto a quello di variabile aleatoria discreta, anche se poi come vedremo operativamente analogo, è quello di variabile aleatoria assolutamente continua.

Definizione 2.4.1 (Variabili aleatorie assolutamente continue) *La variabile aleatoria X definita su di uno spazio di probabilità (Ω, \mathcal{F}, P) è una variabile aleatoria assolutamente continua se esiste una funzione $f_X: \mathbb{R} \rightarrow \mathbb{R}_+$ integrabile, tale che la funzione di ripartizione F_X di X si può scrivere come*

$$F_X(x) = \int_{-\infty}^x f_X(s) ds \quad (2.4.1)$$

f_X prende il nome di densità di X .

Dalla definizione data qui sopra si vede subito che F_X è una funzione continua, quindi se X è una variabile aleatoria assolutamente continua, per l'Esercizio 2.1.13, $P(X = x) = F_X(x) - \lim_{y \uparrow x} F_X(y) = 0$ per ogni $x \in \mathbb{R}$! In questo senso le variabili aleatorie assolutamente continue sono molto differenti dalle variabili aleatorie discrete.

Esercizio 2.4.2 Si dimostri che se X è variabile aleatoria assolutamente continua con funzione di ripartizione F_X , allora

$$P(X < x) = F_X(x) \quad \forall x \in \mathbb{R}$$

Esempio 2.4.3 (Segue Esempio 2.1.4) Sia T la variabile aleatoria che rappresenta il tempo di rottura dell'Esempio 2.1.4. Poiché avevamo visto nell'Esempio 2.1.9 che la funzione di ripartizione di T è

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

ne segue che T è una variabile aleatoria assolutamente continua con densità

$$f_T(t) := \begin{cases} 0 & \text{se } t < 0 \\ \lambda e^{-\lambda t} & \text{se } t \geq 0 \end{cases}$$

Infatti si ha che

$$\int_{-\infty}^t f_T(s) ds = \begin{cases} 0 & \text{se } t < 0 \\ \int_0^t \lambda e^{-\lambda s} ds = 1 - e^{-\lambda t} & \text{se } t \geq 0 \end{cases}$$

Per le variabili aleatorie assolutamente continue e le loro densità valgono proprietà analoghe a quelle delle variabili aleatorie discrete elencate nella Proposizione 2.2.4:

Proposizione 2.4.4 *Se f_X è la densità di una variabile aleatoria assolutamente continua X allora*

1. $\int_{\mathbb{R}} f_X(x) dx = 1$;
2. *se F_X è la funzione di ripartizione di X allora $f_X(x) = F'_X(x)$ per tutti gli $x \in \mathbb{R}$ tali che esiste $F'_X(x)$;*
3. *se $-\infty < a < b < +\infty$ allora*

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

Dimostrazione

1. Abbiamo che

$$1 = \lim_{x \rightarrow +\infty} F_X(x) = \lim_{x \rightarrow +\infty} \int_{-\infty}^x f_X(s) ds = \int_{\mathbb{R}} f_X(s) ds$$

2. È conseguenza del *teorema fondamentale del calcolo*.
3. Dal fatto che $P(X = x) = 0 \forall x$ in \mathbb{R} , segue che

$$P(X \in (a, b]) = P(\{X \in (a, b)\} \cup \{X = b\}) = P(X \in (a, b)) + P(X = b) = P(X \in (a, b))$$

Analogamente si dimostra che $P(X \in (a, b)) = P(X \in [a, b)) = P(X \in [a, b])$. Consideriamo ora l'intervallo $(a, b]$. Allora

$$\begin{aligned} P(X \in (a, b]) &= P(\{X \in (-\infty, b]\} \setminus \{X \in (-\infty, a]\}) \\ &= P(\{X \in (-\infty, b]\}) - P(\{X \in (-\infty, a]\}) \\ &= F_X(b) - F_X(a) = \int_{-\infty}^b f_X(x) dx - \int_{-\infty}^a f_X(x) dx \\ &= \int_a^b f_X(x) dx \quad \blacksquare \end{aligned}$$

Il punto 2. della Proposizione 2.4.4 può essere rafforzato opportunamente nel modo seguente:

Proposizione 2.4.5 *Sia X una variabile aleatoria ed F_X la sua funzione di ripartizione. Se F_X è continua ovunque, ed è derivabile con continuità per tutti gli $x \in \mathbb{R}$ eccetto al più in un insieme finito di punti, $B := \{x_1, \dots, x_n\} \subset \mathbb{R}$, allora X è una variabile aleatoria assolutamente continua e la funzione $f_X(x) = F'_X(x)$ per ogni $x \notin B$ e definita in modo arbitrario su B è una densità per X .*

Questo risultato ci dà un metodo operativo per riconoscere alcune variabili aleatorie assolutamente continue a partire dalla funzione di ripartizione e ci dice anche come calcolarne la densità.

Nota* 2.4.6 Si noti che la Proposizione 2.4.5 ci dice di calcolare $f_X(x)$ come $F'_X(x)$ per ogni $x \in \mathbb{R}$ eccetto un numero finito di punti B e di assegnarla in modo arbitrario sull'insieme B . Infatti il valore di $f_X(x)$ se $x \in B$ non è importante: possiamo definire $f_X(x)$ come vogliamo oppure non definirla affatto. Infatti nella Definizione 2.4.12 abbiamo visto che f è la densità di X assolutamente continua se $F_X(x) = \int_{-\infty}^x f(s) ds$. Ma se g è un'altra funzione tale che $g(x) = f(x)$ per ogni $x \in \mathbb{R}$ eccetto che in un numero finito di punti, è chiaro che

$$F_X(x) = \int_{-\infty}^x f(s) ds = \int_{-\infty}^x g(s) ds$$

quindi sia f che g sono densità di X ! Questa non univocità può sorprendere in un primo momento, ma è assolutamente inoffensiva dal punto di vista delle applicazioni. Essa può essere risolta matematicamente, cosa che noi non faremo, dando una definizione più generale del concetto di funzione.

Il punto 3. della Proposizione 2.4.4 può essere opportunamente rafforzato nel seguente:

Corollario 2.4.7 *Sia X una variabile aleatoria assolutamente continua con densità f_X e $B \subset \mathbb{R}$ tale che $B = B_1 \cup B_2 \cup \dots$ dove i B_k , $k = 1, 2, \dots$ sono intervalli disgiunti. Allora*

$$P(X \in B) = \int_B f_X(x) dx = \sum_{k=1}^{+\infty} \int_{B_k} f_X(x) dx$$

Esercizio* 2.4.8 Dimostrare il Corollario 2.4.7.

Nota* 2.4.9 Al lettore più attento verrà naturale chiedersi se il Corollario 2.4.7 possa essere generalizzato ad un insieme arbitrario B , se cioè è vero che

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

per ogni $B \subset \mathbb{R}$. La risposta a questa domanda è non banale³ e fuori dalla portata di questo corso. D'altro canto, chi ci garantisce che per un insieme arbitrario B , $\{X \in B\}$ sia un evento?

L'annunciata similitudine operativa tra variabili aleatorie assolutamente continue e variabili aleatorie discrete risiede proprio nel fatto che, se X è una variabile assolutamente continua, allora $P(X \in B)$ si calcola facendo l'integrale su B della densità, mentre, se X è discreta, si calcola $P(X \in B)$ facendo una somma sugli elementi di B (vedi punto 4. della Proposizione 2.2.4). Ritroveremo questa similitudine anche più avanti.

³Dipende dalla teoria degli insiemi che stiamo usando!

Nota 2.4.10 Come abbiamo fatto nella Nota 2.2.7 ci interessa evidenziare la motivazione euristica della parola “densità”. Supponiamo che f_X sia la densità di una variabile aleatoria assolutamente continua X , questo significa che f_X attribuisce un numero $f_X(x)$ ad ogni $x \in \mathbb{R}$. Analogamente a quanto fatto per le variabili aleatorie discrete, possiamo immaginare l’asse reale come un materiale inhomogeneo, in cui la densità di massa è f_X , cioè la massa del segmento *infinitesimo* $(x, x + dx)$ è $f_X(x)dx$.

Nota* 2.4.11 Anche in questo caso, prima di vedere alcuni esempi importanti di densità di variabili aleatorie assolutamente continue, torniamo al punto 1. della Proposizione 2.4.4. Analogamente al caso discreto, data una funzione integrabile $f(x) \geq 0$ che verifica la proprietà 1. della Proposizione 2.4.4 è possibile costruire uno spazio di probabilità (Ω, \mathcal{F}, P) ed una variabile aleatoria X su di esso che ha $f(x)$ come densità, cioè tale che $f_X(x) = f(x)$. Questo, come già osservato per le variabili aleatorie discrete, ci permetterà di parlare di variabili aleatorie assegnandone la densità.

La precedente osservazione giustifica la seguente definizione.

Definizione* 2.4.12 Una funzione $f : \mathbb{R} \rightarrow \mathbb{R}$ è una densità su \mathbb{R} se

1. $f(x)$ è integrabile, $f(x) \geq 0$ per ogni $x \in \mathbb{R}$;
2. $\int_{\mathbb{R}} f(x) dx = 1$.

2.5 Esempi di densità continue notevoli

In questo paragrafo elenchiamo alcune delle densità continue più importanti per le applicazioni.

2.5.1 Densità uniforme continua

Sia X un punto “scelto a caso” in $[0, 1]$. Ci chiediamo che tipo di variabile aleatoria sia X . Ovviamente da un punto di vista formale la domanda è mal posta, ma tuttavia nella sua accezione più immediata si può pensare che se un punto è scelto a caso in $[0, 1]$ la probabilità che questo sia più piccolo o uguale ad $1/2$ sia $1/2$. Questo perché $[0, 1] = [0, 1/2] \cup (1/2, 1]$,

$$P(X \in [0, 1/2]) = P(X \in (1/2, 1]) \quad \text{e} \quad P(X \in [0, 1/2]) + P(X \in (1/2, 1]) = 1$$

Possiamo ripetere il precedente ragionamento dividendo $[0, 1]$ nei quattro intervalli $[0, 1/4]$, $(1/4, 1/2]$, $(1/2, 3/4]$, $(3/4, 1]$ e affermare che la probabilità che X appartenga ad uno fissato di essi sia $1/4$. Questo implica anche che

$$P(X \leq 1/4) = P(X \in [0, 1/4]) = \frac{1}{4}$$

$$P(X \leq 1/2) = P(X \in [0, 1/4]) + P(X \in (1/4, 1/2]) = \frac{1}{2}$$

$$P(X \leq 3/4) = P(X \in [0, 1/4]) + P(X \in (1/4, 1/2]) + P(X \in (1/2, 3/4]) = \frac{3}{4}$$

Se si continua questo ragionamento, suddividendo $[0, 1]$ in 8, 16, 32, ... sottointervalli ci si convince che $P(X \leq x) = x$, per $x \in [0, 1]$. Inoltre, poiché X è un numero in $[0, 1]$, abbiamo che $P(X \leq x) = 0$ se $x < 0$ e $P(X \leq x) = 1$ se $x \geq 1$. Ne segue che la funzione di ripartizione di X è

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

che è funzione derivabile con continuità tranne nei punti 0 e 1. Segue dalla Proposizione 2.4.5 che X è una variabile aleatoria assolutamente continua e la sua densità si ottiene derivando la funzione di ripartizione:

$$f'_X(x) = \begin{cases} 0 & \text{se } x < 0 \\ 1 & \text{se } 0 < x < 1 \\ 0 & \text{se } x > 1 \end{cases}$$

Pertanto $f_X = \mathbf{1}_{(0,1)}$ o anche $f_X = \mathbf{1}_{[0,1]}$. Tale densità è detta *densità uniforme continua sull'intervallo* $[0, 1]$, la variabile aleatoria X è detta *uniforme su* $[0, 1]$ e si scrive $X \sim \mathcal{U}(0, 1)$.

2.5.2 Densità esponenziale

La densità esponenziale è l'analogo continuo della densità geometrica. Supponiamo di avere un'apparecchiatura *non soggetta ad usura* ed inizialmente funzionante, ma che si può guastare per motivi contingenti. Sia T l'istante, in minuti secondi, in cui l'apparecchiatura si guasta. La probabilità che l'apparecchiatura sia ancora funzionante dopo s secondi è $P(T > s)$. Quindi, se $s \leq 0$, allora $P(T > s) = 1$. Supponiamo ora $s > 0$. Osserviamo che, se l'apparecchiatura è funzionante al tempo $t > 0$, allora la probabilità che l'apparecchiatura sia ancora funzionante dopo s secondi, cioè al tempo $t + s$, è $P(T > s)$. Infatti, per l'assenza di usura, la probabilità che l'apparecchiatura non si guasti nell'intervallo di tempo $(t, t + s]$, se l'apparecchiatura funziona al tempo t , è uguale alla probabilità che l'apparecchiatura non si guasti nell'intervallo di tempo $(0, s]$. In formule $P(T > t + s | T > t) = P(T > s)$. Ma allora

$$P(T > s) = \frac{P(T > t + s, T > t)}{P(T > t)} = \frac{P(T > t + s)}{P(T > t)}$$

e quindi $P(T > t + s) = P(T > t)P(T > s)$. Se definiamo $\bar{F}(t) := P(T > t)$, per ogni $t \geq 0$, abbiamo che

$$\bar{F}(t + s) = \bar{F}(t)\bar{F}(s) \quad \forall t, s > 0$$

Una funzione⁴ che verifica questa *equazione funzionale* è $e^{\alpha t}$, dove $\alpha \in \mathbb{R}$. Quindi $P(T > t) = e^{\alpha t}$ e $P(T \leq t) = 1 - e^{\alpha t}$ per $t \geq 0$. Inoltre, poiché $P(T \leq t) \leq 1$, allora necessariamente $\alpha \leq 0$ e, per evitare situazioni banali, $\alpha < 0$. Quindi la funzione di ripartizione di

⁴In realtà l'unica funzione continua.

T è data da:

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

Sempre per la Proposizione 2.4.5 sappiamo che T è una variabile aleatoria assolutamente continua e la sua densità si ottiene derivando la funzione di ripartizione:

$$f'_T(t) = \begin{cases} 0 & \text{se } t < 0 \\ \lambda e^{-\lambda t} & \text{se } t > 0 \end{cases}$$

Pertanto $f_T(t) = \lambda e^{-\lambda t} \mathbf{1}_{[0, +\infty)}(t)$ è una densità per T . Questa densità è detta *densità esponenziale di parametro λ* e la variabile aleatoria T è detta *variabile esponenziale di parametro λ* . Si scrive anche $T \sim \mathcal{E}(\lambda)$.

La Figura 2.6 mostra l'andamento di densità e funzione di ripartizione $\mathcal{E}(\lambda)$ al variare di λ : al diminuire di λ aumenta la probabilità che la variabile aleatoria esponenziale assuma valori grandi.

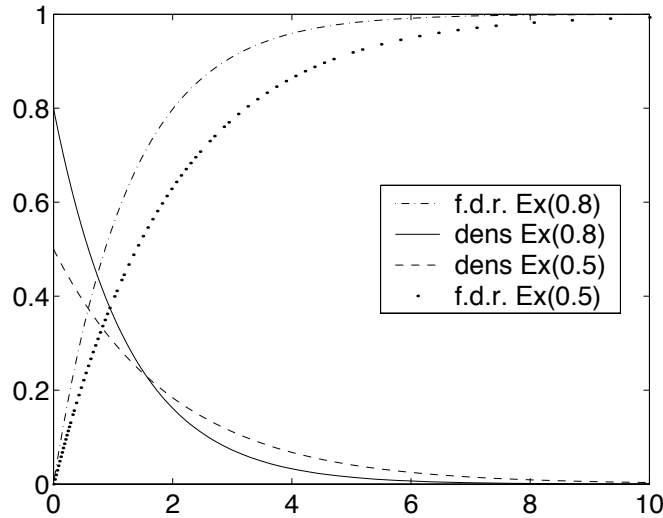


Figura 2.6: Densità e funzione di ripartizione $\mathcal{E}(\lambda)$

2.5.3 Densità gaussiana standard

In molte librerie dei più diffusi linguaggi di programmazione (C, Fortran, R) è disponibile la “funzione degli errori” (“error function” o “error integral”) erf:

$$\text{erf}(u) = \frac{2}{\sqrt{\pi}} \int_0^u e^{-y^2} dy$$

Tale funzione fornisce i valori di probabilità legate ad una particolare variabile aleatoria assolutamente continua detta *gaussiana standard*. La variabile gaussiana fornisce un utile

modello probabilistico per gli errori che si commettono per esempio nei procedimenti di misurazione. Il ruolo fondamentale in probabilità della densità gaussiana standard sarà più chiaro quando verrà presentato il “Teorema centrale del limite”. Per ora limitiamoci a definirla e a descriverne qualche proprietà.

Definizione 2.5.1 *Una variabile aleatoria assolutamente continua Z definita su uno spazio di probabilità (Ω, \mathcal{F}, P) è detta avere densità gaussiana standard (e scriveremo $Z \sim \mathcal{N}(0, 1)$) se ha densità*

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \quad \forall x \in \mathbb{R}$$

Esercizio* 2.5.2 Dimostrare che φ è una densità di probabilità continua, cioè soddisfa le proprietà 1. e 2. della Definizione 2.4.12.

Trovate in Figura 2.7(a) il grafico della funzione φ che ha andamento a campana con punto di massimo in 0 ed è simmetrico rispetto all’asse delle ordinate (ossia φ è funzione “pari”, cioè $\varphi(-z) = \varphi(z) \forall z > 0$). In termini di φ , la probabilità dell’evento $\{-z < Z < z\}$ è

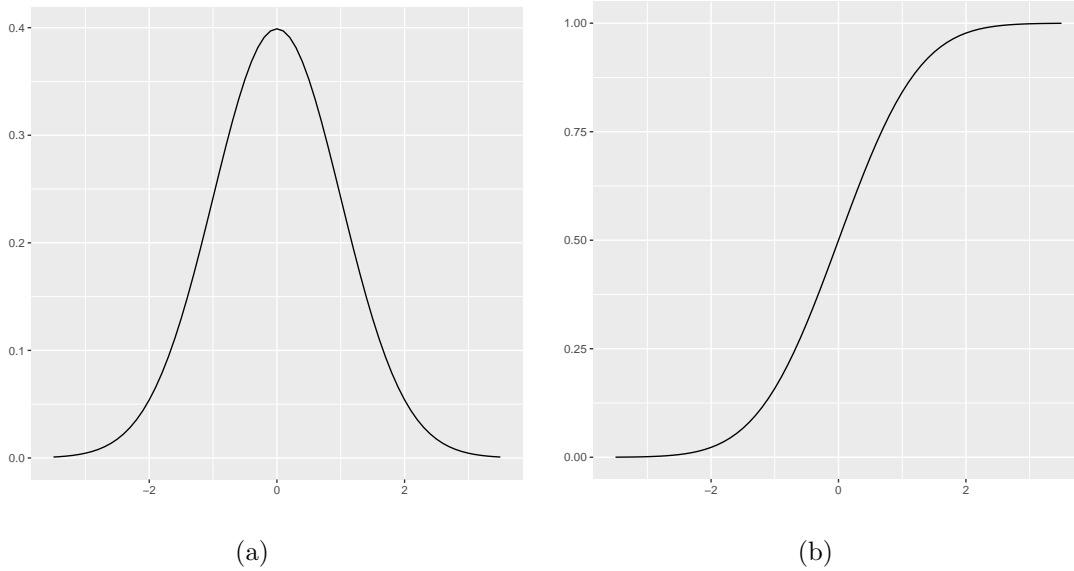


Figura 2.7: Densità di probabilità in 2.7 (a) e funzione di ripartizione in 2.7 (b) $\mathcal{N}(0, 1)$

$$P(-z < Z < z) = \int_{-z}^z \varphi(x) dx = \frac{1}{\sqrt{2\pi}} \int_{-z}^z e^{-x^2/2} = \frac{2}{\sqrt{2\pi}} \int_0^z e^{-x^2/2} dx$$

Posto $y = x/\sqrt{2}$ e operando il cambio di variabile nell’integrale si ottiene

$$P(-z < Z < z) = \frac{2}{\sqrt{\pi}} \int_0^{z/\sqrt{2}} e^{-y^2} dy = \text{erf}(z/\sqrt{2}) \quad \forall z > 0$$

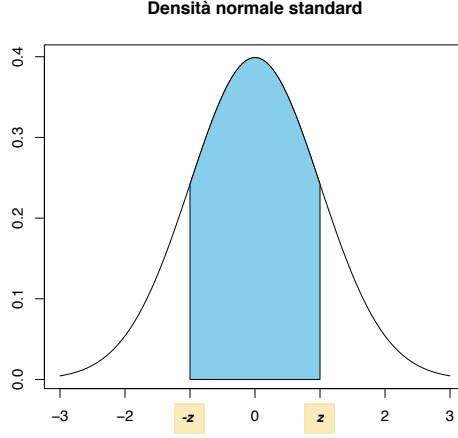


Figura 2.8: Densità normale standard; l'area celeste rappresenta $P(-z < Z < z)$.

Quindi la funzione erf fornisce la probabilità che una variabile aleatoria gaussiana standard assuma valori in un intervallo simmetrico rispetto all'origine. Dal significato dell'operazione di integrazione, segue che graficamente $P(-z < Z < z)$ è rappresentata dall'area colorata di celeste tra le due linee nere in Figura 2.8.

La funzione di ripartizione di una variabile aleatoria $\mathcal{N}(0, 1)$ rappresentata in Figura 2.7 (b) viene indicata di solito con $\Phi(z)$:

$$\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z e^{-x^2/2} dx$$

Non è possibile calcolare Φ analiticamente, ma troverete Φ tabulata in quasi tutti i libri di probabilità. Tipicamente sono tabulati i valori di $\Phi(z)$ per $z \geq 0$. Se $z < 0$, $\Phi(z)$ si può ottenere usando la seguente formula

$$\Phi(z) = 1 - \Phi(-z) \quad \forall z \in \mathbb{R} \quad (2.5.1)$$

La formula 2.5.1 deriva dalla simmetria di φ nel seguente modo:

$$\begin{aligned} \Phi(z) &= 1 - P(Z > z) = 1 - P(Z > z) - P(Z = z) \\ &= 1 - P(Z \geq z) = 1 - P(Z \leq -z) = 1 - \Phi(-z) \end{aligned}$$

In particolare, per $z = 3$,

$$P(|Z| \geq 3) = 1 - P(-3 \leq Z \leq 3) = 1 - [\Phi(3) - \Phi(-3)] = 2(1 - \Phi(3)) \simeq 0.0027$$

cioè $Z \sim \mathcal{N}(0, 1)$ ha probabilità trascurabile di assumere valori all'esterno dell'intervallo $[-3, 3]$. Tutte queste proprietà forniscono una parziale giustificazione al fatto che la densità gaussiana venga usata come modello probabilistico per gli errori.

2.6 Funzioni di variabili aleatorie

Sappiamo che una variabile aleatoria va pensata come un numero casuale. Ora se X è una variabile aleatoria e $g: \mathbb{R} \rightarrow \mathbb{R}$ è una funzione, allora $Y := g(X)$ è ancora un numero causale. Precisamente $Y := g(X)$ è il numero che si ottiene applicando la funzione g al numero casuale X . Per avere un esempio concreto, se $X \sim \mathcal{U}(0, 1)$, allora $Y := \pi X^2$ indica l'area di un cerchio per il quale la lunghezza del raggio è “scelta a caso” in $(0, 1)$, ovvero Y è l'area di un cerchio “scelto a caso” tra i cerchi di raggio più piccolo di 1. In questo caso $g(x) = \pi x^2$.

È naturale ora chiedersi se $Y = g(X)$ è una variabile aleatoria nel senso della Definizione 2.1.2, cioè ci chiediamo se è *sempre* vero che, se X è una variabile aleatoria definita su (Ω, \mathcal{F}, P) , allora $\{\omega \in \Omega: g[X(\omega)] \leq x\} \in \mathcal{F}$ per ogni $x \in \mathbb{R}$. Questo non è vero in generale, ma nei casi importanti per le applicazioni che tratteremo è sempre vero. Per esempio è vero se g è una funzione continua a tratti. Ci interesserà in particolare capire se e come sia possibile determinare la densità della variabile aleatoria $Y := g(X)$ a partire da X . Prima di procedere vediamo qualche ulteriore esempio.

Esempio 2.6.1 In molti procedimenti industriali è routine monitorare il livello di danni o fattori indesiderati. Per esempio rappresentiamo con X il numero di batteri in un campione di liquido preso da un bacino di lavorazione alimentare. Se X supera un livello critico c , il procedimento viene arrestato e si attua una procedura di rinnovo e pulizia del sistema di depurazione. Definiamo

$$Y := \begin{cases} 1 & \text{se } X \geq c \\ 0 & \text{se } X < c \end{cases}$$

Allora $Y = 1$ se e solo se il processo produttivo viene arrestato. La variabile aleatoria Y è funzione della variabile aleatoria X e si può scrivere $Y = g(X)$ dove $g(x) = \mathbf{1}_{[c, +\infty)}(x)$. Poiché Y assume solo i valori 0 e 1, Y è una variabile aleatoria di Bernoulli di parametro $p = P(Y = 1)$, con

$$P(Y = 1) = P(X \geq c) = \sum_{k \geq c} p_X(k)$$

Quindi

$$p_Y(h) = \begin{cases} 1 - \sum_{k \geq c} p_X(k) & \text{se } h = 0 \\ \sum_{k \geq c} p_X(k) & \text{se } h = 1 \\ 0 & \text{se } h \notin \{0, 1\} \end{cases}$$

cioè la densità di Y è calcolabile a partire dalla densità di X .

Esempio 2.6.2 Sia T la variabile aleatoria che denota la temperatura in una stanza climatizzata. Se $T < a$ l'impianto di condizionamento riscalda. Se $T > b$ refrigera. Altrimenti, si spegne. Quindi lo stato dell'impianto di condizionamento, in funzione della temperatura, può essere descritto mediante una variabile aleatoria S che assume valore 1 se l'impianto refrigera, 0 se è spento e -1 se riscalda, cioè:

$$S = \begin{cases} -1 & \text{se } T < a \\ 0 & \text{se } a \leq T \leq b \\ 1 & \text{se } T > b \end{cases}$$

Volendo calcolare per esempio la probabilità che l'impianto sia spento, cioè $P(S = 0)$, possiamo procedere nel seguente modo:

$$P(S = 0) = P(a \leq T \leq b) = \begin{cases} \sum_{a \leq t \leq b} p_T(t) & \text{se } T \text{ è discreta} \\ \int_a^b f_T(t) dt & \text{se } T \text{ è assolutamente continua} \end{cases}$$

Quello che i precedenti esempi evidenziano è che se conosciamo la densità di una variabile aleatoria X è possibile (in alcuni casi) determinare la densità di $Y := g(X)$.

Per essere più specifici inizialmente supponiamo che X sia una variabile aleatoria discreta, con densità $p_X(x)$ e $P(X \in S) = 1$, dove $S = \{x_k : k \in I\}$, $I \subset \mathbb{Z}$. Siano $g: S \rightarrow \mathbb{R}$ e $g(S) = \{g(x) : x \in S\}$. Se definiamo $Y := g(X)$, allora Y assume valori in $g(S)$, in particolare $P(X \in S) = 1$ implica che $P(Y \in g(S)) = 1$, cioè Y è una variabile aleatoria discreta e la sua densità è nulla se $y \notin g(S)$. Inoltre se $y \in g(S)$ abbiamo

$$\begin{aligned} P(Y = y) &= P(g(X) = y) = P\left(\bigcup_{k: g(x_k)=y} \{X = x_k\}\right) = \\ &= \sum_{k: g(x_k)=y} P(X = x_k) = \sum_{k: g(x_k)=y} p_X(x_k) \end{aligned}$$

In definitiva vale la seguente proposizione.

Proposizione 2.6.3 *Sia X una variabile aleatoria discreta, con densità $p_X(x)$ e $P(X \in S) = 1$, dove $S = \{x_k : k \in I\}$, $I \subset \mathbb{Z}$. Sia $g: S \rightarrow \mathbb{R}$ e sia $g(S) = \{g(x) : x \in S\}$. Se definiamo $Y := g(X)$, allora Y è una variabile aleatoria discreta a valori in $g(S)$, cioè $P(Y \in g(S)) = 1$, e la sua densità è*

$$p_Y(y) = \begin{cases} \sum_{k: g(x_k)=y} p_X(x_k) & \text{se } y \in g(S) \\ 0 & \text{se } y \notin g(S) \end{cases}$$

Esempio 2.6.4 La probabilità di vincere giocando a una slot machine è $p = 0.2$ e per partecipare a n giocate si paga una posta iniziale di n €. Se si effettuano 10 giocate e ad ogni giocata o si totalizza 0 o si vincono 2 €, qual è la probabilità di vincere 4 € (al netto della posta iniziale)?

Siano X la variabile aleatoria che indica il numero di vittorie su 10 giocate e Y quella che indica la vincita accumulata dopo 10 giocate. Allora $X \sim \mathbf{Bi}(10, 0.2)$ e $Y = 2X - 10$. Inoltre, la densità di probabilità di Y è

$$p_Y(k) = \begin{cases} \binom{10}{\frac{10+k}{2}} 0.2^{\frac{10+k}{2}} 0.8^{\frac{10-k}{2}} & k = 0, \pm 2, \pm 4, \pm 6, \pm 8, \pm 10 \\ 0 & \text{altrove} \end{cases}$$

In particolare:

$$P(\text{"vincere 4 €"}) = P(Y = 4) = \binom{10}{\frac{10+4}{2}} 0.2^{\frac{10+4}{2}} \times 0.8^{\frac{10-4}{2}} = \binom{10}{7} 0.2^7 \times 0.8^3 \simeq 0.0008.$$

La Proposizione 2.6.3 afferma che, se X è una variabile aleatoria discreta, allora $g(X)$ è una variabile aleatoria discreta e la sua densità è univocamente determinata dalla densità di X . Questo fatto implica, tra l'altro, che se X e W sono due variabili aleatorie discrete che hanno la stessa densità, lo stesso vale per $g(X)$ e $g(W)$. Inoltre la proposizione mostra un metodo per calcolare la densità di $g(X)$ a partire dalla densità di X .

Nel caso di variabili aleatorie assolutamente continue vale un risultato analogo, questa volta però sotto ipotesi restrittive su g :

Proposizione 2.6.5 *Sia $Y = g(X)$, con X variabile aleatoria assolutamente continua con densità f_X . Supponiamo che esista un intervallo aperto $S \subset \mathbb{R}$ tale che: $P(X \in S) = 1$, g sia differenziabile con continuità su S e $g'(x) \neq 0$ per ogni $x \in S$. Sia g^{-1} la funzione inversa di g e $g(S) = \{g(x) : x \in S\}$. Allora Y è una variabile aleatoria assolutamente continua con densità data da*

$$f_Y(y) = \begin{cases} f_X(g^{-1}(y)) |(g^{-1})'(y)| & \text{se } y \in g(S) \\ 0 & \text{se } y \notin g(S) \end{cases} \quad (2.6.1)$$

Non vedremo la dimostrazione di questa proposizione in generale, ma osserviamo che nel caso particolare in cui $S = \mathbb{R} = g(S)$ e g è crescente si ha:

$$F_Y(y) = P(g(X) \leq y) = P(X \leq g^{-1}(y)) = \int_{-\infty}^{g^{-1}(y)} f_X(s) ds = \int_{-\infty}^y f_X(g^{-1}(t)) (g^{-1})'(t) dt$$

dove, nell'ultima uguaglianza, abbiamo utilizzato il cambiamento di variabile $t = g(s)$. Quindi

$$f_Y(y) = F_Y'(y) = f_X(g^{-1}(y)) (g^{-1})'(y)$$

Esempio 2.6.6 (Densità uniforme su un intervallo) Sia $X \sim \mathcal{U}(0, 1)$, $\alpha, \beta \in \mathbb{R}$ e $g(x) := \alpha x + \beta$. Ci chiediamo che tipo di variabile aleatoria è $Y := g(X) = \alpha X + \beta$. Innanzi tutto osserviamo che se $\alpha = 0$, allora $Y \equiv \beta$, cioè Y è la variabile aleatoria degenerare che vale sempre β . Se $\alpha \neq 0$, possiamo utilizzare la Proposizione 2.6.5, infatti in questo caso $g'(x) = \alpha \neq 0$, $g^{-1}(y) = (y - \beta)/\alpha$ e si ha $|(g^{-1})'(y)| = 1/|\alpha|$. Per la Proposizione 2.6.5 Y è assolutamente continua e ha densità data da

$$f_Y(y) = f_X\left(\frac{y - \beta}{\alpha}\right) \frac{1}{|\alpha|} \mathbf{1}_{g((0,1))}(y) = \frac{1}{|\alpha|} \mathbf{1}_{g(0,1)}(y)$$

Se $\alpha > 0$ allora $g(0, 1) = (\beta, \alpha + \beta)$. Infatti

$$y \in g(0, 1) \iff g^{-1}(y) \in (0, 1) \iff 0 < \frac{y - \beta}{\alpha} < 1 \iff \beta < y < \alpha + \beta$$

Quindi, $f_Y(y) = (1/\alpha) \mathbf{1}_{(\beta, \alpha + \beta)}(y)$. Osservando la definizione di Y si vede che Y si ottiene da X mediante una dilatazione di fattore α seguita da una traslazione di ragione β . Questa trasformazione fa corrispondere all'intervallo $(0, 1)$ l'intervallo $(\beta, \alpha + \beta)$, quindi risulta intuitivo che se X è un numero scelto a caso in $(0, 1)$ allora Y è un numero scelto a caso in $(\beta, \alpha + \beta)$. Ora, per $a, b \in \mathbb{R}$ con $a < b$, poniamo $\beta = a$ e $\alpha = b - a$ otteniamo che un numero scelto a caso in (a, b) è una variabile aleatoria Y con densità

$$f_Y(y) = \frac{1}{b - a} \mathbf{1}_{(a, b)}(y) \quad (2.6.2)$$

Se $\alpha < 0$, analoghi ragionamenti portano a dire che $f_Y(y) = -\frac{1}{\alpha} \mathbf{1}_{(\beta + \alpha, \beta)}(y)$.

L'esempio precedente ci porta ad una generalizzazione della densità uniforme vista nel Paragrafo 2.5.1. Una variabile aleatoria Y assolutamente continua è detta *uniforme su (a, b)* se la sua densità è data da (2.6.2). La densità f_Y è detta *densità uniforme su (a, b)* e si può scrivere $Y \sim \mathcal{U}(a, b)$.

Esercizio 2.6.7 Mostrare che se $X \sim \mathcal{U}(a, b)$, allora $\frac{X - a}{b - a} \sim \mathcal{U}(0, 1)$.

Esercizio* 2.6.8 Verificare che f_Y definita dalla (2.6.2) è una densità, cioè valgono 1. e 2. della Definizione 2.4.12.

Esercizio 2.6.9 Sia $X \sim \mathcal{U}(a, b)$: determinare e disegnare F_X .

Osserviamo che se $X \sim \mathcal{U}(a, b)$, allora la probabilità che X cada in un intervallo $[c, d]$ con $a \leq c < d \leq b$ è *proporzionale* alla lunghezza di $[c, d]$ con costante di proporzionalità data da $(b - a)^{-1}$, cioè

$$P(c < Y < d) = \frac{d - c}{b - a}$$

Esempio 2.6.10 Sia $X \sim \mathcal{E}(\lambda)$ e $a > 0$, allora $Y = aX \sim \mathcal{E}(\frac{\lambda}{a})$. Infatti per la (2.6.1)

$$f_Y(y) = f_X\left(\frac{y}{a}\right) \frac{1}{a} = \lambda e^{-\lambda \frac{y}{a}} \mathbf{1}_{(0,+\infty)}\left(\frac{y}{a}\right) \frac{1}{a} = \frac{\lambda}{a} e^{-\frac{\lambda}{a} y} \mathbf{1}_{(0,+\infty)}(y)$$

Nota 2.6.11 L'esempio precedente mostra che la famiglia delle variabili aleatorie esponenziali è “chiusa” rispetto all'operazione di cambiamento di scala (passaggio da X ad aX con $a > 0$).

Esempio 2.6.12 Sia $X \sim \mathcal{N}(0, 1)$ e $Y := X^2$. Vogliamo capire se Y è variabile aleatoria assolutamente continua e, se la risposta è positiva, determinarne la densità. In questo caso non possiamo applicare direttamente la Proposizione 2.6.5 in quanto $S = \mathbb{R}$ e $g(x) = x^2$ non è biettiva. Procediamo direttamente a scrivere la funzione di ripartizione F_Y e vedere se ammette densità. Osserviamo innanzi tutto che se $y < 0$ allora $F_Y(y) = P(Y \leq y) = P(X^2 \leq y) = 0$ semplicemente perché $X^2 \geq 0$. Se invece $y \geq 0$ allora

$$\begin{aligned} F_Y(y) &= P(Y \leq y) = P(X^2 \leq y) = P(-\sqrt{y} \leq X \leq \sqrt{y}) = \\ &= P(X \leq \sqrt{y}) - P(X < -\sqrt{y}) = F_X(\sqrt{y}) - F_X(-\sqrt{y}) \end{aligned}$$

Quindi per la Proposizione 2.4.5 abbiamo che se $y > 0$ allora

$$\begin{aligned} f_Y(y) &= F'_Y(y) = \frac{d}{dy} F_X(\sqrt{y}) - \frac{d}{dy} F_X(-\sqrt{y}) = F'_X(\sqrt{y}) \frac{1}{2\sqrt{y}} + F'_X(-\sqrt{y}) \frac{1}{2\sqrt{y}} = \\ &= \frac{1}{2\sqrt{y}} [F'_X(\sqrt{y}) + F'_X(-\sqrt{y})] = \frac{1}{2\sqrt{y}} [f_X(\sqrt{y}) + f_X(-\sqrt{y})] \end{aligned}$$

Sostituendo la densità gaussiana standard in questa formula otteniamo

$$f_Y(y) = \frac{e^{-y/2}}{\sqrt{y}2\pi}$$

In definitiva

$$f_Y(y) = \frac{e^{-y/2}}{\sqrt{y}2\pi} \mathbf{1}_{(0,+\infty)}(y)$$

Una variabile aleatoria Y con questa densità è detta variabile aleatoria chi-quadrato con 1 grado di libertà e si scrive $X \sim \chi^2(1)$.

Nota 2.6.13 Si noti che quanto visto sopra ci dice che per ogni variabile aleatoria X assolutamente continua e con densità f_X , allora $Y := X^2$ è assolutamente continua e si ha

$$f_Y(y) = \frac{1}{2\sqrt{y}} [f_X(\sqrt{y}) + f_X(-\sqrt{y})] \mathbf{1}_{(0,+\infty)}(y) \quad (2.6.3)$$

Esempio 2.6.14 Siano $X \sim \mathcal{U}(0, 1)$ e $F(y) = (1 - e^{-\lambda y})\mathbf{1}_{(0, +\infty)}(y)$ con $\lambda > 0$ (F è la funzione di ripartizione esponenziale di parametro λ). Introduciamo la funzione $g : (0, 1) \rightarrow \mathbb{R}$ definita da

$$g(x) := \frac{-\log(1-x)}{\lambda} \quad \forall x \in (0, 1)$$

Allora g è una funzione iniettiva tale che

$$g^{-1}(x) = 1 - e^{-\lambda x} = F(x) \quad \forall x \in (0, \infty)$$

e $Y := g(X) \sim \mathcal{E}(\lambda)$. Infatti, per la Proposizione 2.6.5 si ha che Y è assolutamente continua e la sua densità è data da

$$f_Y(y) = f_X(F(y))|F'(y)|\mathbf{1}_{g((0,1))}(y) = \lambda e^{-\lambda y}\mathbf{1}_{(0, +\infty)}(y)$$

2.6.1 *Cenno alla simulazione di variabili aleatorie

L'Esempio 2.6.14 mostra la possibilità di rappresentare una variabile aleatoria esponenziale come una trasformazione di una variabile aleatoria $\mathcal{U}(0, 1)$, mediante la funzione di ripartizione stessa. Un'importante conseguenza di questo risultato è che per generare una variabile aleatoria $\mathcal{E}(\lambda)$ è sufficiente generare $U \sim \mathcal{U}(0, 1)$ e poi calcolare $-\log(1-U)/\lambda$. Tutto ciò è facilmente attuabile perchè nelle librerie dei linguaggi di programmazione esistono routine che generano valori di variabili “pseudo-aleatorie” e $\mathcal{U}(0, 1)$. Il risultato di rappresentare una variabile aleatoria come una trasformazione di una $\mathcal{U}(0, 1)$ non riguarda soltanto le variabili aleatorie esponenziali, ma anche tutte le altre, siano esse discrete o continue. Qui accenniamo soltanto al risultato per il caso delle variabili aleatorie che hanno funzione di ripartizione $F(x)$ strettamente crescente sull'insieme $\{x: 0 < F(x) < 1\}$. In tal caso, l'equazione $F(x) = u$ ammette un'unica soluzione per ogni $u \in (0, 1)$, cioè $x = F^{-1}(u)$.

Proposizione 2.6.15 (della trasformata integrale) Sia F una funzione di ripartizione strettamente crescente sull'insieme $\{x: 0 < F(x) < 1\}$ e F^{-1} la funzione definita da $F(F^{-1}(u)) = u$ per ogni $u \in (0, 1)$. Se $U \sim \mathcal{U}(0, 1)$ allora $X = F^{-1}(U)$ ha funzione di ripartizione F .

Dimostrazione Poiché F è funzione strettamente crescente su $\{x: 0 < F(x) < 1\}$ e per ogni $u \in (0, 1)$ $F(F^{-1}(u)) = u$ allora, per ogni $u \in (0, 1)$ e per ogni $x \in F^{-1}((0, 1))$ vale che $F^{-1}(u) \leq x$ se e solo se $u = F(F^{-1}(u)) \leq F(x)$ e quindi

$$F_X(x) = P(F^{-1}(U) \leq x) = P(U \leq F(x)) = F(x) \quad \blacksquare$$

Il precedente lemma dà una prima idea del perché la densità $\mathcal{U}(0, 1)$ giochi un ruolo chiave nelle simulazioni: teoricamente, per generare una qualunque variabile aleatoria continua X avente funzione di ripartizione F invertibile, potremmo procedere a generare $U \sim \mathcal{U}(0, 1)$ e fare la trasformazione $F^{-1}(U)$. Praticamente, questo metodo si applica soltanto nei casi in cui siamo in grado di determinare esplicitamente F^{-1} . Ma questi casi sono pochi. Ad esempio, rimane fuori la variabile aleatoria gaussiana standard. Nei casi non coperti dall'uso della *trasformata integrale* potremo procedere o con metodi generali alternativi e validi per diverse famiglie di variabili aleatorie, o con metodi *ad hoc* che usano in modo pesante proprietà specifiche delle variabili aleatorie da simulare. Il lettore interessato veda ad esempio [6].

Esercizio 2.6.16 Scrivete un programma in un linguaggio a voi noto per ottenere $n = 100$ simulazioni dalla densità $\mathcal{U}(-2, 2)$.

Capitolo 3

Media varianza e momenti

Abbiamo visto nel Capitolo 2 che nella teoria assiomatica della probabilità i numeri causali vengono modellizzati dalle variabili aleatorie. In questo capitolo vedremo come sia possibile associare ad una variabile aleatoria alcune grandezze deterministiche, cioè alcuni numeri, che ci daranno informazioni qualitative e quantitative sulla variabile aleatoria. Per chiarire meglio la situazione facciamo subito un esempio.

Esempio 3.0.1 Supponiamo di giocare alla roulette puntando sul rosso¹. Sia X il guadagno netto che otteniamo puntando 1€. Si vede subito che X è una variabile discreta, che assume solo i due valori -1 (cioè abbiamo perso 1€) e 1 (cioè abbiamo vinto 1€), con probabilità $19/37$ e $18/37$ rispettivamente. In particolare si vede che il gioco della roulette è favorevole al banco, infatti la probabilità di vincere 1€ è più piccola di quella perdere 1€. Supponiamo ora invece di giocare a testa e croce con un amico. Se Y è il guadagno netto che otteniamo puntando 1€, allora anche in questo caso Y è una variabile discreta, che assume solo i due valori -1 e 1 , ma con probabilità $1/2$ questa volta. In questo caso la probabilità di vincere 1€ è uguale alla probabilità di perderlo, quindi questo gioco è in un certo senso più “giusto” della roulette, o come si dice è un gioco *equo*. Vedremo che è possibile associare a ciascuna delle variabili aleatorie X e Y un numero, chiamato *media*. Vedremo che la media di X è negativa (ad indicare che il guadagno netto medio al gioco della roulette è un numero negativo) mentre quella di Y è nulla (ad indicare che il guadagno netto medio a testa e croce è nullo).

3.1 Valore atteso (o media)

In questa sezione viene introdotta la nozione di media per variabili aleatorie discrete e assolutamente continue. Avendo a disposizione strumenti matematici più avanzati si potrebbe introdurre tale nozione per qualsiasi variabile aleatoria.

Per le variabili aleatorie discrete abbiamo:

¹Una roulette europea “onesta” è costituita da un congegno che seleziona casualmente un numero tra 37 disponibili, 18 dei quali sono rossi, 18 neri ed uno (lo “zero”) verde.

Definizione 3.1.1 Sia X una variabile aleatoria discreta a valori in $S = \{x_k : k \in I\}$ con $I \subset \mathbb{Z}$ e sia p_X la sua densità. Se

$$\sum_{k \in I} |x_k| p_X(x_k) < +\infty$$

si definisce media di X o valore atteso di X il numero

$$E(X) := \sum_{k \in I} x_k p_X(x_k)$$

altrimenti si dice che X non ammette valore atteso.

Prima di procedere con gli esempi, facciamo qualche osservazione sulla definizione appena data. Innanzitutto osserviamo che se X è una variabile aleatoria discreta, non è detto che X abbia valore atteso. Questo dipende dalla convergenza della “serie” $\sum_{k \in I} |x_k| p_X(x_k)$. Ovviamente se X assume un numero finito di valori, cioè se I è un insieme finito, questa serie diventa una somma finita che è sicuramente convergente e la media di X in questo caso esiste.

Osserviamo che se $\sum_{k \in I} |x_k| p_X(x_k) < +\infty$, allora per un noto teorema dell’analisi (convergenza assoluta \Rightarrow convergenza) segue che $\sum_{k \in I} x_k p_X(x_k)$ converge. Quindi $E(X)$ è un numero finito. Il motivo per il quale si richiede la convergenza assoluta, invece della semplice convergenza, è essenzialmente tecnico e non lo discuteremo in questo contesto.

Il valore atteso di una variabile aleatoria X è un oggetto legato alla densità p_X di X piuttosto che alla funzione che definisce la variabile aleatoria X . Questo significa che due variabili aleatorie con la stessa densità hanno lo stesso valore atteso (oppure non hanno valore atteso).

Se torniamo all’interpretazione della densità di una variabile aleatoria discreta, come densità di massa, vista nella Nota 2.2.7, abbiamo che il valore atteso di X può essere visto come il baricentro del sistema di masse descritto.

Esempio 3.1.2 (Segue Esempio 3.0.1) Calcoliamo i valori attesi di X ed Y . Per X otteniamo

$$E(X) = -1 \times \frac{19}{37} + 1 \times \frac{18}{37} = -\frac{1}{37}$$

mentre per Y abbiamo

$$E(Y) = -1 \times \frac{1}{2} + 1 \times \frac{1}{2} = 0$$

Esempio 3.1.3 Se X è uniforme su $\{1, \dots, n\}$, cioè $p_X(k) = 1/n$ per ogni $k = 1, \dots, n$, allora

$$E(X) = \frac{1 + \dots + n}{n} = \frac{n(n+1)}{2n} = \frac{n+1}{2}$$

Esempio 3.1.4 Se $X \sim \mathbf{Be}(p)$ allora

$$E(X) = 1 \times p + 0 \times (1-p) = p$$

Esempio 3.1.5 Se $X \sim \mathcal{P}(\lambda)$ allora la media esiste e vale:

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

Per le variabili aleatorie assolutamente continue vale una definizione di valore atteso analoga alla Definizione 3.1.1:

Definizione 3.1.6 Sia X una variabile aleatoria assolutamente continua e sia f_X la sua densità. Se

$$\int_{\mathbb{R}} |x| f_X(x) dx < +\infty$$

si definisce media di X o valore atteso di X il numero

$$E(X) := \int_{\mathbb{R}} x f_X(x) dx$$

altrimenti si dice che X non ammette valore atteso.

Anche in questo caso valgono osservazioni analoghe a quelle fatte dopo la Definizione 3.1.1, che ometteremo.

Esercizio 3.1.7 Fare le osservazioni che abbiamo omissso.

Esempio 3.1.8 Se $X \sim \mathcal{U}(0, 1)$ allora la media esiste e vale:

$$E(X) = \int_0^1 x dx = \frac{1}{2}$$

Esempio 3.1.9 Se $X \sim \mathcal{E}(\lambda)$ allora la media esiste e vale:

$$E(X) = \int_0^{+\infty} x \lambda e^{-\lambda x} dx = - \int_0^{+\infty} x \cdot \frac{d}{dx} e^{-\lambda x} dx = x e^{-\lambda x} \Big|_0^{+\infty} + \int_0^{+\infty} e^{-\lambda x} dx = \frac{1}{\lambda}$$

Esempio 3.1.10 Se $X \sim \mathcal{N}(0, 1)$ allora la media esiste e vale:

$$E(X) = \int_{-\infty}^{+\infty} \frac{x}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = 0$$

poiché la funzione integranda è dispari e l'insieme di integrazione è simmetrico.

Esercizio* 3.1.11 Si fornisca un esempio di variabile aleatoria X discreta che non ha media finita.

Soluzione Sia p definita da

$$p(x) = \begin{cases} \frac{1}{x(x+1)} & \text{se } x \in \{1, 2, \dots\} \\ 0 & \text{se } x \notin \{1, 2, \dots\} \end{cases}$$

p è una densità, infatti $p(x) \geq 0$ e

$$\sum_{x=1}^{+\infty} p(x) = \sum_{x=1}^{\infty} \frac{1}{x(x+1)} = \sum_{x=1}^{+\infty} \left(\frac{1}{x} - \frac{1}{x+1} \right) = 1$$

per la proprietà telescopica. Se quindi X è una variabile aleatoria di densità $p_X(x) = p(x)$, allora

$$\sum_{x=1}^{+\infty} |x| p_X(x) = \sum_{x=1}^{+\infty} \frac{x}{x(x+1)} = \sum_{x=1}^{+\infty} \frac{1}{x+1} = +\infty \quad \blacksquare$$

3.1.1 Valore atteso di funzioni di variabili aleatorie

Sia X una variabile aleatoria, g una funzione reale e $Y := g(X)$. Nella Sezione 2.6 abbiamo studiato la densità di Y . Qui ci poniamo il problema di calcolare $E(Y)$. Supponiamo per un momento che X sia discreta, allora mediante la Proposizione 2.6.3, possiamo calcolare la densità p_Y di Y e poi affermare che $E(Y) = \sum_y y p_Y(y)$. In realtà se siamo interessati solamente a $E(Y)$ e non alla densità di Y , possiamo evitare di determinare esplicitamente p_Y . Vale infatti la seguente proposizione:

Proposizione 3.1.12 *Siano X una variabile aleatoria discreta a valori in $S = \{x_k : k \in I\}$ con $I \subset \mathbb{Z}$ e densità p_X , g una funzione reale e $Y := g(X)$. Se $\sum_{k \in I} |g(x_k)| p_X(x_k) < +\infty$, allora Y ammette valore atteso e*

$$E(Y) = \sum_{k \in I} g(x_k) p_X(x_k). \quad (3.1.1)$$

Siano X una variabile aleatoria assolutamente continua con densità f_X e g una funzione reale tale che $Y := g(X)$ è una variabile aleatoria. Se $\int_{\mathbb{R}} |g(x)| f_X(x) dx < +\infty$, allora Y ammette valore atteso e

$$E(Y) = \int_{\mathbb{R}} g(x) f_X(x) dx. \quad (3.1.2)$$

Non dimostriamo questa proposizione, ma illustriamone il suo senso con qualche esempio.

Esempio 3.1.13 Siano $X \sim \mathcal{U}(-1, 1)$ e $Y = X^2$. Allora $f_Y(y) = 1/(2\sqrt{y}) \mathbf{1}_{(0,1)}(y)$ (per ottenerla si applichi la formula in (2.6.3)) e quindi

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

Con maggior economia di calcoli possiamo arrivare allo stesso risultato applicando la Proposizione 3.1.12:

$$E(Y) = \int_{-1}^1 x^2 \cdot \frac{1}{2} dx = \frac{2}{2} \int_0^1 x^2 dx = \frac{1}{3}$$

3.1.2 Proprietà del valore atteso

Nella seguente proposizione elenchiamo alcune proprietà che discendono direttamente dalla definizione di valore atteso.

Proposizione 3.1.14 *Sia X una variabile aleatoria definita sullo spazio di probabilità (Ω, \mathcal{F}, P) .*

1. *Se $P(X = c) = 1$ allora $E(X) = c$.*
2. *Se X è una variabile aleatoria e $B \subset \mathbb{R}$ tale che $\{X \in B\} \in \mathcal{F}$ allora $E(\mathbf{1}_B(X)) = P(X \in B)$.*
3. *Se X è una variabile aleatoria tale che $E(X)$ esiste e α è una costante, allora $E(\alpha X) = \alpha E(X)$.*
4. *Se X è una variabile aleatoria, g e h sono funzioni tali che $E(g(X))$ ed $E(h(X))$ esistono, allora $E(g(X) + h(X)) = E(g(X)) + E(h(X))$.*
5. *Se X è una variabile aleatoria tale che $P(X \geq 0) = 1$ e $E(X)$ esiste, allora $E(X) \geq 0$. Se in aggiunta $E(X) = 0$ allora $P(X = 0) = 1$.*
6. *Se $a, b \in \mathbb{R}$ sono tali che $P(a \leq X \leq b) = 1$, allora $a \leq E(X) \leq b$.*
7. *Siano g e h funzioni tali che $E(g(X))$ ed $E(h(X))$ esistono. Se $P(h(X) \geq g(X)) = 1$, allora $E(h(X)) \geq E(g(X))$.*

Dimostrazione

1. Se $P(X = c) = 1$, allora $E(X) = c \cdot P(X = c) = c$.
2. Sia $Y := \mathbf{1}_B(X)$. Allora $Y \sim \mathbf{Be}(p)$ con $p = P(Y = 1) = P(X \in B)$ e quindi $E(Y) = P(X \in B)$.
3. Supponiamo ad esempio che X sia assolutamente continua. Allora per la Proposizione 3.1.12 vale che:

$$E(\alpha X) = \int_{\mathbb{R}} \alpha x f_X(x) dx = \alpha \int_{\mathbb{R}} x f_X(x) dx = \alpha E(X).$$

La dimostrazione nel caso discreto procede in modo analogo.

4. La dimostrazione di questo punto è del tutto analoga a quella del punto precedente e viene lasciata al lettore per esercizio.

5. Supponiamo X discreta. Poiché $P(X \geq 0) = 1$, allora $p_X(x) = 0$ per ogni $x < 0$ da cui:

$$E(X) = \sum_{k: x_k \geq 0} x_k p_X(x_k) \geq 0$$

Il caso continuo si tratta analogamente.

Tralasciamo la dimostrazione della seconda parte perché più delicata.

6. Innanzi tutto osserviamo che se $P(a \leq X \leq b) = 1$, allora dalla definizione di valore atteso segue che sicuramente esiste $E(X)$. Poi osserviamo che $P(X - a \geq 0) = 1$ e che per le proprietà 5. 4. e 1. si ha $0 \leq E(X - a) = E(X) + E(-a) = E(X) - a$, cioè $E(X) \geq a$. Per ottenere $E(X) \leq b$ basta osservare che $P(b - X \geq 0) = 1$ e ripetere il ragionamento.
7. La dimostrazione procede analogamente al punto 6. ed è lasciata al lettore per esercizio. ■

La proprietà 1. della Proposizione 3.1.14 ci dice che il valore atteso di una costante è la costante stessa; questa proprietà è talvolta chiamata proprietà di *coerenza* del valore atteso. La proprietà 2. sottolinea un ovvio legame tra valore atteso e probabilità. Le proprietà 3. e 4. ci dicono come si comporta il valore atteso quando si effettuano operazioni lineari sulla variabile aleatoria sottostante, queste proprietà sono dette proprietà di *linearità* del valore atteso. La proprietà 5. è detta *positività* del valore atteso. La proprietà 6. prende il nome di *internalità* del valore atteso.

Nota 3.1.15 Tutte le proprietà del valore atteso enunciate nella Proposizione 3.1.14 valgono sia nel caso discreto che nel caso assolutamente continuo, cioè in tutti i casi per cui in questo corso abbiamo definito il valore atteso. Questo ci autorizzerà nel seguito ad applicarle a tutte le variabili aleatorie che prenderemo in considerazione senza ulteriormente specificare di quale natura sia la loro funzione di ripartizione.

3.2 Varianza

Abbiamo visto nella sezione precedente che, in alcuni casi, è possibile associare a una variabile aleatoria una grandezza deterministica che abbiamo chiamato valore atteso. Tuttavia la media della variabile aleatoria non riassume tutte le proprietà qualitative di una variabile aleatoria, nel senso che ci sono variabili aleatorie che hanno la stessa media ma che sono qualitativamente molto differenti.

Esempio 3.2.1 (Segue Esempio 3.0.1) Supponiamo ancora di giocare a testa e croce con un amico. Sia ora Z il guadagno netto che otteniamo puntando 1000€: Z è una variabile discreta che assume solo i due valori -1000 e 1000 con probabilità $1/2$. Anche in questo caso il gioco è *equo*, cioè $E(Z) = 0$. Ma questo gioco è molto più rischioso rispetto a puntare 1 € come nell'Esempio 3.0.1. Eppure $E(Y) = E(Z) = 0$.

La differenza fondamentale fra Y e Z , rispettivamente degli Esempi 3.0.1 e 3.2.1, è che mentre Y assume valori vicini alla propria media, Z assume valori piuttosto lontani da $E(Z)$; $E(Y)$ rappresenta meglio Y di quanto non faccia $E(Z)$ per Z .

Il ragionamento appena fatto ci porta a considerare la distanza $|X - E(X)|$ tra una variabile aleatoria X e la sua media. Un oggetto matematicamente più facile da studiare è però la distanza al quadrato $[X - E(X)]^2$; questa è ancora una variabile aleatoria, che in alcuni casi ammette valore atteso.

Definizione 3.2.2 *Sia X una variabile aleatoria discreta o assolutamente continua, tale che esista $E(X)$. Se inoltre esiste $E((X - E(X))^2)$, allora si pone*

$$\text{Var}(X) := E((X - E(X))^2)$$

e $\text{Var}(X)$ si chiama varianza di X . La radice quadrata della varianza $\sqrt{\text{Var}(X)}$ prende il nome di deviazione standard di X .

Vale per la (3.1.1) che se X è una variabile aleatoria discreta con densità p_X e media $E(X) = \mu$ allora $\text{Var}(X) = \sum_k (x_k - \mu)^2 p_X(x_k)$; mentre, da (3.1.2) deduciamo che se X è una variabile aleatoria assolutamente continua con densità f_X e media $E(X) = \mu$ allora $\text{Var}(X) = \int (x - \mu)^2 f_X(x) dx$.

Esempio 3.2.3 (Seguono Esempi 3.0.1 e 3.2.1) Sia Y il guadagno netto che si ha giocando a testa e croce puntando 1€ e Z quello che si ha puntando 1000€. Allora $P(Y = -1) = P(Y = 1) = P(Z = 1000) = P(Z = -1000) = 1/2$ ed $E(Y) = E(Z) = 0$. Per quanto riguarda la varianza di Y si ha

$$\text{Var}(Y) = E((Y - E(Y))^2) = E(Y^2) = \sum_{k \in \{-1, 1\}} k^2 p_Y(k) = (-1)^2 \times \frac{1}{2} + 1^2 \times \frac{1}{2} = 1,$$

mentre per quella di Z si ha

$$\text{Var}(Z) = E((Z - E(Z))^2) = E(Z^2) = \sum_{k \in \{-10^3, 10^3\}} k^2 p_Z(k) = (-10^3)^2 \times \frac{1}{2} + (10^3)^2 \times \frac{1}{2} = 10^6.$$

Come già anticipato, $\text{Var}(Z)$ è (molto) più grande di $\text{Var}(Y)$ ad indicare che Z si discosta da $E(Z)$ (molto) più di quanto non faccia Y da $E(Y)$.

Esercizio 3.2.4 Calcolare la varianza della variabile aleatoria X dell'Esempio 3.0.1.

3.2.1 Proprietà della varianza

La seguente proposizione fornisce alcune proprietà elementari della varianza.

Proposizione 3.2.5 *Sia X una variabile aleatoria, allora*

1. $\text{Var}(X) = 0$ se e solo se $P(X = c) = 1$ per qualche costante c . In questo caso $c = E(X)$.
2. Se X ammette varianza ed $\alpha \in \mathbb{R}$ allora $\text{Var}(\alpha X) = \alpha^2 \text{Var}(X)$.
3. Se X ammette varianza e $\beta \in \mathbb{R}$ allora $\text{Var}(X + \beta) = \text{Var}(X)$.
4. Se X ammette varianza allora X^2 ammette media e $\text{Var}(X) = E(X^2) - E(X)^2$.

Dimostrazione Utilizzeremo le proprietà della media contenute nella Proposizione 3.1.14.

1. Se $P(X = c) = 1$ allora $E(X) = c$ e $\text{Var}(X) = E((c - c)^2) = E(0) = 0$. Viceversa, se $\text{Var}(X) = E((X - E(X))^2) = 0$, poiché $P(X - E(X))^2 \geq 0) = 1$ allora $P((X - E(X))^2 = 0) = 1$ che è possibile solo se $P(X = E(X)) = 1$.
2. Poiché $E(\alpha X) = \alpha E(X)$ allora

$$\begin{aligned} \text{Var}(\alpha X) &= E((\alpha X - E(\alpha X))^2) = E((\alpha X - \alpha E(X))^2) \\ &= E(\alpha^2 (X - E(X))^2) = \alpha^2 E((X - E(X))^2) = \alpha^2 \text{Var}(X). \end{aligned}$$

3. Osserviamo che per la linearità del valore atteso $E(X + \beta) = E(X) + \beta$, quindi

$$\begin{aligned} \text{Var}(X + \beta) &= E((X + \beta - E(X + \beta))^2) \\ &= E((X + \beta - E(X) - \beta)^2) \\ &= E((X - E(X))^2) \\ &= \text{Var}(X). \end{aligned}$$

4. Se X ammette varianza allora

$$\begin{aligned} E(X^2) &= E((X - E(X) + E(X))^2) \\ &\leq E(2(X - E(X))^2 + 2E(X)^2) \\ &= 2E((X - E(X))^2) + 2E(X)^2 \\ &= 2\text{Var}(X) + 2E(X)^2 < +\infty \end{aligned}$$

Quindi X^2 ammette media. Inoltre:

$$\begin{aligned} \text{Var}(X) &= E((X - E(X))^2) \\ &= E(X^2 - 2XE(X) + E(X)^2) \\ &= E(X^2) - 2E(XE(X)) + E(E(X)^2) \\ &= E(X^2) - 2E(X)^2 + E(X)^2 \\ &= E(X^2) - E(X)^2 \quad \blacksquare \end{aligned}$$

Commentiamo brevemente la proposizione appena dimostrata. Il punto 1. afferma che le uniche variabili aleatorie con varianza nulla sono le costanti. Questo è in pieno accordo con il concetto intuitivo di varianza come misura di quanto una variabile aleatoria si discosta dalla propria media. Il punto 2. ci dice che la varianza è quadratica (mentre la media è lineare). Il punto 3. mostra come la varianza sia invariante per traslazioni. Infatti sommando ad una variabile aleatoria un numero, cioè traslandola, anche la media viene traslata dello stesso numero e lo scostamento della variabile dalla sua media non cambia. Il punto 4. mostra una formula molto utile nelle applicazioni e negli esercizi per calcolare la varianza.

A titolo d'esempio calcoliamo la varianza di alcune delle variabili aleatorie precedentemente introdotte.

Esempio 3.2.6 Se X è variabile aleatoria uniforme discreta su $\{1, \dots, n\}$ sappiamo che $E(X) = (n+1)/2$ e

$$E(X^2) = \sum_{k=1}^n \frac{k^2}{n} = \frac{n(n+1)(2n+1)}{6n} = \frac{(n+1)(2n+1)}{6}$$

$$\text{Var}(X) = \frac{(n+1)(2n+1)}{6} - \left(\frac{n+1}{2}\right)^2 = \frac{n^2-1}{12}$$

Esempio 3.2.7 Se $X \sim \mathcal{P}(\lambda)$ sappiamo che $E(X) = \lambda$ e

$$E(X^2) = \sum_{k=0}^{+\infty} k^2 \frac{e^{-\lambda} \lambda^k}{k!} = \sum_{k=1}^{+\infty} k^2 \frac{e^{-\lambda} \lambda^k}{k!} = \sum_{k=1}^{+\infty} k \frac{e^{-\lambda} \lambda^k}{(k-1)!} = \sum_{h=0}^{+\infty} (h+1) \frac{e^{-\lambda} \lambda^{h+1}}{h!} =$$

$$= \lambda \sum_{h=0}^{+\infty} (h+1) \frac{e^{-\lambda} \lambda^h}{h!} = \lambda \sum_{h=0}^{+\infty} h \frac{e^{-\lambda} \lambda^h}{h!} + \lambda \sum_{h=0}^{+\infty} \frac{e^{-\lambda} \lambda^h}{h!} = \lambda E(X) + \lambda = \lambda^2 + \lambda$$

Segue che $\text{Var}(X) = E(X^2) - E(X)^2 = \lambda$.

Esercizio 3.2.8 Mostrare che, se X ha densità binomiale di parametri n e p , $E(X) = np$ e $\text{Var}(X) = np(1-p)$.

Esempio 3.2.9 Se $U \sim \mathcal{U}(0, 1)$ sappiamo che $E(U) = 1/2$; inoltre

$$E(U^2) = \int_0^1 u^2 du = \frac{1}{3}.$$

Segue che $\text{Var}(U) = E(U^2) - E(U)^2 = 1/3 - 1/4 = 1/12$.

Se $X \sim \mathcal{U}(a, b)$, allora $X = (b-a)U + a$, dove $U \sim \mathcal{U}(0, 1)$ (si veda Esempio 2.6.6) e

$$E(X) = (b-a)E(U) + a = \frac{b-a}{2} + a = \frac{a+b}{2}$$

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

Esempio 3.2.10 Se $X \sim \mathcal{E}(\lambda)$ sappiamo che $E(X) = 1/\lambda$ e

$$\begin{aligned} E(X^2) &= \int_0^{+\infty} x^2 \lambda e^{-\lambda x} dx \\ &= - \int_0^{+\infty} x^2 \frac{d}{dx} e^{-\lambda x} dx \\ &= - x^2 e^{-\lambda x} \Big|_0^{+\infty} + 2 \int_0^{+\infty} x e^{-\lambda x} dx \\ &= \frac{2}{\lambda} E(X) = \frac{2}{\lambda^2} \end{aligned}$$

Segue che $\text{Var}(X) = E(X^2) - E(X)^2 = 2/\lambda^2 - 1/\lambda^2 = 1/\lambda^2$.

Esempio 3.2.11 Se $Z \sim \mathcal{N}(0, 1)$ sappiamo che $E(Z) = 0$, quindi

$$\begin{aligned} \text{Var}(Z) = E(Z^2) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} z^2 e^{-\frac{z^2}{2}} dz = -\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} z \frac{d}{dz} e^{-\frac{z^2}{2}} dz = \\ &= -\frac{1}{\sqrt{2\pi}} z e^{-\frac{z^2}{2}} \Big|_{-\infty}^{+\infty} + \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{z^2}{2}} dz = 1 \end{aligned}$$

3.3 Disuguaglianza di Chebychev

La successiva importante disuguaglianza, nota come *disuguaglianza di Chebychev*, precisa in che senso una variabile X con varianza “piccola” è concentrata intorno alla sua media.

Proposizione 3.3.1 (Disuguaglianza di Chebychev) *Sia X una variabile aleatoria che ammette media e varianza. Allora per ogni $\epsilon > 0$:*

$$P(|X - E(X)| > \epsilon) \leq \frac{\text{Var}(X)}{\epsilon^2}$$

Dimostrazione Osserviamo che o $|X - E(X)| \leq \epsilon$ oppure $|X - E(X)| > \epsilon$; quindi $\mathbf{1}_{(-\infty, \epsilon]}(|X - E(X)|) + \mathbf{1}_{(\epsilon, +\infty)}(|X - E(X)|) \equiv 1$, da cui

$$\begin{aligned} \text{Var}(X) &= E((X - E(X))^2) \\ &= E((X - E(X))^2 \mathbf{1}_{(-\infty, \epsilon]}(|X - E(X)|)) + E((X - E(X))^2 \mathbf{1}_{(\epsilon, +\infty)}(|X - E(X)|)) \\ &\geq E((X - E(X))^2 \mathbf{1}_{(\epsilon, +\infty)}(|X - E(X)|)) \\ &\geq E(\epsilon^2 \mathbf{1}_{(\epsilon, +\infty)}(|X - E(X)|)) \\ &= \epsilon^2 P(|X - E(X)| > \epsilon) \quad \blacksquare \end{aligned}$$

Esercizio 3.3.2 Dimostrare che se X è una variabile aleatoria positiva tale che la k -esima potenza X^k ammette media per un intero positivo k , allora vale che

$$P(X > \epsilon) \leq \frac{E(X^k)}{\epsilon^k} \quad \forall \epsilon > 0.$$

Questa disuguaglianza è nota con il nome di *Disuguaglianza di Markov*.

3.4 Standardizzazione di una variabile aleatoria

In questa sezione ci occuperemo di una particolare trasformazione affine di una variabile aleatoria, detta standardizzazione.

Sia X una variabile aleatoria non costante che ammette media $E(X) = m$ e varianza $\text{Var}(X) = \sigma^2$. Poiché X non è costante $\text{Var}(X) > 0$. Consideriamo la variabile aleatoria Y ottenuta mediante la seguente trasformazione affine di X :

$$Y := \frac{X - m}{\sigma} \quad (3.4.1)$$

dove σ è la deviazione standard di X . Segue dalle Proposizioni 3.1.14 e 3.2.5 che Y ammette media e varianza finite. Inoltre segue dalla linearità della media che

$$E(Y) = E\left(\frac{X - m}{\sigma}\right) = \frac{E(X) - m}{\sigma} = 0$$

mentre, dalle proprietà della varianza otteniamo che

$$\text{Var}(Y) = \text{Var}\left(\frac{X - m}{\sigma}\right) = \frac{\text{Var}(X - m)}{\sigma^2} = \frac{\text{Var}(X)}{\sigma^2} = 1$$

Quindi, qualunque siano la media e la varianza di X ci siamo ricondotti a una variabile aleatoria Y con media uguale a 0 (diremo che è centrata) e varianza uguale a 1. Per questo motivo Y è detta *standardizzata della variabile X* e l'operazione che trasforma la variabile X nella corrispondente variabile Y è detta *standardizzazione*. Inoltre la funzione di ripartizione F_X della variabile X è legata alla funzione di ripartizione F_Y della sua standardizzata Y dalla semplice relazione

$$F_X(t) = P(X \leq t) = P\left(\frac{X - m}{\sigma} \leq \frac{t - m}{\sigma}\right) = F_Y\left(\frac{t - m}{\sigma}\right)$$

L'operazione di standardizzazione gioca un ruolo fondamentale nel teorema di De Moivre Laplace e nel Teorema centrale del limite che vedremo più avanti.

3.5 Densità gaussiana $\mathcal{N}(\mu, \sigma^2)$

Siano $Z \sim \mathcal{N}(0, 1)$, $\sigma > 0$ e $\mu \in \mathbb{R}$. Consideriamo la variabile aleatoria $X = \sigma Z + \mu$. Segue dalle proprietà di media e varianza che

$$E(X) = \sigma E(Z) + \mu = \mu$$

e

$$\text{Var}(X) = \sigma^2 \text{Var}(Z) = \sigma^2$$

Inoltre, grazie alla Proposizione 2.6.5, X è una variabile aleatoria assolutamente continua con densità

$$f_X(x) = \varphi\left(\frac{x-\mu}{\sigma}\right) \frac{1}{\sigma} = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\} \quad \forall x \in \mathbb{R} \quad (3.5.1)$$

dove φ rappresenta la densità gaussiana standard.

Definizione 3.5.1 Una variabile aleatoria assolutamente continua X con densità (3.5.1) è detta gaussiana di parametri μ e σ^2 e si indica $X \sim \mathcal{N}(\mu, \sigma^2)$.

Deduciamo dagli ultimi calcoli fatti che i due parametri di una variabile aleatoria $\mathcal{N}(\mu, \sigma^2)$ hanno una precisa interpretazione: μ è la media e σ^2 la varianza. Quindi, come messo in evidenza nella Figura 3.1, μ è un polo di riferimento e σ un indice della concentrazione (o dispersione) della densità $\mathcal{N}(\mu, \sigma^2)$ intorno a μ .

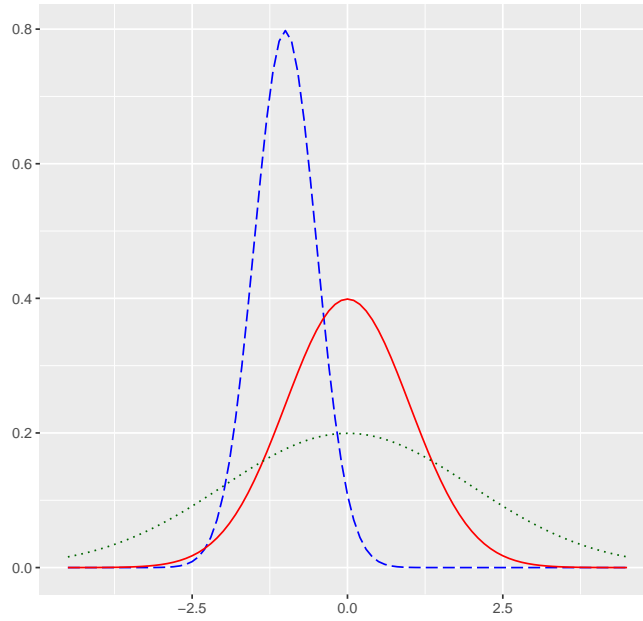


Figura 3.1: Grafico delle densità di probabilità $\mathcal{N}(0, 1)$ (linea rossa solida), $\mathcal{N}(0, 2)$ (linea verde puntinata) e $\mathcal{N}(-1, 0.5)$ (linea blu tratteggiata)

Esercizio 3.5.2 Sia $X \sim \mathcal{N}(\mu, \sigma^2)$. Mostrare che $Y = (X - \mu)/\sigma \sim \mathcal{N}(0, 1)$. Cioè la standardizzata di una variabile aleatoria gaussiana è una variabile aleatoria gaussiana standard.

Esercizio 3.5.3 Sia $X \sim \mathcal{N}(\mu, \sigma^2)$ e sia F_X la sua funzione di ripartizione. Mostrare che

$$F_X(x) = \Phi\left(\frac{x-\mu}{\sigma}\right) \quad \forall x \in \mathbb{R}$$

dove Φ è la funzione di ripartizione della densità gaussiana standard.

3.6 Approssimazione gaussiana della funzione di ripartizione binomiale

Un risultato² molto importante della teoria delle probabilità è il teorema di De Moivre-Laplace. Questo teorema afferma che, se standardizziamo una variabile aleatoria con densità binomiale di parametri n e p , la funzione di ripartizione della variabile così ottenuta converge, per $n \rightarrow +\infty$ e p fissato, alla funzione di ripartizione di una variabile aleatoria gaussiana standard. Vedremo nell'ultimo capitolo che questo risultato è un caso particolare del Teorema centrale del limite, ma la sua formulazione e dimostrazione è stata fornita in modo indipendente e molto tempo prima. Diamo qui di seguito l'enunciato del teorema di De Moivre-Laplace e ne illustriamo il suo utilizzo con un esempio. Ricordiamo che n prove di Bernoulli sono n esperimenti, con due possibili risultati, successo e insuccesso, i risultati di ciascuna prova sono eventi tra loro indipendenti e infine in ogni singola prova è uguale la probabilità che si verifichi il successo. (Si confronti la Sezione 1.5.3 del Capitolo 1).

Teorema 3.6.1 (di De Moivre-Laplace) *Sia S_n il numero di successi in n prove di Bernoulli, in ognuna delle quali il successo ha probabilità $p \in (0, 1)$. Allora, per ogni $a < b$,*

$$P\left(a < \frac{S_n - np}{\sqrt{np(1-p)}} \leq b\right) \rightarrow \Phi(b) - \Phi(a) \quad \text{per } n \rightarrow +\infty$$

dove Φ è la funzione di ripartizione di una gaussiana standard.

Nota 3.6.2 La variabile $(S_n - np)/\sqrt{np(1-p)}$ è la standardizzata di una variabile aleatoria binomiale. Infatti, come abbiamo già visto nel Capitolo 2, S_n ha densità binomiale di parametri n e p e, come vi si è chiesto di verificare nell'Esercizio 3.2.8, np e $\sqrt{np(1-p)}$ sono, rispettivamente, la sua media e la sua deviazione standard.

Nota 3.6.3 Si noti che abbiamo due possibili approssimazioni per le probabilità collegate ad una densità binomiale. Possiamo utilizzare un'approssimazione di Poisson se n è “grande” e p è “piccolo”, mentre si può vedere che vale un'approssimazione gaussiana se n è “grande” e p è “lontano” dai valori estremi 0 e 1. Esistono varie “ricette” per stabilire quanto n deve essere grande e p lontano da 0 e da 1. Per esempio, l'approssimazione gaussiana è buona se $np > 5$ e $n(1-p) > 5$, oppure per $np(1-p) \geq 10$. Invece, in [7] si indica che l'approssimazione di Poisson risulta soddisfacente se $n \geq 100$ e $p \leq 0.05$ oppure $n \geq 100$ e $np \leq 10$.

Esempio 3.6.4 Calcolare in modo approssimato la probabilità di ottenere in 100 lanci di una moneta equa un numero di teste compreso fra 45 e 55 (inclusi).

²Questa sezione è in parte tratta da [9]

Sia S_{100} la variabile aleatoria che conta il numero di teste nei 100 lanci. Allora $S_{100} \sim \mathbf{Bi}(100, 1/2)$ e la probabilità richiesta è

$$\begin{aligned}
 P(45 \leq S_{100} \leq 55) &= P(44 < S_{100} \leq 55) \\
 &= P\left(\frac{44 - \frac{100}{2}}{\sqrt{\frac{100}{4}}} < \frac{S_{100} - \frac{100}{2}}{\sqrt{\frac{100}{4}}} \leq \frac{55 - \frac{100}{2}}{\sqrt{\frac{100}{4}}}\right) \\
 &\simeq \Phi\left(\frac{55 - 50}{5}\right) - \Phi\left(\frac{44 - 50}{5}\right) \\
 &= \Phi(1) - \Phi(-1.2) \\
 &= \Phi(1) + \Phi(1.2) - 1 \simeq 0.84135 + 0.88493 - 1 = 0.72628
 \end{aligned}$$

D'altra parte, poiché S_{100} è variabile aleatoria discreta con funzione di ripartizione costante a tratti sull'intervallo $[k, k+1)$, per $k = 0, \dots, 99$, allora:

$$\begin{aligned}
 P(45 \leq S_{100} \leq 55) &= P(44 < S_{100} \leq 55) \\
 &= P(S_{100} \leq 55) - P(S_{100} \leq 44) \\
 &= P(S_{100} \leq 55.5) - P(S_{100} \leq 44.5) \\
 &\simeq \Phi\left(\frac{55.5 - 50}{5}\right) - \Phi\left(\frac{44.5 - 50}{5}\right) \\
 &= \Phi(1.1) - \Phi(-1.1) \\
 &= 2\Phi(1.1) - 1 \simeq 2 \times 0.86433 - 1 = 0.72867
 \end{aligned}$$

Nell'ultima equazione per calcolare un valore approssimato di $P(45 \leq S_{100} \leq 55) = P(S_{100} \leq 55) - P(S_{100} \leq 44)$, abbiamo apportato una *correzione di continuità* sostituendo a 55 il valore $55+0.5$ e a 44 il valore $44+0.5$. Calcolando ora esattamente $P(45 \leq S_{100} \leq 55)$ mediante la densità binomiale, otteniamo che $P(45 \leq S_{100} \leq 55) = 0.728747$. Quindi, senza la correzione di continuità l'approssimazione gaussiana produce un errore in percentuale pari a $(0.728747 - 0.726275)/0.728747 \simeq 0.34\%$, mentre, con la correzione di continuità, l'errore è $(0.728747 - 0.728668)/0.728747 \simeq 0.011\%$: l'introduzione della correzione di continuità ha ridotto l'errore di approssimazione di un fattore 31.

In generale, se n è grande e $S_n \sim \mathbf{Bi}(n, p)$, con $p \in (0, 1)$, la correzione di continuità si apporta nel seguente modo:

$$P(S_n \leq r) \simeq \Phi\left(\frac{r + 0.5 - np}{\sqrt{np(1-p)}}\right)$$

per ogni $r = 0, 1, \dots$.

Supponiamo ora di lanciare 100 monete con trucco $p = 1/5$ e sia S_{100} il numero di teste su 100 lanci. Allora $E(S_{100}) = 100/5 = 20$ e $\text{Var}(S_{100}) = 100 \times (1/5) \times (4/5) = 16$ e

$$\begin{aligned}
 P(16 \leq S_{100} \leq 24) &= P(15 < S_{100} \leq 24) = P\left(-1.25 \leq \frac{S_{100} - 20}{4} \leq 1\right) \\
 &\simeq \Phi(1) + \Phi(1.25) - 1 \simeq 0.841345 + 0.894350 - 1 = 0.735695.
 \end{aligned}$$

Con la correzione di continuità abbiamo:

$$\begin{aligned}
 P(16 \leq S_{100} \leq 24) &= P(S_{100} \leq 24.5) - P(S_{100} \leq 15.5) \\
 &= P\left(\frac{S_{100} - 20}{4} \leq 1.125\right) - P\left(\frac{S_{100} - 20}{4} \leq -1.125\right) \\
 &\simeq 2\Phi(1.125) - 1 \simeq 0.739411
 \end{aligned}$$

Si noti che $P(16 \leq S_{100} \leq 24)$ vale esattamente 0.7401413. In questo caso, l'errore in percentuale è pari a $(0.7401413 - 0.735695)/0.7401413 \simeq 0.6\%$ senza la correzione di continuità e $(0.7401413 - 0.739411)/0.7401413 \simeq 0.1\%$ con la correzione di continuità. Notiamo che con la correzione di continuità l'approssimazione è migliorata ma di misura inferiore rispetto al caso della moneta equa. D'altro canto l'errore relativo di approssimazione è comunque più alto rispetto al caso della moneta equa. Infine, trovate in Figura 3.2 il grafico della funzione di ripartizione gaussiana standard Φ e della standardizzata di una variabile aleatoria $S_{20} \sim \mathbf{Bi}(20, 0.5)$, a confronto.

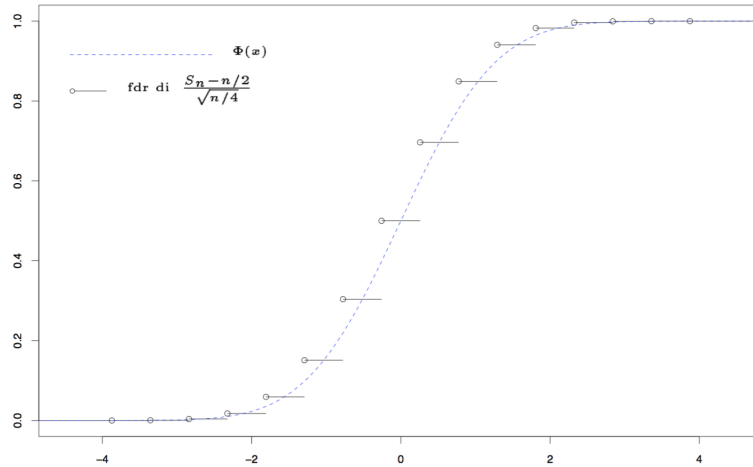


Figura 3.2: Funzione di ripartizione esatta di $(S_n - np(1 - p))/\sqrt{np(1 - p)}$ e funzione di ripartizione Φ a confronto per $n = 20$ e $p = 0.5$

3.7 *Momenti e funzione generatrice dei momenti

Nella Sezione 3.3 abbiamo visto come trarre informazioni sulla variabile aleatoria X conoscendo $E(X)$ ed $E(X^2)$, per esempio usando la disuguaglianza di Chebychev. Poi, nell'Esercizio 3.3.2 è stato preso in considerazione il numero $E(X^k)$. In generale si possono trarre maggiori informazioni conoscendo $E(X^k)$ per $k = 1, 2, \dots$. Non indagheremo questo punto a fondo, ma data l'importanza diamo la seguente definizione

Definizione 3.7.1 Sia X una variabile aleatoria assolutamente continua o discreta tale che $|X|^k$ ammetta valore atteso. Allora il numero $E(X^k)$ è detto momento k -esimo o momento di ordine k della variabile aleatoria X .

Secondo questa definizione la media di una variabile aleatoria è il suo momento primo, mentre la varianza è la differenza tra il suo momento secondo ed il suo momento primo al quadrato.

Proposizione 3.7.2 Sia X una variabile aleatoria che ammette momento k -esimo, per qualche $k \geq 2$. Allora X ammette momento h -esimo per ogni $1 \leq h < k$.

Dimostrazione Sappiamo per ipotesi che $E(|X|^k)$ è un numero finito, allora

$$\begin{aligned} E(|X|^k) &= E(|X|^k [\mathbf{1}_{[0,1]}(|X|) + \mathbf{1}_{(1,+\infty)}(|X|)]) \\ &= E(|X|^k \mathbf{1}_{[0,1]}(|X|)) + E(|X|^k \mathbf{1}_{(1,+\infty)}(|X|)) \\ &\geq E(|X|^k \mathbf{1}_{(1,+\infty)}(|X|)) = E(|X| |X|^{k-1} \mathbf{1}_{(1,+\infty)}(|X|)) \\ &\geq E(|X|^{k-1} \mathbf{1}_{(1,+\infty)}(|X|)) \end{aligned}$$

Quindi $|X|^{k-1} \mathbf{1}_{(1,+\infty)}(|X|)$ ha media finita. Poiché

$$|X|^{k-1} = |X|^{k-1} \mathbf{1}_{[0,1]}(|X|) + |X|^{k-1} \mathbf{1}_{(1,+\infty)}(|X|)$$

e

$$|X|^{k-1} \mathbf{1}_{[0,1]}(|X|) \leq 1$$

segue che anche $|X|^k$ ha media finita. Risulta così dimostrato che se esiste il momento k -esimo esiste anche il momento $(k-1)$ -esimo. Per concludere basta iterare il procedimento. ■

Esercizio 3.7.3 Mostrare che una variabile aleatoria gaussiana standard ammette momenti di ogni ordine; quindi verificare che quelli di ordine dispari sono nulli.

Più in generale, mostrare che se X è una variabile aleatoria *simmetrica* (cioè X e $-X$ hanno la stessa funzione di ripartizione) ed ammette momento di ordine n , allora tutti i momenti di ordine dispari $E(X^{2k+1})$ con $2k+1 \leq n$ sono nulli.

Esercizio 3.7.4 Sia X una variabile aleatoria assolutamente continua con densità

$$f_X(x) = \begin{cases} 0 & x < 1 \\ 3x^{-4} & x \geq 1 \end{cases}$$

Mostrare che X ammette momenti primo e secondo ma non ammette momento terzo.

Uno strumento molto utile nel calcolo dei momenti di una variabile aleatoria, quindi anche nel calcolo di media e varianza, è la *funzione generatrice dei momenti*:

Definizione 3.7.5 Sia X una variabile aleatoria per la quale esiste un intervallo aperto \mathcal{O} contenente lo 0 tale che e^{tX} ammette media per ogni t in \mathcal{O} . Allora la funzione

$$m_X(t) := E(e^{tX})$$

definita (almeno) per ogni $t \in \mathcal{O}$ è detta funzione generatrice dei momenti di X .

Nota 3.7.6 La funzione generatrice dei momenti di una variabile aleatoria X coincide con la trasformata di Laplace della densità di probabilità di X .

Esercizio 3.7.7 Sia $X \sim \mathcal{N}(\mu, \sigma^2)$. Mostrare che la funzione generatrice dei momenti di X è

$$m_X(t) = e^{\mu t + \sigma^2 t^2 / 2}$$

per ogni $t \in \mathbb{R}$.

Il nome di funzione generatrice dei momenti è dovuto alla seguente proprietà di m_X .

Proposizione 3.7.8 Sia X una variabile aleatoria che ammette funzione generatrice dei momenti m_X . Allora esistono tutti i momenti di X e

$$E(X) = m'_X(0), \quad E(X^2) = m''_X(0), \dots$$

Non dimostreremo questa proposizione, ma per ricordarla meglio conviene tenere presente la seguente *dimostrazione formale*:

$$m'_X(t) = \frac{d}{dt} E(e^{tX}) = E\left(\frac{d}{dt} e^{tX}\right) = E(X e^{tX})$$

quindi $m'_X(0) = E(X)$. Lo stesso ragionamento si può ripetere per i momenti successivi.

Esercizio 3.7.9 Calcolare il momento quarto di una variabile aleatoria X avente densità gaussiana standard.

Capitolo 4

Vettori Aleatori

4.1 Variabili aleatorie indipendenti

Nel Capitolo 2 abbiamo introdotto la variabile aleatoria per modellare il concetto di numero casuale. Spesso nelle applicazioni accade che sia necessario considerare simultaneamente più variabili aleatorie definite sullo stesso spazio campione, cioè relative ad uno stesso esperimento aleatorio. Per esempio possiamo pensare di essere interessati alla misurazione di altezza e peso degli individui di una certa popolazione, oppure siamo interessati ai tempi di vita dei componenti che costituiscono un'apparecchiatura complessa. È importante quindi conoscere il comportamento congiunto di più variabili aleatorie. Cosa significa fare ciò dal punto di vista probabilistico?

Esempio 4.1.1 Consideriamo l'esperimento aleatorio consistente nel lanciare dieci volte due monete equilibrate. Sia X il numero di teste nei dieci lanci della prima moneta e Y quello nella seconda. L'evento $\{X = 5\}$ è costituito da tutte le possibili sequenze dei dieci lanci delle due monete compatibili con il fatto che la prima moneta abbia mostrato esattamente cinque volte testa. Analogamente, $\{Y \leq 8\}$ è l'evento che si verifica se la seconda moneta ha mostrato testa al più otto volte. Considerare contemporaneamente il verificarsi di questi due eventi (cioè l'intersezione) riguarda il comportamento congiunto delle due variabili aleatorie e scriveremo:

$$\{X = 5, Y \leq 8\} = \{\omega : X(\omega) = 5\} \cap \{\omega : Y(\omega) \leq 8\}.$$

D'ora in avanti si userà questa notazione per indicare intersezione di eventi espressi in termini di variabili aleatorie. In questo caso è chiaro che gli eventi $\{X = 5\}$ e $\{Y \leq 8\}$ sono indipendenti secondo la Definizione 1.5.18 e quindi

$$P(X = 5, Y \leq 8) = P(X = 5)P(Y \leq 8).$$

Il conto procede considerando che $X \sim \mathbf{Bi}(10, 1/2)$ e $Y \sim \mathbf{Bi}(10, 1/2)$. Allo stesso modo si prova che

$$P(X \in A, Y \in B) = P(X \in A)P(Y \in B)$$

per ogni $A, B \subset \{0, 1, \dots, 10\}$ e le due probabilità a secondo membro si calcolano facilmente ricorrendo alla densità binomiale.

Sia ora Z la v.a. che indica il numero totale di teste nei lanci di entrambe le monete. Allora $Z \sim \mathbf{Bi}(20, 1/2)$ ma, evidentemente, gli eventi $\{Z \leq 8\}$ e $\{X = 5\}$ non sono indipendenti: ad esempio sull'insieme $\{\omega : X(\omega) = 5\}$ Z non può assumere valori inferiori a 5. Questo fatto implica che per calcolare $P(X = 5, Z \leq 8)$ non è sufficiente conoscere le densità di X e di Z , ma è necessario analizzare più a fondo il loro comportamento congiunto. Comunque, in questo caso il calcolo è facile:

$$\begin{aligned} P(X = 5, Z \leq 8) &= P(\{\omega : X(\omega) = 5\} \cap \{Z(\omega) \leq 8\}) \\ &= P(\{\omega : X(\omega) = 5\} \cap \{Y(\omega) \leq 3\}) \\ &= P(X = 5, Y \leq 3) = p_X(5)F_Y(3) \simeq 0.0423 \end{aligned}$$

D'altro canto si noti che

$$P(X = 5)P(Z \leq 8) = p_X(5)F_Z(8) \simeq 0.0619 \neq 0.0423 = P(X = 5, Z \leq 8)$$

Appare naturale chiamare le variabili aleatorie X e Y *indipendenti* in quanto generano eventi indipendenti. Se X, Y sono indipendenti, tutte le probabilità che riguardano la coppia si deducono dalle densità delle singole variabili, cioè se p_X e p_Y rappresentano le densità di X e Y , rispettivamente, allora

$$P(X = x, Y = y) = p_X(x)p_Y(y) = \begin{cases} \binom{10}{x}\binom{10}{y}\frac{1}{2^{20}} & x, y = 0, 1, \dots, 10 \\ 0 & \text{altrove} \end{cases}$$

Diversamente, per la coppia (X, Z) , la sola conoscenza di p_X e p_Z non porta direttamente a quella di $P(X = x, Z = z)$.

La seguente definizione di indipendenza fra variabili aleatorie formalizza ed estende i concetti introdotti con l'Esempio 4.1.1. Come nel caso bidimensionale, anche nel caso di un numero qualunque n di variabili aleatorie X_1, \dots, X_n definite sullo stesso spazio di probabilità (Ω, \mathcal{F}, P) , useremo la scrittura $\{X_1 \in B_1, \dots, X_n \in B_n\}$ per indicare l'intersezione degli eventi $\{\omega \in \Omega : X_1(\omega) \in B_1\}, \dots, \{\omega \in \Omega : X_n(\omega) \in B_n\}$:

$$\{X_1 \in B_1, \dots, X_n \in B_n\} = \{\omega \in \Omega : X_1(\omega) \in B_1\} \cap \dots \cap \{\omega \in \Omega : X_n(\omega) \in B_n\}.$$

Definizione 4.1.2 *Siano X_1, \dots, X_n n variabili aleatorie definite sullo stesso spazio di probabilità (Ω, \mathcal{F}, P) . Diciamo che sono indipendenti se*

$$P(X_1 \in B_1, \dots, X_n \in B_n) = P(X_1 \in B_1) \cdots P(X_n \in B_n) \quad (4.1.1)$$

per ogni scelta di domini regolari¹ $B_1, \dots, B_n \subset \mathbb{R}$.

¹Gli insiemi B_1, \dots, B_n che dobbiamo considerare sono solo quelli ottenuti con un numero al più infinito numerabile di operazioni fra intervalli.

Si può verificare che le variabili aleatorie X, Y dell'Esempio 4.1.1 sono indipendenti secondo la Definizione 4.1.2.

Nota 4.1.3 Nell'equazione (4.1.1) prendiamo $B_i = (-\infty, x_i]$ per ogni $i = 1, \dots, n$ con $x_i \in \mathbb{R}$. Allora (4.1.1) diventa

$$P(X_1 \leq x_1, \dots, X_n \leq x_n) = P(X_1 \leq x_1) \cdots P(X_n \leq x_n) \quad (4.1.2)$$

In altri termini, se le variabili aleatorie sono indipendenti, allora vale (4.1.2). In realtà, vale anche il viceversa, cioè

Proposizione 4.1.4 *Le variabili aleatorie X_1, \dots, X_n sono indipendenti se e solo se per ogni scelta di $x_1, \dots, x_n \in \mathbb{R}$ vale (4.1.2).*

Nota 4.1.5 Siano X_1, \dots, X_n variabili aleatorie discrete indipendenti con densità rispettivamente p_{X_1}, \dots, p_{X_n} . Allora, prendendo $B_1 = \{x_1\}, \dots, B_n = \{x_n\}$ in (4.1.1) risulta che

$$P(X_1 = x_1, \dots, X_n = x_n) = P(X_1 = x_1) \cdots P(X_n = x_n), \quad \forall x_i \in \mathbb{R}, \forall i = 1, \dots, n. \quad (4.1.3)$$

Anche in questo caso vale il viceversa:

Proposizione 4.1.6 *Le variabili aleatorie discrete X_1, \dots, X_n sono indipendenti se e solo se vale (4.1.3).*

Rivisitiamo alla luce di quanto ora introdotto l'esperimento di n prove di Bernoulli.

Esempio 4.1.7 Sia (Ω, \mathcal{F}, P) lo spazio di n prove di Bernoulli di parametro p e siano X_1, \dots, X_n le variabili aleatorie definite su questo spazio da:

$$X_i(\omega) = a_i, \quad \forall \omega := (a_1, \dots, a_n) \in \Omega = \{0, 1\}^n, \quad \forall i = 1, \dots, n$$

È immediato verificare che X_1, \dots, X_n sono indipendenti. Infatti, $X_i \sim \mathbf{Be}(p)$ e

$$\begin{aligned} P(X_1 = a_1, \dots, X_n = a_n) &= p^{\sum_{i=1}^n a_i} (1-p)^{n-\sum_{i=1}^n a_i} \\ &= \prod_{i=1}^n p^{a_i} (1-p)^{1-a_i} = P(X_1 = a_1) \cdots P(X_n = a_n) \end{aligned}$$

Esempio 4.1.8 Siano S e T due variabili aleatorie che descrivono i tempi di guasto, in minuti secondi, di due componenti elettronici. Supponiamo che il modello probabilistico assegnato sia il seguente: la probabilità che il primo componente funzioni nei primi s secondi e il secondo nei primi t secondi (per ogni $s \geq 0$ e $t \geq 0$) sia

$$P(S > s, T > t) = \int_s^{+\infty} \int_t^{+\infty} \lambda^2 e^{-\lambda(u+v)} du dv$$

Segue che per ogni $s > 0$:

$$P(S > s) = P(S > s, T > 0) = \int_s^{+\infty} \lambda e^{-\lambda u} du \int_0^{+\infty} \lambda e^{-\lambda v} dv = \int_s^{+\infty} \lambda e^{-\lambda u} du = e^{-\lambda s}.$$

Analogamente, $P(T > t) = e^{-\lambda t}$ e quindi:

$$P(S > s, T > t) = P(S > s)P(T > t)$$

cioè, gli eventi $\{S > s\}$ e $\{T > t\}$ sono indipendenti. Ma allora anche gli eventi complementari $\{S \leq s\}$ e $\{T \leq t\}$ sono indipendenti, da cui:

$$\begin{aligned} P(S \leq s, T \leq t) &= [1 - P(S > s)][1 - P(T > t)] = \\ &= F_S(s)F_T(t) = \int_0^s \lambda e^{-\lambda u} du \int_0^t \lambda e^{-\lambda v} dv \quad (4.1.4) \end{aligned}$$

Deduciamo da (4.1.4) che S e T sono indipendenti.

4.2 Vettori aleatori

È opportuno a questo punto introdurre alcune definizioni in cui ritroviamo gli oggetti considerati nella sezione precedente.

Definizione 4.2.1 (Vettore aleatorio) Sia (Ω, \mathcal{F}, P) uno spazio di probabilità. Un vettore aleatorio n -dimensionale è una funzione vettoriale $\mathbf{X} := (X_1, \dots, X_n)$, $\mathbf{X}: \Omega \rightarrow \mathbb{R}^n$ tale che ogni X_i (per $i = 1, \dots, n$) è una variabile aleatoria.

Esempio 4.2.2 (Continuazione degli Esempi 4.1.1 e 4.1.8) Le coppie (X, Y) , (X, Z) considerate nell'Esempio 4.1.1 e (S, T) nell'Esempio 4.1.8 sono vettori aleatori bidimensionali.

Discutendo della nozione di indipendenza, nell'equazione (4.1.2), abbiamo considerato probabilità della forma:

$$P(X_1 \leq x_1, \dots, X_n \leq x_n) \quad \forall x_i \in \mathbb{R}, \forall i = 1, \dots, n \quad (4.2.1)$$

La precedente, per $n = 1$, definisce la funzione di ripartizione di una variabile aleatoria. La considerazione di (4.2.1) per n qualsiasi, al variare di x_i in \mathbb{R} per $i = 1, \dots, n$, porta a introdurre la funzione di ripartizione di vettori aleatori:

Definizione 4.2.3 (Funzione di ripartizione multidimensionale) Sia $\mathbf{X} = (X_1, \dots, X_n)$ un vettore aleatorio n -dimensionale definito su uno spazio di probabilità (Ω, \mathcal{F}, P) . Si chiama funzione di ripartizione di \mathbf{X} (o funzione di ripartizione congiunta di X_1, \dots, X_n) la funzione $F_{\mathbf{X}} = F_{(X_1, \dots, X_n)}: \mathbb{R}^n \rightarrow [0, 1]$ definita per ogni $(x_1, \dots, x_n) \in \mathbb{R}^n$ come $F_{(X_1, \dots, X_n)}(x_1, \dots, x_n) := P(X_1 \leq x_1, \dots, X_n \leq x_n)$.

La funzione di ripartizione di un vettore aleatorio gode di alcune proprietà analoghe a quelle della funzione di ripartizione di una variabile aleatoria (cfr. Proposizione 2.1.10).

Dato un vettore aleatorio $\mathbf{X} = (X_1, \dots, X_n)$ è interessante vedere che legame c'è tra $F_{\mathbf{X}}$ e le funzioni di ripartizione F_{X_1}, \dots, F_{X_n} delle componenti che spesso vengono chiamate *funzioni di ripartizione marginali*.

Proposizione 4.2.4 *Sia $\mathbf{X} = (X_1, \dots, X_n)$ un vettore aleatorio che ha funzione di ripartizione $F_{\mathbf{X}}$ e sia $\mathbf{x} = (x_1, \dots, x_n)$. Allora per ogni $k = 1, \dots, n$, $\lim_{x_k \rightarrow -\infty} F_{\mathbf{X}}(\mathbf{x}) = 0$, mentre*

$$\begin{aligned} \lim_{x_k \rightarrow +\infty} F_{\mathbf{X}}(\mathbf{x}) &= P(X_1 \leq x_1, \dots, X_{k-1} \leq x_{k-1}, X_{k+1} \leq x_{k+1}, \dots, X_n \leq x_n) \\ &= F_{(X_1, \dots, X_{k-1}, X_{k+1}, \dots, X_n)}(x_1, \dots, x_{k-1}, x_{k+1}, \dots, x_n) \end{aligned}$$

Esercizio 4.2.5 Dimostrare la Proposizione 4.2.4.

La precedente proposizione ci dice che se $x_k \rightarrow +\infty$, $F_{\mathbf{X}}(\mathbf{x})$ converge alla funzione di ripartizione del vettore aleatorio $(n-1)$ -dimensionale $(X_1, \dots, X_{k-1}, X_{k+1}, \dots, X_n)$ che si ottiene da \mathbf{X} eliminando la k -esima componente.

Nel caso di un vettore aleatorio bidimensionale (X, Y) che ha funzione di ripartizione $F_{X,Y}$, la Proposizione 4.2.4 afferma che:

$$\begin{aligned} \lim_{x \rightarrow +\infty} F_{X,Y}(x, y) &= P(Y \leq y) = F_Y(y) \\ \lim_{y \rightarrow +\infty} F_{X,Y}(x, y) &= P(X \leq x) = F_X(x) \end{aligned}$$

Nel caso di un vettore aleatorio n -dimensionale, applicando iterativamente la Proposizione 4.2.4, si ottiene che per ogni $x \in \mathbb{R}$ e per ogni $i = 1, \dots, n$

$$F_{X_i}(x) = \lim_{\substack{x_j \rightarrow +\infty \\ \forall j \neq i}} F_{\mathbf{X}}(x_1, \dots, x_{i-1}, x, x_{i+1}, \dots, x_n)$$

Quindi dalla funzione di ripartizione congiunta siamo in grado di calcolare tutte le funzioni di ripartizione marginali. Ma il viceversa è falso, come mostra il seguente esempio.

Esempio 4.2.6 Siano (X_1, Y_1) un vettore aleatorio con funzione di ripartizione

$$F_{X_1, Y_1}(x, y) = \begin{cases} 0 & x < 0 \text{ o } y < 0 \\ e^{-y}(1 - e^{-x} - x) - (1 + x)e^{-x} + 1 & 0 \leq x \leq y \\ e^{-x}(1 - e^{-y} - y) - (1 + y)e^{-y} + 1 & x > y \geq 0 \end{cases}$$

e (X_2, Y_2) un vettore aleatorio con funzione di ripartizione

$$F_{X_2, Y_2}(x, y) = \{1 + (1 + x)e^{-x}(e^{-y}(1 + y) - 1) - e^{-y}(1 + y)\} \mathbf{1}_{(0, \infty)}(x) \mathbf{1}_{(0, \infty)}(y)$$

Verificate che le funzioni di ripartizione marginali di F_{X_1, Y_1} e F_{X_2, Y_2} coincidono e sono date da

$$F_{X_i}(x) = F_{Y_i}(x) = (1 - (1 + x)e^{-x}) \mathbf{1}_{(0, \infty)}(x), \quad i = 1, 2$$

Riusciamo a ricostruire la funzione di ripartizione congiunta dalle marginali nel caso di variabili aleatorie indipendenti.

Alla luce della definizione appena introdotta, possiamo rinunciare la Proposizione 4.1.4 nel seguente modo

Proposizione 4.2.7 *Le componenti di un vettore aleatorio $\mathbf{X} = (X_1, \dots, X_n)$ sono indipendenti se e solo se la funzione di ripartizione di \mathbf{X} coincide con il prodotto delle funzioni di ripartizione marginali, cioè*

$$F_{\mathbf{X}} = F_{X_1} F_{X_2} \cdots F_{X_n}$$

Esempio 4.2.8 (Continuazione dell'Esempio 4.1.8) Le variabili aleatorie S e T che rappresentano i tempi di guasto dei componenti elettronici dell'Esempio 4.1.8 hanno funzione di ripartizione congiunta

$$F_{S,T}(s, t) = (1 - e^{-\lambda s})(1 - e^{-\lambda t})\mathbf{1}_{(0,+\infty) \times (0,+\infty)}(s, t) = F_S(s)F_T(t)$$

e S e T sono indipendenti.

Per esempi significativi di vettori aleatori con componenti non indipendenti rimandiamo alle prossime sezioni. Seguendo lo schema del caso unidimensionale, di seguito considereremo le due classi di vettori aleatori *discreti* e *assolutamente continui*.

4.3 Vettori aleatori discreti

Definizione 4.3.1 (Vettori aleatori discreti) *Un vettore aleatorio \mathbf{X} n -dimensionale è discreto se le sue componenti X_1, \dots, X_n sono variabili aleatorie discrete.*

Esempi di vettori aleatori discreti sono i vettori (X, Y) e (X, Z) dell'Esempio 4.1.1.

Per un vettore aleatorio discreto è possibile definire una *densità discreta* nel modo seguente:

Definizione 4.3.2 *Sia \mathbf{X} un vettore aleatorio discreto su di uno spazio di probabilità (Ω, \mathcal{F}, P) . La funzione $p_{\mathbf{X}}(\mathbf{x}) := P(X_1 = x_1, \dots, X_n = x_n)$, dove $\mathbf{x} = (x_1, \dots, x_n)$, si chiama densità discreta del vettore aleatorio \mathbf{X} (o densità congiunta di X_1, \dots, X_n).*

Si noti che se $p_{\mathbf{X}}$ è la densità di un vettore aleatorio discreto \mathbf{X} allora $p_{\mathbf{X}}(\mathbf{x}) = 0$ tranne che per una quantità al più numerabile di $\mathbf{x} \in \mathbb{R}^n$.

Esempio 4.3.3 (Densità multinomiale) Supponiamo che una popolazione contenga oggetti di $k \geq 2$ tipi diversi e che la proporzione degli oggetti di tipo i nella popolazione sia p_i per $i = 1, \dots, k$ ($p_i > 0, \sum_{i=1}^k p_i = 1$). Inoltre, supponiamo che n oggetti siano scelti a caso dalla popolazione con reimmissione. Sia X_i il numero di oggetti di tipo i estratti, per $i = 1, \dots, k$ e sia \mathbf{X} il vettore aleatorio che ha componenti X_1, \dots, X_k . Allora il vettore aleatorio \mathbf{X} è discreto, la somma delle sue componenti è pari al numero di estrazioni ($X_1 + \dots + X_k = n$) e la sua densità è detta *multinomiale* di parametri n, p_1, \dots, p_k .

Per scrivere esplicitamente la densità multinomiale, possiamo pensare di estrarre gli elementi dalla popolazione uno alla volta. Poiché le n scelte sono indipendenti, la probabilità che la sequenza delle n estrazioni contenga n_1 elementi di tipo 1, ..., n_k elementi di tipo k (in un ordine prefissato) è $p_1^{n_1} \cdots p_k^{n_k}$. Inoltre, il numero di modi differenti in cui l'ordine degli n oggetti può essere specificato è pari al numero di partizioni ordinate di classe (n_1, \dots, n_k) , cioè

$$\binom{n}{n_1 \dots n_k} := \frac{n!}{n_1! \times n_2! \times \cdots \times n_k!}$$

Segue che la probabilità di ottenere esattamente n_1 elementi di tipo 1, ..., n_k elementi di tipo k è

$$P(X_1 = n_1, \dots, X_k = n_k) = \binom{n}{n_1 \dots n_k} p_1^{n_1} \cdots p_k^{n_k} \quad \forall n_1, \dots, n_k = 0, \dots, n: n_1 + \cdots + n_k = n$$

Si osservi che per $k = 2$, il vettore \mathbf{X} si riduce al vettore $(X_1, n - X_1)$ e $X_1 \sim \mathbf{Bi}(n, p_1)$.

Se $p_{\mathbf{X}}$ è la densità di \mathbf{X} allora valgono proprietà analoghe a quelle della densità discreta unidimensionale (cfr. Proposizione 2.2.4). Per definire queste proprietà penseremo \mathbb{R}^n dotato delle consuete operazioni di somma e prodotto per uno scalare e della seguente relazione di ordine parziale " \leq ": se $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, allora $\mathbf{x} \leq \mathbf{y}$ se e solo se $x_k \leq y_k$ per ogni $k = 1, \dots, n$.

Proposizione 4.3.4 *Sia $p_{\mathbf{X}}$ la densità di un vettore aleatorio n -dimensionale \mathbf{X} che assume valori in un insieme al più numerabile S con probabilità 1 (cioè $P(\mathbf{X} \in S) = 1$). Allora*

1. $0 \leq p_{\mathbf{X}}(\mathbf{x}) \leq 1$ per ogni $\mathbf{x} \in \mathbb{R}^n$ e $p_{\mathbf{X}}(\mathbf{x}) = 0$ per ogni $\mathbf{x} \notin S$;
2. $\sum_{\mathbf{x} \in S} p_{\mathbf{X}}(\mathbf{x}) = 1$;
3. se $F_{\mathbf{X}}$ è la funzione di ripartizione di \mathbf{X} allora

$$F_{\mathbf{X}}(\mathbf{x}) = \sum_{\mathbf{y} \in S: \mathbf{y} \leq \mathbf{x}} p_{\mathbf{X}}(\mathbf{y}) \quad \forall \mathbf{x} \in \mathbb{R}^n$$

4. se $B \subset \mathbb{R}^n$ allora

$$P(\mathbf{X} \in B) = \sum_{\mathbf{x} \in B \cap S} p_{\mathbf{X}}(\mathbf{x})$$

Dimostrazione La dimostrazione è analoga a quella della Proposizione 2.2.4 e viene lasciata per esercizio al lettore. ■

Un'applicazione particolarmente importante del punto 4. della Proposizione 4.3.4 riguarda il calcolo delle densità delle componenti X_i del vettore aleatorio discreto \mathbf{X} dette *densità*

marginali. Supponiamo che $\mathbf{X} = (X_1, \dots, X_n)$ sia un vettore aleatorio n -dimensionale a valori in S con densità $p_{\mathbf{X}}$. Vogliamo calcolare la densità di X_1 . A tal fine osserviamo che

$$p_{X_1}(x_1) = P(X_1 = x_1) = P(X_1 = x_1, X_2 \in \mathbb{R}, \dots, X_n \in \mathbb{R}) = P(\mathbf{X} \in B)$$

dove $B := \{x_1\} \times \mathbb{R}^{n-1}$; quindi

$$p_{X_1}(x_1) = \sum_{\mathbf{x} \in B \cap S} p_{\mathbf{X}}(\mathbf{x}) = \sum_{x_2, \dots, x_n} p_{\mathbf{X}}(x_1, x_2, \dots, x_n)$$

Esercizio 4.3.5 Fornire l'espressione della densità marginale della generica componente X_i del vettore \mathbf{X} .

Esercizio 4.3.6 Fornire l'espressione della densità congiunta delle prime due componenti X_1 e X_2 del vettore \mathbf{X} .

Alla luce delle definizioni ora introdotte, rinunciando la Proposizione 4.1.6 nei seguenti termini:

Proposizione 4.3.7 *Le componenti di un vettore aleatorio discreto $\mathbf{X} = (X_1, \dots, X_n)$ sono indipendenti se e solo se la densità di \mathbf{X} coincide con il prodotto delle densità marginali p_{X_1}, \dots, p_{X_n} di X_1, \dots, X_n , rispettivamente, cioè*

$$p_{\mathbf{X}} = p_{X_1} \cdots p_{X_n} \quad (4.3.1)$$

Esercizio 4.3.8 Dimostrare la Proposizione 4.3.7.

Esempio 4.3.9 (Continuazione dell'Esempio 4.1.1) Il vettore aleatorio (X, Y) dell'Esempio 4.1.1 soddisfa l'equazione (4.3.1).

4.4 Vettori aleatori assolutamente continui

In questa sezione introduciamo l'analogo multidimensionale del concetto di variabile aleatoria assolutamente continua.

Definizione 4.4.1 (Vettori aleatori assolutamente continui) *Un vettore aleatorio \mathbf{X} n -dimensionale è assolutamente continuo se esiste una funzione $f_{\mathbf{X}}: \mathbb{R}^n \rightarrow \mathbb{R}^+$ integrabile, tale che la funzione di ripartizione $F_{\mathbf{X}}$ di \mathbf{X} si può scrivere come*

$$F_{\mathbf{X}}(\mathbf{x}) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} f_{\mathbf{X}}(s_1, \dots, s_n) ds_n \cdots ds_1 \quad \forall \mathbf{x} = (x_1, \dots, x_n)$$

$f_{\mathbf{X}}$ prende il nome di densità del vettore aleatorio assolutamente continuo \mathbf{X} (o densità congiunta di X_1, \dots, X_n).

Per i vettori aleatori assolutamente continui e le loro densità valgono proprietà analoghe a quelle delle variabili aleatorie assolutamente continue date nella Proposizione 2.4.4:

Proposizione 4.4.2 Sia $f_{\mathbf{X}}$ la densità di un vettore aleatorio n -dimensionale assolutamente continuo \mathbf{X} . Allora

1. $\int_{\mathbb{R}^n} f_{\mathbf{X}}(x_1, \dots, x_n) dx_1 \cdots dx_n = 1$

2. Se $F_{\mathbf{X}}$ è la funzione di ripartizione di \mathbf{X} allora

$$\frac{\partial^n F_{\mathbf{X}}(\mathbf{x})}{\partial x_1 \cdots \partial x_n} = f_{\mathbf{X}}(\mathbf{x})$$

per tutti gli $\mathbf{x} \in \mathbb{R}^n$ tali che esiste la derivata parziale al primo membro;

3. se $B \subset \mathbb{R}^n$ è un “dominio regolare” allora

$$P(\mathbf{X} \in B) = \int_B f_{\mathbf{X}}(x_1, \dots, x_n) dx_1 \cdots dx_n$$

Dimostrazione La dimostrazione è analoga a quella della Proposizione 2.4.4 e viene lasciata come esercizio. ■

Proposizione 4.4.3 Se $f_{\mathbf{X}}$ è la densità di un vettore aleatorio n -dimensionale assolutamente continuo $\mathbf{X} = (X_1, \dots, X_n)$ allora X_i è una variabile aleatoria assolutamente continua e la sua densità è

$$f_{X_i}(x_i) = \int_{\mathbb{R}^{n-1}} f_{\mathbf{X}}(s_1, \dots, s_{i-1}, x_i, s_{i+1}, \dots, s_n) ds_1 \cdots ds_{i-1} ds_{i+1} \cdots ds_n$$

Dimostrazione Per semplicità di notazioni, consideriamo il caso $i = 1$. Bisogna dimostrare che

$$F_{X_1}(x) = \int_{-\infty}^x \left\{ \int_{\mathbb{R}^{n-1}} f_{\mathbf{X}}(s_1, \dots, s_n) ds_2 \cdots ds_n \right\} ds_1$$

che è vero in quanto, se $B := (-\infty, x] \times \mathbb{R}^{n-1}$, allora per il punto 3. della Proposizione 4.4.2 abbiamo:

$$\begin{aligned} \int_{-\infty}^x \left\{ \int_{\mathbb{R}^{n-1}} f_{\mathbf{X}}(s_1, \dots, s_n) ds_2 \cdots ds_n \right\} ds_1 &= P(\mathbf{X} \in B) = \\ &= P(X_1 \leq x, X_2 \in \mathbb{R}, \dots, X_n \in \mathbb{R}) = P(X_1 \leq x) = F_{X_1}(x) \quad \blacksquare \end{aligned}$$

Le densità delle componenti di un vettore assolutamente continuo sono dette *densità marginali*.

Esempio 4.4.4 (Densità uniforme sul cerchio) [Tratto da [1]] Siano (X, Y) le coordinate di un punto “scelto a caso” nel cerchio C di raggio r : $C = \{(x, y) \in \mathbb{R}^2: x^2 + y^2 \leq$

$r^2\}$. Questo significa che il vettore aleatorio (X, Y) è assolutamente continuo con densità costante su C e nulla al di fuori di C :

$$f_{X,Y}(x, y) = \begin{cases} a & (x, y) \in C \\ 0 & \text{altrove} \end{cases}$$

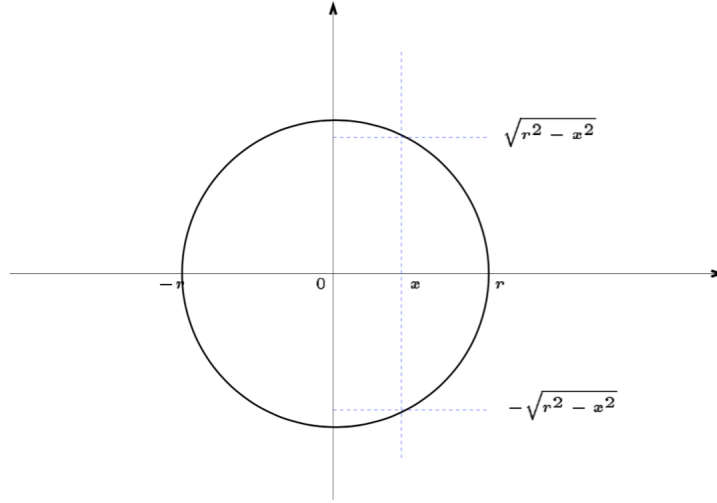
Dalla proprietà 1. della Proposizione 4.4.2 segue che il valore della costante a deve essere tale che

$$\int_C a \, dx \, dy = 1$$

cioè a è il reciproco dell'area del cerchio C : $a = 1/(\pi r^2)$. Pertanto, se (X, Y) è un punto “scelto a caso” nel cerchio C , allora (X, Y) è un vettore aleatorio assolutamente continuo con densità

$$f_{X,Y}(x, y) = \frac{1}{\pi r^2} \mathbf{1}_C(x, y)$$

Calcoliamo ora le densità marginali f_X, f_Y . Sia $x \in (-r, r)$. Allora



$$f_X(x) = \int_{\mathbb{R}} \frac{1}{\pi r^2} \mathbf{1}_C(x, y) dy = \int_{-\sqrt{r^2-x^2}}^{\sqrt{r^2-x^2}} \frac{1}{\pi r^2} dy = \frac{2}{r^2\pi} \sqrt{r^2 - x^2}$$

Se invece $x \notin (-r, r)$ allora $f_X(x) = 0$. In definitiva,

$$f_X(x) = \frac{2}{r^2\pi} \sqrt{r^2 - x^2} \mathbf{1}_{(-r,r)}(x)$$

Per motivi di simmetria vale anche che

$$f_Y(y) = \frac{2}{r^2\pi} \sqrt{r^2 - y^2} \mathbf{1}_{(-r,r)}(y)$$

Esercizio 4.4.5 (Continuazione dell'Esempio 4.1.8) Siano S, T i tempi di guasto rispettivamente del primo e del secondo componente introdotti nell'Esempio 4.1.8. Calcolare la probabilità che il primo componente si guasti prima del secondo.

(S, T) è un vettore aleatorio bidimensionale a componenti indipendenti e continuo di densità

$$f_{S,T}(s, t) = \lambda^2 e^{\lambda(s+t)} \mathbf{1}_{(0,+\infty) \times (0,+\infty)}(s, t).$$

La probabilità richiesta è

$$P(S < T) = P((S, T) \in A)$$

dove $A = \{(s, t) \in (0, +\infty) \times (0, +\infty) : s < t\}$. Quindi:

$$P(S < T) = \int_A f_{S,T}(s, t) ds dt = \lambda^2 \int_0^{+\infty} e^{-\lambda s} \int_s^{+\infty} e^{-\lambda t} dt ds = \lambda \int_0^{+\infty} e^{-2\lambda s} ds = \frac{1}{2}$$

Notiamo che la funzione di densità congiunta del vettore (S, T) verifica la condizione: $f_{S,T}(s, t) = f_S(s)f_T(t)$ per ogni $s, t > 0$. Questo fatto è comune a tutti i vettori aleatori assolutamente continui a componenti indipendenti. Si può infatti dimostrare la seguente proposizione.

Proposizione 4.4.6 *Le componenti di un vettore aleatorio assolutamente continuo sono indipendenti se e solo se ammettono una densità congiunta che può essere scritta come prodotto delle densità marginali.*

Esercizio 4.4.7 Dimostrare la Proposizione 4.4.6.

4.5 Funzioni di vettori aleatori

Siano $\mathbf{X} = (X_1, \dots, X_n)$ un vettore aleatorio n -dimensionale e $\mathbf{g} = (g_1, \dots, g_m): \mathbb{R}^n \rightarrow \mathbb{R}^m$ una funzione vettoriale. Sia inoltre $\mathbf{Y} := \mathbf{g}(\mathbf{X})$. Allora \mathbf{Y} ha componenti

$$\begin{aligned} Y_1 &= g_1(X_1, \dots, X_n) \\ Y_2 &= g_2(X_1, \dots, X_n) \\ &\vdots \\ Y_m &= g_m(X_1, \dots, X_n) \end{aligned}$$

Ci chiediamo: \mathbf{Y} è un vettore aleatorio?

Come nel caso unidimensionale, se \mathbf{X} è vettore assolutamente continuo, allora sono necessarie alcune ipotesi sulla regolarità di \mathbf{g} affinché \mathbf{Y} sia vettore aleatorio, per esempio \mathbf{g} continua a tratti.

In questa sezione ci occupiamo di determinare, quando è possibile, la densità di \mathbf{Y} a partire da quella di \mathbf{X} . In particolare, determiniamo la densità di somme di variabili aleatorie, a partire dalla loro densità congiunta. Al solito, trattiamo separatamente le funzioni di vettori aleatori discreti e assolutamente continui.

4.5.1 Funzioni di vettori aleatori discreti

Sia \mathbf{X} un vettore aleatorio discreto con densità $p_{\mathbf{X}}(\mathbf{x})$ e $P(\mathbf{X} \in S) = 1$, con S al più numerabile. Consideriamo $\mathbf{g}: S \rightarrow \mathbb{R}^m$ e $\mathbf{Y} := \mathbf{g}(\mathbf{X})$. \mathbf{Y} è chiaramente un vettore aleatorio discreto a valori in $g(S) = \{\mathbf{y} = \mathbf{g}(\mathbf{x}), \mathbf{x} \in S\}$, cioè $P(\mathbf{Y} \in g(S)) = 1$.

Per determinare la densità $p_{\mathbf{Y}}$ di \mathbf{Y} osserviamo che per ogni $\mathbf{y} \in \mathbb{R}^m$:

$$\begin{aligned} p_{\mathbf{Y}}(\mathbf{y}) &= P(\mathbf{Y} = \mathbf{y}) = P(\mathbf{g}(\mathbf{X}) = \mathbf{y}) = P\left(\bigcup_{\mathbf{x} \in S: \mathbf{g}(\mathbf{x}) = \mathbf{y}} \{\mathbf{X} = \mathbf{x}\}\right) \\ &= \sum_{\mathbf{x} \in S: \mathbf{g}(\mathbf{x}) = \mathbf{y}} P(\mathbf{X} = \mathbf{x}) = \sum_{\mathbf{x} \in S: \mathbf{g}(\mathbf{x}) = \mathbf{y}} p_{\mathbf{X}}(\mathbf{x}) \end{aligned}$$

Si noti che se $\mathbf{y} \notin g(S)$, la somma non contiene termini e si intende $p_{\mathbf{Y}}(\mathbf{y}) = 0$. Rimane così dimostrata la seguente proposizione.

Proposizione 4.5.1 *Sia \mathbf{X} un vettore aleatorio discreto con densità $p_{\mathbf{X}}(\mathbf{x})$ e $P(\mathbf{X} \in S) = 1$ e sia $\mathbf{g}: S \rightarrow \mathbb{R}^m$. Allora $\mathbf{Y} := \mathbf{g}(\mathbf{X})$ è un vettore aleatorio discreto tale che $P(\mathbf{Y} \in g(S)) = 1$ e la densità di \mathbf{Y} è*

$$p_{\mathbf{Y}}(\mathbf{y}) = \sum_{\mathbf{x} \in S: \mathbf{g}(\mathbf{x}) = \mathbf{y}} p_{\mathbf{X}}(\mathbf{x}) \quad (4.5.1)$$

Somme di variabili aleatorie discrete. In questo paragrafo deriviamo dalla Proposizione 4.5.1 una formula per densità di somme di variabili aleatorie discrete, nota la loro densità congiunta. Per maggiore semplicità espositiva, studiamo la somma di due variabili aleatorie, $X_1 + X_2$. Il risultato si estende per ricorrenza alla somma di n variabili aleatorie $X_1 + \dots + X_n$.

Sia (X_1, X_2) un vettore aleatorio discreto con densità $p_{X_1, X_2}(x_1, x_2)$ e sia Y la variabile aleatoria somma data da $Y = X_1 + X_2$.

Segue dalla formula (4.5.1) che

$$p_{X_1+X_2}(y) = \sum_{x_1, x_2: x_1+x_2=y} p_{X_1, X_2}(x_1, x_2) = \sum_{x_2} p_{X_1, X_2}(y - x_2, x_2)$$

In particolare, se X_1, X_2 sono indipendenti allora

$$p_{X_1+X_2}(y) = \sum_{x_2} p_{X_1}(y - x_2) p_{X_2}(x_2)$$

Esempio 4.5.2 (Somma di variabili aleatorie di Poisson indipendenti)

Siano X_1, X_2 variabili aleatorie indipendenti con densità di Poisson di parametri λ_1, λ_2 , rispettivamente ($X_i \sim \mathcal{P}(\lambda_i)$, $i = 1, 2$). La loro somma $X_1 + X_2$ è una variabile aleatoria

discreta che assume i valori $0, 1, \dots$ e per ogni $k = 0, 1, \dots$ abbiamo

$$\begin{aligned} P(X_1 + X_2 = k) &= \sum_{j=0}^k p_{X_1, X_2}(k-j, j) = \sum_{j=0}^k p_{X_1}(k-j) p_{X_2}(j) \\ &= \sum_{j=0}^k e^{-\lambda_1} \frac{\lambda_1^{k-j}}{(k-j)!} e^{-\lambda_2} \frac{\lambda_2^j}{j!} = \frac{e^{-(\lambda_1 + \lambda_2)}}{k!} \sum_{j=0}^k \binom{k}{j} \lambda_1^{k-j} \lambda_2^j \\ &= \frac{e^{-(\lambda_1 + \lambda_2)} (\lambda_1 + \lambda_2)^k}{k!} \end{aligned}$$

Quindi $X_1 + X_2 \sim \mathcal{P}(\lambda_1 + \lambda_2)$

Iterando il procedimento ora visto otteniamo che se X_1, \dots, X_n sono variabili aleatorie indipendenti con $X_1 \sim \mathcal{P}(\lambda_1), \dots, X_n \sim \mathcal{P}(\lambda_n)$, allora $X_1 + \dots + X_n \sim \mathcal{P}(\lambda_1 + \dots + \lambda_n)$.

Esercizio 4.5.3 (Variabile binomiale come somma di bernoulliane indipendenti)

Siano X_1, \dots, X_n n variabili aleatorie indipendenti con densità di Bernoulli di parametro $p \in (0, 1)$. Dimostrate che $X_1 + \dots + X_n \sim \mathbf{Bi}(n, p)$.

Esercizio 4.5.4 Siano X_1, \dots, X_k variabili aleatorie indipendenti con $X_1 \sim \mathbf{Bi}(n_1, p), \dots, X_k \sim \mathbf{Bin}(n_k, p)$, $p \in (0, 1)$. Dimostrate che $X_1 + \dots + X_k \sim \mathbf{Bi}(n_1 + \dots + n_k, p)$.

Trasformazioni affini di vettori aleatori discreti Siano \mathbf{A} una matrice $n \times n$ invertibile, \mathbf{b} un vettore colonna di dimensione n ($\mathbf{b} \in \mathbb{R}^n$), \mathbf{X} un vettore aleatorio discreto n -dimensionale di densità $p_{\mathbf{X}}$ e $\mathbf{Y} = \mathbf{A}\mathbf{X} + \mathbf{b}$. (I vettori \mathbf{X} , e di conseguenza \mathbf{Y} , qui vanno intesi come vettori colonna.)

Per calcolare la densità di \mathbf{Y} applichiamo (4.5.1) alla trasformazione biunivoca di \mathbb{R}^n in \mathbb{R}^n : $\mathbf{g}(\mathbf{x}) = \mathbf{A}\mathbf{x} + \mathbf{b}$ con inversa $\mathbf{g}^{-1}(\mathbf{y}) = \mathbf{A}^{-1}(\mathbf{y} - \mathbf{b})$. Poiché \mathbf{g} è biunivoca otteniamo

$$p_{\mathbf{Y}}(\mathbf{y}) = P(\mathbf{Y} = \mathbf{y}) = p_{\mathbf{X}}(\mathbf{A}^{-1}(\mathbf{y} - \mathbf{b})).$$

4.5.2 Funzioni di vettori aleatori assolutamente continui

Ci occupiamo ora di funzioni di vettori aleatori assolutamente continui. Siano \mathbf{X} un vettore aleatorio n -dimensionale assolutamente continuo con densità $f_{\mathbf{X}}$ e $\mathbf{Y} = \mathbf{g}(\mathbf{X})$, con $\mathbf{g}: \mathbb{R}^n \rightarrow \mathbb{R}^m$. Come sopra accennato e diversamente dal caso di funzioni di vettori aleatori discreti, non è detto che, applicando una funzione qualsiasi \mathbf{g} a un vettore aleatorio assolutamente continuo \mathbf{X} , la funzione $\mathbf{g}(\mathbf{X})$ sia ancora un vettore aleatorio. Perché \mathbf{Y} sia un vettore aleatorio, \mathbf{g} deve soddisfare opportune condizioni di regolarità, per esempio \mathbf{g} continua a tratti.

Se $\mathbf{Y} = \mathbf{g}(\mathbf{X})$ è un vettore aleatorio, per calcolare la funzione di ripartizione di \mathbf{Y} è sufficiente osservare che $F_{\mathbf{Y}}(\mathbf{y}) = P(\mathbf{Y} \leq \mathbf{y}) = P(\mathbf{g}(\mathbf{X}) \leq \mathbf{y})$ e applicare il punto 3. della

Proposizione 4.4.2². In questo modo riusciamo a esprimere $F_{\mathbf{Y}}(\mathbf{y})$ in funzione della densità di \mathbf{X} come:

$$F_{\mathbf{Y}}(\mathbf{y}) = P(\mathbf{X} \in A) = \int_A f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x} \quad \text{con } A := \{\mathbf{x} : \mathbf{g}(\mathbf{x}) \leq \mathbf{y}\} \quad (4.5.2)$$

Nel prossimo paragrafo useremo l'equazione (4.5.2) per studiare la somma di variabili aleatorie.

Somme di variabili aleatorie: caso di un vettore assolutamente continuo. Sia (X_1, X_2) un vettore aleatorio assolutamente continuo con densità f_{X_1, X_2} . L'equazione (4.5.2) applicata alla funzione $g(x_1, x_2) = x_1 + x_2$ fornisce per la funzione di ripartizione di $X_1 + X_2$:

$$\begin{aligned} F_{X_1+X_2}(y) &= \int_{\{(x_1, x_2) : x_1+x_2 \leq y\}} f_{X_1, X_2}(x_1, x_2) dx_1 dx_2 \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{y-x_1} f_{X_1, X_2}(x_1, x_2) dx_1 dx_2 \\ &= \int_{-\infty}^y \left(\int_{-\infty}^{\infty} f_{X_1, X_2}(x_1, x_2 - x_1) dx_1 \right) dx_2 \end{aligned}$$

Quindi, $X_1 + X_2$ è una variabile aleatoria assolutamente continua e ha densità

$$f_{X_1+X_2}(y) = \int_{-\infty}^{+\infty} f_{X_1, X_2}(x_1, y - x_1) dx_1$$

Inoltre, se X_1, X_2 sono indipendenti allora

$$f_{X_1+X_2}(y) = \int_{-\infty}^{+\infty} f_{X_1}(x_1) f_{X_2}(y - x_1) dx_1 \quad (4.5.3)$$

Calcoliamo ora le densità delle somme di alcune variabili aleatorie indipendenti assolutamente continue.

Esempio 4.5.5 (Somme di variabili aleatorie gaussiane indipendenti) Cominciamo sommando due variabili aleatorie Z_1, Z_2 gaussiane indipendenti e a media nulla, cioè $Z_1 \sim \mathcal{N}(0, \sigma_1^2)$ e $Z_2 \sim \mathcal{N}(0, \sigma_2^2)$. Segue dalla (4.5.3) che

$$\begin{aligned} f_{Z_1+Z_2}(y) &= \int_{-\infty}^{+\infty} \frac{1}{2\pi\sigma_1\sigma_2} e^{-\frac{x^2}{2\sigma_1^2} - \frac{(y-x)^2}{2\sigma_2^2}} dx \\ &= \frac{1}{\sqrt{2\pi(\sigma_1^2 + \sigma_2^2)}} e^{-\frac{y^2}{2(\sigma_1^2 + \sigma_2^2)}} \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi\tau}} e^{-\frac{(x-\nu)^2}{2\tau}} dx \end{aligned}$$

²Ricordate che “ \leq ” è la seguente relazione di ordine parziale: se $\mathbf{x}, \mathbf{y} \in \mathbb{R}^m$, allora $\mathbf{x} \leq \mathbf{y}$ se e solo se $x_k \leq y_k$ per ogni $k = 1, \dots, m$

dove

$$\nu := \frac{y\sigma_1^2}{\sigma_1^2 + \sigma_2^2} \quad \text{e} \quad \tau := \frac{\sigma_1^2\sigma_2^2}{\sigma_1^2 + \sigma_2^2}$$

D'altro canto

$$\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi\tau}} e^{-\frac{(x-\nu)^2}{2\tau}} dx = 1$$

Quindi

$$f_{Z_1+Z_2}(y) = \frac{1}{\sqrt{2\pi(\sigma_1^2 + \sigma_2^2)}} e^{-\frac{y^2}{2(\sigma_1^2 + \sigma_2^2)}}$$

cioè $Z_1 + Z_2 \sim \mathcal{N}(0, \sigma_1^2 + \sigma_2^2)$.

Siano ora X_1, X_2 due variabili aleatorie indipendenti con $X_1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$ e $X_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$. Allora $X_1 + X_2$ ha la stessa densità di $(Z_1 + Z_2) + (\mu_1 + \mu_2)$, che è trasformazione lineare della variabile aleatoria gaussiana $Z_1 + Z_2$, come abbiamo appena dimostrato. Quindi: $X_1 + X_2 \sim \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$.

Iterando il procedimento ora visto otteniamo che se X_1, \dots, X_n sono variabili aleatorie indipendenti e gaussiane con $X_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$, $\forall i = 1, \dots, n$, allora $\sum_{i=1}^n X_i \sim \mathcal{N}(\sum_{i=1}^n \mu_i, \sum_{i=1}^n \sigma_i^2)$. In breve: la somma di variabili aleatorie gaussiane indipendenti è gaussiana di parametri la somma dei parametri.

Esempio 4.5.6 Siano X_1, X_2 due variabili aleatorie indipendenti entrambe con densità esponenziale di parametro $\lambda > 0$. Calcoliamo la densità di $X_1 + X_2$.

Applicando (4.5.3) abbiamo:

$$f_{X_1+X_2}(y) = \begin{cases} \int_0^y \lambda e^{-\lambda u} \lambda e^{-\lambda(y-u)} du = \lambda^2 e^{-\lambda y} y & \text{se } y > 0 \\ 0 & \text{se } y \leq 0 \end{cases}$$

Procedendo per induzione su n si può dimostrare che se X_1, \dots, X_n sono variabili aleatorie indipendenti tali che $X_i \sim \mathcal{E}(\lambda) \forall i = 1, \dots, n$, allora la densità di $X_1 + \dots + X_n$ è

$$f_{X_1+\dots+X_n}(x) = \frac{\lambda^n}{(n-1)!} x^{n-1} e^{-\lambda x} \mathbf{1}_{(0,+\infty)}(x) \quad (4.5.4)$$

Definizione 4.5.7 La densità (4.5.4) è detta densità Gamma di parametri n e λ e scriveremo $\Gamma(n, \lambda)$.

Concludiamo questa sezione con la seguente proposizione sulle funzioni biunivoche di vettori aleatori assolutamente continui. Essa è l'analogo della Proposizione 2.6.5 per funzioni di variabili aleatorie assolutamente continue.

Proposizione 4.5.8 Siano U e V due insiemi aperti di \mathbb{R}^n e sia \mathbf{g} un'applicazione biunivoca da U su V differenziabile con continuità insieme alla sua inversa \mathbf{g}^{-1} . Sia \mathbf{X} un vettore

aleatorio n -dimensionale assolutamente continuo con densità $f_{\mathbf{X}}$ e tale che $P(\mathbf{X} \in U) = 1$. Allora $\mathbf{Y} := \mathbf{g}(\mathbf{X})$ è un vettore aleatorio assolutamente continuo con densità data da

$$f_{\mathbf{Y}}(\mathbf{y}) = \mathbf{1}_V(\mathbf{y}) f_{\mathbf{X}}[\mathbf{g}^{-1}(\mathbf{y})] |\det [\mathbf{J}(\mathbf{g}^{-1}(\mathbf{y}))]| \quad (4.5.5)$$

dove $\mathbf{J}(\mathbf{g}^{-1}(\mathbf{y}))$ indica la matrice jacobiana associata alla funzione \mathbf{g}^{-1} calcolata in \mathbf{y} :

$$\mathbf{J}(\mathbf{g}^{-1}) = \begin{pmatrix} \frac{\partial g_1^{-1}}{\partial y_1} & \frac{\partial g_1^{-1}}{\partial y_2} & \cdots & \frac{\partial g_1^{-1}}{\partial y_n} \\ \frac{\partial g_2^{-1}}{\partial y_1} & \frac{\partial g_2^{-1}}{\partial y_2} & \cdots & \frac{\partial g_2^{-1}}{\partial y_n} \\ \vdots & \vdots & \cdots & \vdots \\ \frac{\partial g_n^{-1}}{\partial y_1} & \frac{\partial g_n^{-1}}{\partial y_2} & \cdots & \frac{\partial g_n^{-1}}{\partial y_n} \end{pmatrix}$$

Esercizio 4.5.9 Dimostrare la Proposizione 4.5.8 per il caso di un vettore bidimensionale (X_1, X_2) per cui $U = V = \mathbb{R}^2$. (Usare l'Equazione 4.5.2).

Esercizio 4.5.10 Il lettore scriva la formula (4.5.5) per $n = 1$ e la confronti con la formula (2.6.1) fornita nel caso di variabili aleatorie assolutamente continue.

Esempio 4.5.11 (Trasformazioni affini di vettori aleatori assolutamente continui)

Siano \mathbf{A} una matrice $n \times n$ invertibile, \mathbf{b} un vettore colonna di \mathbb{R}^n , \mathbf{X} un vettore aleatorio (colonna) assolutamente continuo n -dimensionale di densità $f_{\mathbf{X}}$ e $\mathbf{Y} = \mathbf{A}\mathbf{X} + \mathbf{b}$. Per calcolare la densità di \mathbf{Y} possiamo applicare (4.5.5) alla trasformazione biunivoca di \mathbb{R}^n in \mathbb{R}^n : $\mathbf{g}(\mathbf{x}) = \mathbf{A}\mathbf{x} + \mathbf{b}$ con inversa $\mathbf{g}^{-1}(\mathbf{y}) = \mathbf{A}^{-1}(\mathbf{y} - \mathbf{b})$. Infatti tutte le ipotesi della Proposizione 4.5.8 sono soddisfatte e la densità $f_{\mathbf{Y}}$ di $\mathbf{Y} = \mathbf{A}\mathbf{X} + \mathbf{b}$ risulta

$$f_{\mathbf{Y}}(\mathbf{y}) = f_{\mathbf{X}}(\mathbf{A}^{-1}(\mathbf{y} - \mathbf{b})) |\det(\mathbf{A}^{-1})| \quad (4.5.6)$$

Esercizio 4.5.12 Siano X_1, X_2 due variabili aleatorie indipendenti e uniformi sull'intervallo $(0, 1)$ e siano $Y_1 = X_1 + X_2$ e $Y_2 = X_1 - X_2$. Verificate che il vettore aleatorio (Y_1, Y_2) è uniforme sul quadrato di vertici $(0, 0), (1, 1), (2, 0), (1, -1)$.

4.6 *Vettori aleatori indipendenti

Siano X_1, \dots, X_n n variabili aleatorie indipendenti. Fissato $m < n$, consideriamo due funzioni $\mathbf{g}: \mathbb{R}^m \rightarrow \mathbb{R}^k$, $\mathbf{h}: \mathbb{R}^{n-m} \rightarrow \mathbb{R}^l$ tali che $\mathbf{W} = \mathbf{g}(X_1, \dots, X_m)$ e $\mathbf{Z} = \mathbf{h}(X_{m+1}, \dots, X_n)$ sono ancora vettori aleatori. È facile mostrare che gli eventi esprimibili in termini di \mathbf{W} e quelli esprimibili in termini di \mathbf{Z} sono indipendenti in quanto i primi dipendono soltanto da X_1, \dots, X_m e i secondi soltanto da X_{m+1}, \dots, X_n che sono gruppi di variabili tra di loro tutte indipendenti. Per i vettori aleatori \mathbf{W}, \mathbf{Z} vale quindi che

$$P(\mathbf{W} \in A, \mathbf{Z} \in B) = P(\mathbf{W} \in A)P(\mathbf{Z} \in B)$$

per ogni scelta di domini regolari $A \subset \mathbb{R}^k$ e $B \subset \mathbb{R}^l$.

Alla luce di quanto fin qui detto, appare naturale la seguente definizione di vettori aleatori indipendenti:

Definizione 4.6.1 Siano $\mathbf{X}_1, \dots, \mathbf{X}_n$ n vettori aleatori definiti sullo stesso spazio di probabilità (Ω, \mathcal{F}, P) di dimensione rispettivamente m_1, \dots, m_n . Diciamo che sono indipendenti se

$$P(\mathbf{X}_1 \in B_1, \dots, \mathbf{X}_n \in B_n) = P(\mathbf{X}_1 \in B_1) \cdots P(\mathbf{X}_n \in B_n) \quad (4.6.1)$$

per ogni scelta di domini regolari $B_1 \in \mathbb{R}^{m_1}, \dots, B_n \in \mathbb{R}^{m_n}$.

Per funzioni vettoriali di vettori aleatori indipendenti, si può inoltre dimostrare la seguente proposizione.

Proposizione 4.6.2 Siano $\mathbf{X}_1, \dots, \mathbf{X}_n$ vettori aleatori indipendenti di dimensione rispettivamente m_1, \dots, m_n e siano $\mathbf{g}_1, \dots, \mathbf{g}_n$ delle funzioni definite da $\mathbf{g}_1: \mathbb{R}^{m_1} \rightarrow \mathbb{R}^{k_1}, \dots, \mathbf{g}_n: \mathbb{R}^{m_n} \rightarrow \mathbb{R}^{k_n}$. Allora i vettori aleatori $\mathbf{Y}_1 = \mathbf{g}_1(\mathbf{X}_1), \dots, \mathbf{Y}_n = \mathbf{g}_n(\mathbf{X}_n)$ sono indipendenti.

4.7 Valore atteso di funzioni di vettori aleatori

Sia \mathbf{X} un vettore n -dimensionale, $g: \mathbb{R}^n \rightarrow \mathbb{R}$ una funzione a valori reali tale che $Y = g(\mathbf{X})$ è una variabile aleatoria. Analogamente al caso di funzioni di variabili aleatorie, possiamo calcolare $E(Y)$ evitando di determinare esplicitamente la densità di Y . Infatti la Proposizione 3.1.12 si estende al caso di variabili aleatorie definite come funzioni di vettori aleatori nel seguente modo.

Proposizione 4.7.1 Sia \mathbf{X} un vettore aleatorio discreto che assume valori in S e ha densità $p_{\mathbf{X}}$. Siano $g: \mathbb{R}^n \rightarrow \mathbb{R}$ e $Y := g(\mathbf{X})$ una variabile aleatoria. Se $\sum_{\mathbf{x} \in S} |g(\mathbf{x})| p_{\mathbf{X}}(\mathbf{x}) < +\infty$, allora Y ammette valore atteso e

$$E(Y) = \sum_{\mathbf{x} \in S} g(\mathbf{x}) p_{\mathbf{X}}(\mathbf{x}) \quad (4.7.1)$$

Sia \mathbf{X} un vettore aleatorio assolutamente continuo con densità $f_{\mathbf{X}}$, $g: \mathbb{R}^n \rightarrow \mathbb{R}$ e $Y = g(\mathbf{X})$ una variabile aleatoria. Se $\int_{\mathbb{R}^n} |g(x_1 \cdots x_n)| f_{\mathbf{X}}(x_1, \dots, x_n) dx_1 \cdots dx_n < +\infty$, allora Y ammette valore atteso e

$$E(Y) = \int_{\mathbb{R}^n} g(x_1 \cdots x_n) f_{\mathbf{X}}(x_1, \dots, x_n) dx_1 \cdots dx_n \quad (4.7.2)$$

Due interessanti applicazioni della precedente proposizione riguardano il calcolo di media e varianza della somma di variabili aleatorie.

Corollario 4.7.2 Siano X_1 e X_2 variabili aleatorie definite sul medesimo spazio di probabilità e che ammettono media. Allora anche $X_1 + X_2$ ammette media e

$$E(X_1 + X_2) = E(X_1) + E(X_2).$$

Dimostrazione Supponiamo che il vettore aleatorio (X_1, X_2) sia assolutamente continuo con densità di probabilità f_{X_1, X_2} . Dalla disuguaglianza triangolare: $|x + y| \leq |x| + |y|$ discende

$$\begin{aligned} \int_{\mathbb{R}^2} |x + y| f_{X_1, X_2}(x, y) dx dy &\leq \int_{\mathbb{R}^2} |x| f_{X_1, X_2}(x, y) dx dy + \int_{\mathbb{R}^2} |y| f_{X_1, X_2}(x, y) dx dy \\ &= \int_{\mathbb{R}} |x| f_{X_1}(x) dx + \int_{\mathbb{R}} |y| f_{X_2}(y) dy < +\infty \end{aligned}$$

e quindi $X_1 + X_2$ ammette media. Applicando ora la Proposizione 4.7.1 a $g(x, y) = x + y$ risulta:

$$\begin{aligned} E(X_1 + X_2) &= \int_{\mathbb{R}^2} (x + y) f_{X_1, X_2}(x, y) dx dy \\ &= \int_{\mathbb{R}} x \left(\int_{\mathbb{R}} f_{X_1, X_2}(x, y) dy \right) dx + \int_{\mathbb{R}} y \left(\int_{\mathbb{R}} f_{X_1, X_2}(x, y) dx \right) dy \\ &= \int_{\mathbb{R}} x f_{X_1}(x) dx + \int_{\mathbb{R}} y f_{X_2}(y) dy = E(X_1) + E(X_2) \end{aligned}$$

La dimostrazione procede analogamente se X_1 e X_2 sono variabili aleatorie discrete. ■

Nota 4.7.3 È importante osservare che la media di $X_1 + X_2$ dipende soltanto dalle densità marginali del vettore aleatorio (X_1, X_2) . In generale, la media della somma di $n \geq 2$ variabili aleatorie X_1, \dots, X_n è data dalla somma delle n medie:

$$E(X_1 + \dots + X_n) = E(X_1) + \dots + E(X_n) \quad (4.7.3)$$

Corollario 4.7.4 Siano X_1 e X_2 variabili aleatorie indipendenti e che ammettono media. Allora anche $X_1 X_2$ ammette media e

$$E(X_1 X_2) = E(X_1) E(X_2)$$

Dimostrazione Supponiamo che X_1 e X_2 siano assolutamente continue con densità rispettivamente f_{X_1} e f_{X_2} . Allora:

$$\int_{\mathbb{R}} \int_{\mathbb{R}} |xy| f_{X_1}(x) f_{X_2}(y) dx dy = \int_{\mathbb{R}} |x| f_{X_1}(x) dx \cdot \int_{\mathbb{R}} |y| f_{X_2}(y) dy < +\infty$$

e $E(X_1 X_2)$ esiste per la Proposizione 4.7.1 applicata alla funzione $g(x, y) = xy$. Inoltre, dalla Proposizione 4.7.1 discende che:

$$E(X_1 X_2) = \int_{\mathbb{R}} \int_{\mathbb{R}} xy f_{X_1}(x) f_{X_2}(y) dx dy = \int_{\mathbb{R}} x f_{X_1}(x) dx \cdot \int_{\mathbb{R}} y f_{X_2}(y) dy = E(X_1) E(X_2) \quad \blacksquare$$

Nota 4.7.5 Iterando il procedimento nella dimostrazione del Corollario 4.7.4 è immediato verificare che se X_1, \dots, X_n sono n variabili aleatorie indipendenti che ammettono media allora anche $\prod_{i=1}^n X_i$ ammette media e $E(\prod_{i=1}^n X_i) = \prod_{i=1}^n E(X_i)$.

Occupiamoci ora del problema del calcolo della varianza della somma di variabili aleatorie.

Corollario 4.7.6 Se X_1 e X_2 hanno varianza (finita), rispettivamente $\text{Var}(X_1)$ e $\text{Var}(X_2)$, allora anche $X_1 + X_2$ ha varianza finita e

$$\text{Var}(X_1 + X_2) = \text{Var}(X_1) + \text{Var}(X_2) + 2 E[(X_1 - E(X_1))(X_2 - E(X_2))] \quad (4.7.4)$$

Inoltre, se X_1, X_2 sono indipendenti allora

$$\text{Var}(X_1 + X_2) = \text{Var}(X_1) + \text{Var}(X_2) \quad (4.7.5)$$

Dimostrazione Poiché $((X_1 + X_2) - E(X_1 + X_2))^2 = [(X_1 - E(X_1)) + (X_2 - E(X_2))]^2 \leq 2[(X_1 - E(X_1))^2 + (X_2 - E(X_2))^2]$, allora $\text{Var}(X_1 + X_2) = E[((X_1 + X_2) - E(X_1 + X_2))^2] = E[((X_1 - E(X_1)) + (X_2 - E(X_2)))^2] \leq 2[E(X_1 - E(X_1))^2 + E(X_2 - E(X_2))^2] = 2(\text{Var}(X_1) + \text{Var}(X_2))$. Quindi se X_1 e X_2 ammettono varianza, anche $X_1 + X_2$ la ammette. Inoltre

$$\begin{aligned}\text{Var}(X_1 + X_2) &= E[((X_1 - E(X_1)) + (X_2 - E(X_2)))^2] \\ &= E[(X_1 - E(X_1))^2 + (X_2 - E(X_2))^2 + 2(X_1 - E(X_1))(X_2 - E(X_2))] \\ &= E[(X_1 - E(X_1))^2] + E[(X_2 - E(X_2))^2] + 2E[(X_1 - E(X_1))(X_2 - E(X_2))]\end{aligned}$$

[dove l'ultima eguaglianza deriva dal Corollario 4.7.2 applicato alla somma delle variabili $(X_1 - E(X_1))^2$, $(X_2 - E(X_2))^2$ e $(X_1 - E(X_1))(X_2 - E(X_2))$]

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

Per completare la dimostrazione, basta notare che se X_1, X_2 sono indipendenti, allora, per la Proposizione 4.6.2, anche $X_1 - E(X_1), X_2 - E(X_2)$ sono indipendenti e $E[(X_1 - E(X_1))(X_2 - E(X_2))] = 0$ in virtù del Corollario 4.7.4. ■

Esercizio 4.7.7 Si dimostri che la varianza della somma di n variabili aleatorie X_1, \dots, X_n è data da:

$$\text{Var}(X_1 + \dots + X_n) = \sum_{i=1}^n \text{Var}(X_i) + 2 \sum_{i=1}^{n-1} \sum_{j=i+1}^n E[(X_i - E(X_i))(X_j - E(X_j))] \quad (4.7.6)$$

Esempio 4.7.8 Sia $X \sim \mathbf{Bi}(n, p)$. Sappiamo dall'Esercizio 4.5.4 che la variabile aleatoria X ha la stessa densità della somma di n variabili aleatorie –chiamiamole X_1, \dots, X_n – indipendenti con densità di Bernoulli di parametro p . Allora ritroviamo

$$\begin{aligned}E(X) &= E\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n E(X_i) = \sum_{i=1}^n p = np \quad [\text{per il Corollario 4.7.2}] \\ \text{Var}(X) &= \text{Var}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \text{Var}(X_i) = \sum_{i=1}^n p(1-p) = np(1-p)\end{aligned}$$

dove l'ultima eguaglianza deriva dall'indipendenza fra le X_1, \dots, X_n e dall'equazione (4.7.6).

4.8 Covarianza, Coefficiente di correlazione

Abbiamo visto che se X_1, \dots, X_n sono variabili aleatorie con varianza finita allora:

$$\text{Var}(X_1 + \dots + X_n) = \sum_{i=1}^n \text{Var}(X_i) + 2 \sum_{i=1}^{n-1} \sum_{j=i+1}^n E[(X_i - E(X_i))(X_j - E(X_j))]$$

Gli addendi nell'ultima sommatoria sono di per sè rilevanti in probabilità. Quindi introduciamo la seguente definizione.

Definizione 4.8.1 Siano X_1, X_2 due variabili aleatorie definite sul medesimo spazio di probabilità e che ammettono varianza. Si definisce covarianza di X_1, X_2 il numero

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

Se $0 < \text{Var}(X_1)$, $0 < \text{Var}(X_2)$, si definisce coefficiente di correlazione di X_1, X_2 il numero:

$$\rho_{X_1, X_2} = \frac{\text{Cov}(X_1, X_2)}{\sqrt{\text{Var}(X_1) \text{Var}(X_2)}}$$

Osserviamo che la covarianza di X_1 e X_2 è ben definita per variabili aleatorie X_1, X_2 con varianza finita. Infatti, sappiamo dal Corollario 4.7.6 che se X_1, X_2 hanno varianza finita, anche la varianza di $X_1 + X_2$ è finita ed è data da

$$\text{Var}(X_1 + X_2) = \text{Var}(X_1) + \text{Var}(X_2) + 2 \text{Cov}(X_1, X_2)$$

Segue che necessariamente anche $\text{Cov}(X_1, X_2)$ è un numero (finito).

Covarianza e coefficiente di correlazione godono delle proprietà elencate nella seguente proposizione.

Proposizione 4.8.2 Siano X_1, X_2, X_3 variabili aleatorie con varianza finita e $a, b \in \mathbb{R}$. Allora

1. $\text{Cov}(X_1, X_2) = \text{Cov}(X_2, X_1)$;
2. $\text{Cov}(aX_1, X_2) = a \text{Cov}(X_1, X_2)$;
3. $\text{Cov}(X_1 + X_2, X_3) = \text{Cov}(X_1, X_3) + \text{Cov}(X_2, X_3)$;
4. $\text{Cov}(X_1, X_2) = E(X_1 X_2) - E(X_1) E(X_2)$;
5. se X_1, X_2 sono indipendenti allora $\text{Cov}(X_1, X_2) = 0$;
6. $|\rho_{X_1, X_2}| \leq 1$ e $|\rho_{X_1, X_2}| = 1$ se e solo se esistono $a, b \in \mathbb{R}$ tali che $P(X_2 = aX_1 + b) = 1$.
Inoltre in tal caso:

$$a = \frac{\text{Cov}(X_1, X_2)}{\text{Var}(X_1)} \quad e \quad b = E(X_2) - \frac{E(X_1) \text{Cov}(X_1, X_2)}{\text{Var}(X_1)}$$

Dimostrazione Le proprietà 1.–5. seguono immediatamente dalle proprietà della media e la dimostrazione viene lasciata per esercizio al lettore.

La dimostrazione della proprietà 6. è mutuata da [7], pag. 329 e si basa sulle proprietà della varianza. Siano σ_1^2, σ_2^2 le varianze di X_1, X_2 , rispettivamente. Allora

$$\begin{aligned} 0 &\leq \text{Var} \left(\frac{X_1}{\sigma_1} + \frac{X_2}{\sigma_2} \right) = \frac{\text{Var}(X_1)}{\sigma_1^2} + \frac{\text{Var}(X_2)}{\sigma_2^2} + 2 \text{Cov} \left(\frac{X_1}{\sigma_1}, \frac{X_2}{\sigma_2} \right) \\ &= \frac{\sigma_1^2}{\sigma_1^2} + \frac{\sigma_2^2}{\sigma_2^2} + 2 \frac{\text{Cov}(X_1, X_2)}{\sigma_1 \sigma_2} \quad [\text{per il punto 2.}] \\ &= 2(1 + \rho_{X_1, X_2}) \end{aligned}$$

da cui otteniamo

$$\rho_{X_1, X_2} \geq -1$$

Inoltre,

$$0 \leq \text{Var} \left(\frac{X_1}{\sigma_1} - \frac{X_2}{\sigma_2} \right) = \frac{\text{Var}(X_1)}{\sigma_1^2} + \frac{\text{Var}(X_2)}{\sigma_2^2} - 2 \frac{\text{Cov}(X_1, X_2)}{\sigma_1 \sigma_2} = 2(1 - \rho_{X_1, X_2})$$

e quindi

$$\rho_{X_1, X_2} \leq 1$$

Per dimostrare la seconda parte della proprietà 6., osserviamo che $\rho_{X_1, X_2} = 1$ se e solo se $\text{Var}(X_1/\sigma_1 - X_2/\sigma_2) = 0$. Segue quindi dalle proprietà della varianza che

$$\rho_{X_1, X_2} = 1 \text{ se e solo se } P \left(\frac{X_1}{\sigma_1} - \frac{X_2}{\sigma_2} = \frac{E(X_1)}{\sigma_1} - \frac{E(X_2)}{\sigma_2} \right) = 1$$

Inoltre, $\rho_{X_1, X_2} = 1$ se e solo se $\text{Cov}(X_1, X_2) = \sigma_1 \sigma_2$ e quindi $\rho_{X_1, X_2} = 1$ se e solo se

$$X_2 = E(X_2) + \frac{\text{Cov}(X_1, X_2)}{\sigma_1^2} (X_1 - E(X_1))$$

Invece, per $\rho_{X_1, X_2} = -1$ valgono le seguenti equivalenze che completano la dimostrazione: $\rho_{X_1, X_2} = -1$ se e solo se $\text{Cov}(X_1, X_2) = -\sigma_1 \sigma_2$ se e solo se $\text{Var}(X_1/\sigma_1 + X_2/\sigma_2) = 0$ se e solo se $P(X_1/\sigma_1 + X_2/\sigma_2 = E(X_1)/\sigma_1 + E(X_2)/\sigma_2) = 1$ se e solo se

$$X_2 = E(X_2) + \frac{\text{Cov}(X_1, X_2)}{\sigma_1^2} (X_1 - E(X_1)) \quad \blacksquare$$

Nota 4.8.3 Il punto 6. della proposizione precedente illustra un noto risultato della teoria della *regressione lineare*: esiste un legame di tipo lineare fra le variabili aleatorie X_1 e X_2 (cioè $X_2 = aX_1 + b$) se e solo se $\rho(X_1, X_2) = \pm 1$, inoltre $\rho(X_1, X_2) = -1$ implica $\text{Cov}(X_1, X_2) < 0$ e $a < 0$ mentre $\rho(X_1, X_2) = 1$ implica $\text{Cov}(X_1, X_2) > 0$ e $a > 0$.

Nota 4.8.4 La proprietà 5. non può essere invertita come mostra il seguente controesempio:

Esempio 4.8.5 Sia X_1 una variabile aleatoria discreta con densità uniforme su $\{-1, 0, 1\}$ e sia $X_2 = X_1^2$. Allora $E(X_1) = 0$ in quanto X_1 è una variabile aleatoria simmetrica e $E(X_1 X_2) = E(X_1^3) = (-1)^3/3 + 1^3/3 = 0$, da cui $\text{Cov}(X_1, X_2) = 0$. Ma, chiaramente, X_1 e X_2 non sono indipendenti.

Esercizio 4.8.6 Dimostrate che $\text{Cov}(X, a) = 0$ e $\text{Cov}(X + a, Y) = \text{Cov}(X, Y)$ per ogni $a \in \mathbb{R}$.

Esercizio 4.8.7 Siano X_1, \dots, X_m e Y_1, \dots, Y_n variabili aleatorie che ammettono varianza e $a_1, \dots, a_m, b_1, \dots, b_n \in \mathbb{R}$. Dimostrate che

$$\text{Cov} \left(\sum_{i=1}^m a_i X_i, \sum_{j=1}^n b_j Y_j \right) = \sum_{i=1}^m \sum_{j=1}^n a_i b_j \text{Cov}(X_i, Y_j)$$

Esempio* 4.8.8 Da un'urna contenente b biglie bianche e r rosse, si estraggono n biglie senza rimpiazzo e X rappresenta il numero di biglie bianche pescate. Allora X ha densità ipergeometrica $X \sim \mathbf{Iperg}(b + r, r, n)$:

$$p_X(k) = \begin{cases} \frac{\binom{b}{k} \binom{r}{n-k}}{\binom{b+r}{n}} & k = 0 \vee (n-r), \dots, b \wedge n \\ 0 & \text{altrove.} \end{cases}$$

Per calcolare media e varianza di X possiamo procedere analiticamente, calcolando esplicitamente

$$E(X) = \sum_{k=0 \vee (n-r)}^{b \wedge n} \frac{k \binom{b}{k} \binom{r}{n-k}}{\binom{b+r}{n}} \quad \text{e} \quad E(X^2) = \sum_{k=0 \vee (n-r)}^{b \wedge n} \frac{k^2 \binom{b}{k} \binom{r}{n-k}}{\binom{b+r}{n}}.$$

Il conto è fattibile, e il lettore appassionato di proprietà dei coefficienti binomiali è invitato ad eseguirlo come esercizio. Noi daremo qui un procedimento più “probabilistico”.

Supponiamo che le biglie siano estratte sequenzialmente e definiamo le variabili X_1, \dots, X_n come

$$X_i = \begin{cases} 1 & \text{se la } i\text{-esima biglia è bianca} \\ 0 & \text{se la } i\text{-esima biglia è rossa,} \end{cases}$$

ovviamente $X = X_1 + \dots + X_n$. Per calcolare $E(X)$ osserviamo che:

$$E(X) = E(X_1 + \dots + X_n) = E(X_1) + \dots + E(X_n),$$

quindi ci basterà calcolare $E(X_1), \dots, E(X_n)$. Poiché ognuna delle variabili X_i assume solo i valori 0 e 1 (le X_i sono cioè variabili di Bernoulli) abbiamo che $E(X_k) = P(X_k = 1)$, e ci siamo ricondotti a calcolare $P(X_k = 1)$. A tal fine pensiamo di numerare le $b + r$ biglie contenute nell'urna in modo tale che le biglie numerate con i numeri $1, \dots, b$ siano bianche e quelle numerate con i numeri $b + 1, \dots, b + r$ siano rosse. In questo senso possiamo pensare ad ogni risultato del nostro esperimento aleatorio di n estrazioni di biglie dall'urna, come a un punto nello spazio degli eventi elementari

$$\Omega := \{(x_1, x_2, \dots, x_n) : x_i = 1, \dots, b + r, \forall i = 1, \dots, n, \text{ e } x_i \neq x_j \text{ se } i \neq j\}.$$

Chiaramente ogni sequenza di biglie ha la stessa probabilità di essere estratta, cioè Ω è uno spazio equiprobabile finito, e le probabilità possono essere calcolate come casi favorevoli su casi possibili. Per i casi possibili si ha

$$|\Omega| = (b + r)(b + r - 1) \cdots (b + r - n + 1)$$

in quanto la prima biglia può essere scelta in $b + r$ modi e, per ogni scelta della prima, la seconda seconda può essere scelta in $b + r - 1$ modi etc. Per i casi favorevoli all'evento $X_i = 1$, osserviamo che questo si verifica se e solo se l' i -esima biglia pescata è bianca. Quindi fissiamo l' i -esima biglia in b modi, e poi fissiamo le rimanenti $n - 1$ biglie in $(b + r - 1) \cdots (b + r - n + 1)$ modi. In definitiva:

$$P(X_i = 1) = \frac{b(b + r - 1)(b + r - 2) \cdots (b + r - n + 1)}{(b + r)(b + r - 1) \cdots (b + r - n + 1)} = \frac{b}{b + r}$$

Segue che $E(X_i) = b/(b+r)$ per ogni $i = 1, \dots, n$ da cui $E(X) = nb/(b+r)$.

Il risultato $P(X_1 = 1) = P(X_2 = 1) = \dots = P(X_n = 1)$ è in un certo senso stupefacente; si potrebbe infatti pensare che poiché l'estrazione dall'urna della prima biglia modifica il contenuto dell'urna, la probabilità che alla seconda estrazione venga estratta una biglia bianca debba essere necessariamente differente dalla probabilità di ottenere bianca alla prima estrazione. Così non è e il lettore che non si fidasse della precedente deduzione è invitato a calcolare $P(X_2 = 1)$ mediante la formula delle probabilità totali:

$$P(X_2 = 1) = P(X_2 = 1|X_1 = 0)P(X_1 = 0) + P(X_2 = 1|X_1 = 1)P(X_1 = 1)$$

Per quanto riguarda la varianza di X osserviamo che

$$\text{Var}(X) = \text{Var}(X_1 + \dots + X_n) = \sum_{i=1}^n \text{Var}(X_i) + \sum_{i \neq j} \text{Cov}(X_i, X_j)$$

Poiché $X_i \sim \text{Be}(b/(b+r))$ si ha

$$\text{Var}(X_i) = \frac{b}{b+r} \left(1 - \frac{b}{b+r}\right)$$

Ci rimane ora da calcolare $\text{Cov}(X_i, X_j) = E(X_i X_j) - E(X_i)E(X_j)$ per $i \neq j$. Poiché $X_i X_j \neq 0$ se e solo se $X_i = 1$ e $X_j = 1$ e in tal caso $X_i X_j = 1$, allora $E(X_i X_j) = P(X_i = 1, X_j = 1)$. Contiamo ora i casi favorevoli all'evento "l' i -esima e la j -esima biglia sono bianche". Abbiamo b modi di scegliere l' i -esima biglia, per ognuno dei quali ne abbiamo $b-1$ di scegliere la j -esima. Possiamo disporre le rimanenti $b+r-2$ in $(b+r-2) \cdot \dots \cdot (b+r-n+1)$ modi. In definitiva:

$$\begin{aligned} P(X_i = 1, X_j = 1) &= \frac{b(b-1)(b+r-2)(b+r-3) \cdot \dots \cdot (b+r-n+1)}{(b+r)(b+r-1) \cdot \dots \cdot (b+r-n+1)} \\ &= \frac{b(b-1)}{(b+r)(b+r-1)} \end{aligned}$$

e

$$\text{Cov}(X_i, X_j) = \frac{b(b-1)}{(b+r)(b+r-1)} - \frac{b^2}{(b+r)^2}$$

Quindi

$$\begin{aligned} \text{Var}(X) &= \frac{nb}{b+r} \left(1 - \frac{b}{b+r}\right) + (n^2 - n) \left[\frac{b(b-1)}{(b+r)(b+r-1)} - \frac{b^2}{(b+r)^2} \right] \\ &= \frac{nbr}{(b+r)^2} \left(1 - \frac{n-1}{b+r-1}\right) \end{aligned}$$

4.8.1 Matrice di covarianza

Siano X_1, \dots, X_n n variabili aleatorie che ammettono varianza. Per ciascuna coppia (X_i, X_j) , ($i \neq j$) calcoliamo la covarianza $\text{Cov}(X_i, X_j)$ e organizziamo tutte le covarianze in una matrice.

Definizione 4.8.9 Sia $\mathbf{X} = (X_1, \dots, X_n)$ un vettore aleatorio n -dimensionale tale che siano definite $\text{Var}(X_1), \dots, \text{Var}(X_n)$. Si chiama matrice di covarianza di \mathbf{X} la matrice $n \times n$ $\mathbf{C}_{\mathbf{X}} = (c_{ij})_{i,j=1,\dots,n}$ il cui elemento di posto (i, j) è $c_{ij} = \text{Cov}(X_i, X_j)$.

Proposizione 4.8.10 Sia $\mathbf{C}_{\mathbf{X}}$ la matrice di covarianza di un vettore aleatorio \mathbf{X} . Allora

1. $\mathbf{C}_\mathbf{X}$ è una matrice simmetrica e semidefinita positiva³.
2. Se $\mathbf{A} = (a_{ij})_{i,j}$ è una matrice $m \times n$ e \mathbf{b} è un vettore di dimensione m allora la matrice di covarianza di $\mathbf{Y} = \mathbf{AX} + \mathbf{b}$ è

$$\mathbf{C}_\mathbf{Y} = \mathbf{AC}_\mathbf{X}\mathbf{A}^T \quad (4.8.1)$$

Dimostrazione

1. $\mathbf{C}_\mathbf{X}$ è una matrice simmetrica in quanto $c_{ij} = \text{Cov}(X_i, X_j) = \text{Cov}(X_j, X_i) = c_{ji}$. Sia $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_n)^T$ un vettore di \mathbb{R}^n . Allora, per $i = 1, \dots, m$ e $j = 1, \dots, n$:

$$\begin{aligned} \boldsymbol{\lambda}^T \mathbf{C}_\mathbf{X} \boldsymbol{\lambda} &= \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j c_{ij} = \text{E} \left(\sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j (X_i - \text{E}(X_i))(X_j - \text{E}(X_j)) \right) \\ &= \text{E} \left(\left(\sum_{i=1}^n \lambda_i (X_i - \text{E}(X_i)) \right)^2 \right) \geq 0 \end{aligned}$$

2. Se le componenti di \mathbf{X} hanno varianza finita, allora anche le componenti di $\mathbf{Y} = \mathbf{AX} + \mathbf{b}$ hanno varianza finita e quindi ha senso considerarne la matrice di covarianza $\mathbf{C}_\mathbf{Y}$. Sia \tilde{c}_{ij} l'elemento di posto (i, j) di $\mathbf{C}_\mathbf{Y}$. Allora

$$\tilde{c}_{ij} = \text{Cov}(Y_i, Y_j) = \text{Cov} \left(\sum_{k=1}^n a_{ik} X_k + b_i, \sum_{l=1}^n a_{jl} X_l + b_j \right) = \sum_{k=1}^n \sum_{l=1}^n a_{ik} a_{jl} \text{Cov}(X_k, X_l)$$

è l'elemento di posto i, j della matrice $\mathbf{AC}_\mathbf{X}\mathbf{A}^T$. ■

Esempio* 4.8.11 (Continuazione dell'Esempio 4.8.8) Sia (X_1, \dots, X_n) il vettore introdotto nell'Esempio 4.8.8. Allora, la matrice di covarianza di (X_1, \dots, X_n) è

$$\mathbf{C} = \frac{br}{(b+r)^2} \cdot \begin{pmatrix} 1 & -\frac{1}{b+r-1} & -\frac{1}{b+r-1} & \cdots & -\frac{1}{b+r-1} \\ -\frac{1}{b+r-1} & 1 & -\frac{1}{b+r-1} & \cdots & -\frac{1}{b+r-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -\frac{1}{b+r-1} & -\frac{1}{b+r-1} & -\frac{1}{b+r-1} & \cdots & 1 \end{pmatrix}$$

La matrice di covarianza sarà particolarmente utile nella Sezione 4.10 dedicata ai vettori gaussiani.

4.9 *Funzione generatrice dei momenti

La nozione di funzione generatrice dei momenti che abbiamo visto nel caso di variabili aleatorie può essere data anche per vettori aleatori n -dimensionali.

³Una matrice \mathbf{B} $n \times n$ è semidefinita positiva se $\forall \mathbf{x} \in \mathbb{R}^n$ non identicamente nullo $\mathbf{x}^T \mathbf{B} \mathbf{x} \geq 0$.

Definizione 4.9.1 Sia $\mathbf{X} = (X_1, \dots, X_n)$ un vettore aleatorio per il quale esiste un “rettangolo” aperto di \mathbb{R}^n $J = J_1 \times \dots \times J_n$ contenente $\mathbf{0} = (0, \dots, 0)$ tale che $e^{t_1 X_1 + \dots + t_n X_n}$ ammette media per ogni $\mathbf{t} = (t_1, \dots, t_n)$ in J . Allora la funzione

$$m_{\mathbf{X}}(\mathbf{t}) := E(e^{t_1 X_1 + \dots + t_n X_n})$$

definita (almeno) per ogni $\mathbf{t} \in I$ è detta funzione generatrice dei momenti di \mathbf{X} .

Da $m_{\mathbf{X}}$ si possono ottenere le funzioni generatrici marginali di X_1, \dots, X_n , m_{X_1}, \dots, m_{X_n} . Infatti, $m_{\mathbf{X}}(t_1, 0, \dots, 0) = E(e^{t_1 X_1}) = m_{X_1}(t_1)$ e, analogamente, $m_{\mathbf{X}}(0, \dots, 0, t_i, 0, \dots, 0) = E(e^{t_i X_i}) = m_{X_i}(t_i)$.

Non enunceremo qui altre proprietà delle funzioni generatrici dei momenti di vettori aleatori. Ricordiamo solamente due fondamentali risultati: il primo stabilisce una corrispondenza biunivoca fra funzioni di ripartizione e funzioni generatrici dei momenti, il secondo caratterizza la nozione di indipendenza tra variabili aleatorie mediante la funzione generatrice dei momenti.

Proposizione 4.9.2 Siano \mathbf{X}, \mathbf{Y} due vettori aleatori che ammettono funzione generatrice dei momenti $m_{\mathbf{X}}, m_{\mathbf{Y}}$ rispettivamente, e siano $F_{\mathbf{X}}$ la funzione di ripartizione di \mathbf{X} e $F_{\mathbf{Y}}$ quella di \mathbf{Y} . Allora $F_{\mathbf{X}} = F_{\mathbf{Y}}$ se e solo se $m_{\mathbf{X}} = m_{\mathbf{Y}}$.

Proposizione 4.9.3 Sia $\mathbf{X} = (X_1, \dots, X_n)$ un vettore aleatorio che ammette funzione generatrice dei momenti $m_{\mathbf{X}}$ e siano m_{X_i} le funzioni generatrici dei momenti marginali. Allora le componenti di \mathbf{X} sono indipendenti se e solo se $m_{\mathbf{X}} = m_{X_1} \dots m_{X_n}$.

Esercizio 4.9.4 Siano X, Y due variabili aleatorie indipendenti che hanno funzione generatrice dei momenti m_X, m_Y , rispettivamente. Dimostrate che la somma $X + Y$ ammette anche essa funzione generatrice dei momenti ed è data da $m_{X+Y}(s) = m_X(s)m_Y(s)$.

4.10 Vettori gaussiani

Le variabili aleatorie *gaussiane* o *normali* costituiscono probabilmente la più importante famiglia di variabili aleatorie che abbiamo incontrato nel corso. La loro importanza risiede nel fatto che, come vedremo nella Sezione 4.11.2 sul *Teorema del limite centrale*, la densità normale è in un certo senso una “densità naturale universale” e può essere osservata in vari campi delle scienze naturali.

In questa sezione, estendiamo la nozione di variabili aleatorie gaussiane al caso dei vettori aleatori.

Analogamente al caso unidimensionale, iniziamo introducendo la nozione di vettore normale standard o *gaussiano standard multivariato*. In quanto segue, i vettori saranno *vettori colonna* e “ T ” indicherà l’operazione di trasposizione di matrici.

Definizione 4.10.1 *Il vettore aleatorio $\mathbf{Z} = (Z_1, Z_2, \dots, Z_n)^T$ è gaussiano standard n -dimensionale, o n -variato, se le variabili aleatorie Z_1, Z_2, \dots, Z_n sono variabili aleatorie gaussiane standard indipendenti.*

Un vettore aleatorio $\mathbf{Z} = (Z_1, Z_2, \dots, Z_n)^T$ gaussiano standard n -dimensionale è quindi assolutamente continuo ed ha densità

$$f_{\mathbf{Z}}(z_1, z_2, \dots, z_n) = \frac{1}{(2\pi)^{\frac{n}{2}}} e^{-\frac{1}{2} \sum_{k=1}^n z_k^2}$$

Infatti segue dall’indipendenza delle Z_1, \dots, Z_n che

$$f_{\mathbf{Z}}(z_1, z_2, \dots, z_n) = f_{Z_1}(z_1) \cdots f_{Z_n}(z_n) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z_1^2}{2}} \cdots \frac{1}{\sqrt{2\pi}} e^{-\frac{z_n^2}{2}} = \frac{1}{(2\pi)^{\frac{n}{2}}} e^{-\frac{1}{2} \sum_{k=1}^n z_k^2}$$

Nota 4.10.2 Otteniamo facilmente il vettore delle medie e la matrice di covarianza di un vettore gaussiano standard n -dimensionale \mathbf{Z} , osservando che, per definizione, le n componenti di \mathbf{Z} sono indipendenti e gaussiane standard. Quindi il vettore delle medie di \mathbf{Z} è il vettore nullo e la matrice di covarianza di \mathbf{Z} è la matrice identità \mathbf{I} di dimensione n .

Definiamo un vettore aleatorio gaussiano \mathbf{X} n -dimensionale come funzione lineare di un vettore gaussiano standard \mathbf{Z} .

Definizione 4.10.3 *Un vettore aleatorio n -dimensionale \mathbf{X} è gaussiano (o gaussiano n -dimensionale o normale) se esistono una matrice \mathbf{A} $n \times m$, $\boldsymbol{\mu} \in \mathbb{R}^n$ e un vettore gaussiano standard m -dimensionale \mathbf{Z} , tali che $\mathbf{X} = \mathbf{AZ} + \boldsymbol{\mu}$.*

Calcoliamo il vettore delle medie e la matrice di covarianze di \mathbf{X} . Se $\mathbf{X} = \mathbf{AZ} + \boldsymbol{\mu}$ con $\mathbf{Z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$, segue dalla linearità della media che $E(\mathbf{X}) = \mathbf{A}E(\mathbf{Z}) + \boldsymbol{\mu} = \mathbf{0} + \boldsymbol{\mu} = \boldsymbol{\mu}$. Invece, in virtù del punto 2. della Proposizione 4.8.10, la covarianza di \mathbf{X} è data da $\mathbf{AIA}^T = \mathbf{AA}^T$. Notate che \mathbf{AA}^T è simmetrica e semidefinita positiva, come deve essere ogni matrice di covarianza.

Nota 4.10.4 La Definizione 4.10.3 è estremamente concisa in quanto abbiamo utilizzato il linguaggio delle matrici. Poiché, alle volte, questa semplicità formale può nascondere il significato, riscriviamo quanto detto nella Definizione 4.10.3 utilizzando il linguaggio delle coordinate: (X_1, X_2, \dots, X_n) è gaussiano se esistono delle costanti $a_{hk}, b_h \in \mathbb{R}$, per $h = 1, 2, \dots, n$ e $k = 1, 2, \dots, m$ tali che

$$\begin{aligned} X_1 &= a_{11}Z_1 + a_{12}Z_2 + \dots + a_{1m}Z_m + \mu_1 \\ X_2 &= a_{21}Z_1 + a_{22}Z_2 + \dots + a_{2m}Z_m + \mu_2 \\ &\vdots \\ X_n &= a_{n1}Z_1 + a_{n2}Z_2 + \dots + a_{nm}Z_m + \mu_n \end{aligned}$$

dove Z_1, Z_2, \dots, Z_m sono variabili aleatorie gaussiane standard indipendenti.

Nota 4.10.5 Osserviamo che nella definizione di vettore aleatorio gaussiano n -dimensionale nessuna restrizione è posta nella scelta della matrice \mathbf{A} e del vettore $\boldsymbol{\mu}$. Per esempio la matrice \mathbf{A} potrebbe avere prima riga nulla, cioè $a_{1k} = 0 \forall k$, e seconda riga con componenti tutte diverse da zero. Se questo è il caso, allora la prima componente X_1 è una costante cioè $X_1 = \mu_1$, mentre $X_2 = a_{21}Z_1 + a_{22}Z_2 + \dots + a_{2m}Z_m + \mu_2$ è una variabile aleatoria assolutamente continua e gaussiana in quanto somma di variabili aleatorie gaussiane indipendenti (cfr. Esempio 4.5.5).

è quindi chiaro che non sempre un vettore gaussiano n -dimensionale ha densità di probabilità $f_{\mathbf{X}}$ su \mathbb{R}^n . Altrimenti tutte le sue componenti sarebbero necessariamente assolutamente continue come stabilito nella Proposizione 4.4.3.

Tuttavia, in alcuni casi, un vettore gaussiano n -dimensionale \mathbf{X} ha densità in \mathbb{R}^n .

Consideriamo, per esempio, il caso di una matrice \mathbf{A} quadrata, $n \times n$ invertibile, \mathbf{Z} gaussiano standard n -dimensionale e $\mathbf{X} = \mathbf{AZ} + \boldsymbol{\mu}$. Se \mathbf{A} è invertibile, il vettore aleatorio gaussiano \mathbf{X} è una trasformazione affine di \mathbb{R}^n in sé e segue dall'Esempio 4.5.11 che \mathbf{X} è assolutamente continuo con densità

$$\begin{aligned} f_{\mathbf{X}}(\mathbf{x}) &= f_{\mathbf{Z}}(\mathbf{A}^{-1}\mathbf{x} - \mathbf{A}^{-1}\boldsymbol{\mu}) \frac{1}{|\det(\mathbf{A})|} = \frac{1}{(2\pi)^{\frac{n}{2}} |\det(\mathbf{A})|} e^{-\frac{1}{2}[\mathbf{A}^{-1}(\mathbf{x}-\boldsymbol{\mu})]^T [\mathbf{A}^{-1}(\mathbf{x}-\boldsymbol{\mu})]} = \\ &= \frac{1}{(2\pi)^{\frac{n}{2}} |\det(\mathbf{A})|} e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T (\mathbf{A}^T)^{-1} \mathbf{A}^{-1}(\mathbf{x}-\boldsymbol{\mu})} = \frac{1}{(2\pi)^{\frac{n}{2}} |\det(\mathbf{A})|} e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T (\mathbf{AA}^T)^{-1}(\mathbf{x}-\boldsymbol{\mu})} \end{aligned}$$

Osservando che $\det(\mathbf{AA}^T) = \det(\mathbf{A})^2$, otteniamo che la densità di $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{AA}^T)$ è

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^n \det(\mathbf{AA}^T)}} e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T (\mathbf{AA}^T)^{-1}(\mathbf{x}-\boldsymbol{\mu})} \quad (4.10.1)$$

Notate che la densità (4.10.1) dipende soltanto dal vettore delle medie $\boldsymbol{\mu}$ e dalla matrice di covarianza $\mathbf{C} := \mathbf{AA}^T$ che in questo caso è simmetrica e definita positiva⁴.

⁴Cioè tale che $\forall \mathbf{x} \in \mathbb{R}^n$ non identicamente nullo $\mathbf{x}^T \mathbf{C} \mathbf{x} > 0$.

In realtà questo non è l'unico caso in cui c'è una densità su \mathbb{R}^n . Infatti si può dimostrare il seguente risultato per \mathbf{A} non necessariamente quadrata:

Proposizione 4.10.6 *Un vettore gaussiano $\mathbf{X} = \mathbf{AZ} + \boldsymbol{\mu}$ ha densità su \mathbb{R}^n se e solo se la matrice di covarianza $\mathbf{C} = \mathbf{AA}^T$ è non singolare. In questo caso la densità è data da*

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^n \det(\mathbf{C})}} e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T \mathbf{C}^{-1}(\mathbf{x}-\boldsymbol{\mu})} \quad (4.10.2)$$

Nota 4.10.7 Si può dire di più: supponiamo che \mathbf{X} sia un vettore aleatorio assolutamente continuo che ha densità data in (4.10.2) con \mathbf{C} matrice simmetrica e definita positiva. Allora è possibile estrarre la radice di \mathbf{C} , cioè esiste una matrice invertibile \mathbf{A} tale che $\mathbf{C} = \mathbf{AA}^T$. Sia ora $\mathbf{Z} = \mathbf{A}^{-1}(\mathbf{X} - \boldsymbol{\mu})$. Per calcolare la densità di \mathbf{Z} usiamo ancora la formula nell'equazione (4.5.6) da cui otteniamo

$$\begin{aligned} f_{\mathbf{Z}}(\mathbf{z}) &= f_{\mathbf{X}}(\mathbf{Az} + \boldsymbol{\mu}) |\det(\mathbf{A})| = \frac{|\det(\mathbf{A})|}{\sqrt{(2\pi)^n \det(\mathbf{AA}^T)}} e^{-\frac{1}{2}(\mathbf{Az} + \boldsymbol{\mu} - \boldsymbol{\mu})^T (\mathbf{AA}^T)^{-1} (\mathbf{Az} + \boldsymbol{\mu} - \boldsymbol{\mu})} \\ &= \frac{1}{\sqrt{(2\pi)^n}} e^{-\frac{1}{2}\mathbf{z}^T \mathbf{z}} \end{aligned}$$

cioè \mathbf{Z} è gaussiano standard; inoltre chiaramente $\mathbf{X} = \mathbf{AZ} + \boldsymbol{\mu}$. Questo ci dice che se abbiamo un vettore aleatorio n -dimensionale assolutamente continuo con densità (4.10.2), dove $\boldsymbol{\mu} \in \mathbb{R}^n$ e \mathbf{C} è una matrice simmetrica e definita positiva, allora \mathbf{X} è un vettore gaussiano di media $\boldsymbol{\mu}$ e matrice di covarianza \mathbf{C} .

Nel prossimo esempio, sviluppiamo la densità gaussiana bivariata per esteso.

Esempio 4.10.8 (Densità gaussiana bivariata) Sia $\mathbf{X} = (X_1, X_2)$ gaussiano con matrice di covarianza $\mathbf{C} = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix}$ con $\sigma_1^2 \sigma_2^2 > 0$ e vettore delle medie $\boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}$. La matrice \mathbf{C} è invertibile se e solo se $\det(\mathbf{C}) > 0$, cioè

$$\det(\mathbf{C}) = \sigma_1^2 \sigma_2^2 - \sigma_{12}^2 = \sigma_1^2 \sigma_2^2 \left(1 - \frac{\sigma_{12}^2}{\sigma_1^2 \sigma_2^2}\right) = \sigma_1^2 \sigma_2^2 (1 - \rho_{12}^2) > 0$$

dove ρ_{12} è il coefficiente di correlazione tra X_1 e X_2 (quindi $\rho_{12}^2 \neq 1$); inoltre,

$$\mathbf{C}^{-1} = \frac{1}{\sigma_1^2 \sigma_2^2 (1 - \rho_{12}^2)} \begin{pmatrix} \sigma_2^2 & -\sigma_{12} \\ -\sigma_{12} & \sigma_1^2 \end{pmatrix}$$

e

$$\begin{aligned} (\mathbf{x} - \boldsymbol{\mu})^T \mathbf{C}^{-1} (\mathbf{x} - \boldsymbol{\mu}) &= \\ &= \frac{1}{\sigma_1^2 \sigma_2^2 (1 - \rho_{12}^2)} (x_1 - \mu_1, x_2 - \mu_2) \begin{pmatrix} \sigma_2^2 & -\sigma_{12} \\ -\sigma_{12} & \sigma_1^2 \end{pmatrix} \begin{pmatrix} x_1 - \mu_1 \\ x_2 - \mu_2 \end{pmatrix} = \\ &= \frac{1}{1 - \rho_{12}^2} \left[\left(\frac{x_1 - \mu_1}{\sigma_1} \right)^2 - 2\rho_{12} \left(\frac{x_1 - \mu_1}{\sigma_1} \right) \left(\frac{x_2 - \mu_2}{\sigma_2} \right) + \left(\frac{x_2 - \mu_2}{\sigma_2} \right)^2 \right] \end{aligned}$$

Segue che la densità gaussiana bivariata è

$$f_{X_1 X_2}(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho_{12}^2}} e^{-\frac{1}{2(1-\rho_{12}^2)} \left[\left(\frac{x_1-\mu_1}{\sigma_1} \right)^2 - 2\rho_{12} \left(\frac{x_1-\mu_1}{\sigma_1} \right) \left(\frac{x_2-\mu_2}{\sigma_2} \right) + \left(\frac{x_2-\mu_2}{\sigma_2} \right)^2 \right]}$$

Concludiamo la sezione fornendo alcune delle principali proprietà dei vettori aleatori gaussiani.

Proposizione 4.10.9 *Siano $\mathbf{X} = \mathbf{AZ} + \boldsymbol{\mu}$ un vettore gaussiano n -dimensionale e $\mathbf{C} = \mathbf{AA}^T$ la matrice di covarianza di \mathbf{X} . Allora valgono le seguenti proprietà.*

1. *Se $c_{ii} > 0$ allora la componente i -esima X_i è gaussiana con $X_i \sim \mathcal{N}(\mu_i, c_{ii})$. Se invece $c_{ii} = 0$, allora $P(X_i = \mu_i) = 1$.*
2. *Se \mathbf{G} è una matrice $k \times n$, e $\mathbf{h} \in \mathbb{R}^k$ allora $\mathbf{Y} := \mathbf{GX} + \mathbf{h}$ è vettore gaussiano con vettore delle medie $\mathbf{G}\boldsymbol{\mu} + \mathbf{h}$ e matrice di covarianza \mathbf{GCG}^T .*
3. *Se X_1, \dots, X_n sono scorrelate allora sono anche indipendenti.*

Dimostrazione

1. Per quanto discusso nella Nota 4.10.4, ogni X_i si può esprimere come combinazione lineare di variabili aleatorie gaussiane indipendenti più una costante. Segue da quanto svolto nell'Esempio 4.5.5 che anche X_i è gaussiana o costante.

2. $\mathbf{GX} + \mathbf{h} = \mathbf{G}(\mathbf{AZ} + \boldsymbol{\mu}) + \mathbf{h} = (\mathbf{GA})\mathbf{Z} + (\mathbf{G}\boldsymbol{\mu} + \mathbf{h})$. (Notate che possiamo ottenere il risultato 1. anche da 2. per particolari scelte di \mathbf{G} e \mathbf{h}).

3. Dimostriamo questo punto nel caso in cui la matrice di covarianza sia invertibile e quindi \mathbf{X} abbia densità su \mathbb{R}^n . Se X_1, \dots, X_n sono scorrelate la matrice di covarianza \mathbf{C} di \mathbf{X} è una matrice diagonale e la diagonale è costituita dalle varianze $\sigma_1^2, \dots, \sigma_n^2$ di X_1, \dots, X_n . Allora la densità di \mathbf{X} è

$$\begin{aligned} f_{\mathbf{X}}(\mathbf{x}) &= \frac{1}{\sqrt{(2\pi)^n \det(\mathbf{C})}} e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T \mathbf{C}^{-1}(\mathbf{x}-\boldsymbol{\mu})} \\ &= \frac{1}{\sqrt{(2\pi)^n \sigma_1^2 \cdots \sigma_n^2}} e^{-\frac{1}{2} \sum_{i=1}^n \left(\frac{x_i - \mu_i}{\sigma_i} \right)^2} \\ &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma_i^2}} e^{-\frac{1}{2} \left(\frac{x_i - \mu_i}{\sigma_i} \right)^2} \\ &= \prod_{i=1}^n f_{X_i}(x_i) \end{aligned}$$

e quindi le X_i sono indipendenti. ■

Esercizio 4.10.10 Sia $\mathbf{X} = (X_1, \dots, X_n)^T$ un vettore gaussiano con vettore delle medie $\boldsymbol{\mu}$ e matrice di covarianza \mathbf{C} . Mostrare che, per ogni scelta di a_1, \dots, a_n numeri reali di cui almeno uno diverso da 0, $a_1X_1 + \dots + a_nX_n$ è una variabile aleatoria assolutamente continua gaussiana e determinarne i parametri.

Esercizio 4.10.11 Sia $\mathbf{X} = (X_1, \dots, X_n)^T$ un vettore gaussiano con vettore delle medie $\boldsymbol{\mu}$ e matrice di covarianza \mathbf{C} . Usando la proprietà 2. della Proposizione 4.10.9, mostrare che ogni vettore aleatorio (X_i, X_j) ($i \neq j$) estratto da \mathbf{X} è un vettore gaussiano bidimensionale e determinarne i parametri.

Esercizio 4.10.12 Sia $\mathbf{X} = (X_1, \dots, X_n)^T$ un vettore gaussiano con vettore delle medie $\boldsymbol{\mu}$ e matrice di covarianza \mathbf{C} . Mostrare che se X_i, X_j sono scorrelate, allora sono anche indipendenti.

4.11 Teoremi limite per somme di variabili aleatorie

4.11.1 Legge dei grandi numeri

Lanciamo un numero elevato n di volte una moneta (cioè consideriamo un esperimento ripetibile infinite volte) e consideriamo la frequenza relativa di testa negli n lanci:

$$\frac{X_1 + \dots + X_n}{n}$$

dove X_i vale 1 se il risultato della i -esima prova è testa, 0 altrimenti. Se la moneta non è truccata ci aspettiamo che, salvo in casi eccezionali, questa frequenza sia sempre più vicina ad $1/2$, al crescere di n . Tale risultato è confermato dalla “Legge dei grandi numeri”. Si parla di Legge debole dei grandi numeri e di Legge forte dei grandi numeri. La prima è una conseguenza immediata della disuguaglianza di Chebychev:

Proposizione 4.11.1 (Legge debole dei grandi numeri) *Sia X_1, X_2, \dots una successione di variabili aleatorie indipendenti ed identicamente distribuite (i.i.d.) con media μ e varianza σ^2 finite. Sia $S_n = X_1 + \dots + X_n$. Allora, per ogni $\epsilon > 0$*

$$\lim_{n \rightarrow \infty} P\left(\left|\frac{S_n}{n} - \mu\right| > \epsilon\right) = 0$$

Dimostrazione Poiché le X_i sono i.i.d. allora

$$\text{Var}(S_n) = n \text{Var}(X_1) = n\sigma^2$$

da cui

$$\text{Var}\left(\frac{S_n}{n}\right) = \frac{1}{n^2}n\sigma^2 = \frac{\sigma^2}{n}$$

e

$$E\left(\frac{S_n}{n}\right) = \mu.$$

Segue dalla disuguaglianza di Chebychev che per ogni $\epsilon > 0$

$$P\left(\left|\frac{S_n}{n} - \mu\right| > \epsilon\right) \leq \frac{\sigma^2}{n\epsilon^2} \rightarrow 0 \quad (n \rightarrow +\infty) \quad \blacksquare$$

Date n variabili aleatorie X_1, \dots, X_n si chiama *media campionaria* di X_1, \dots, X_n la quantità $(X_1 + \dots + X_n)/n$ e la si indica con \bar{X}_n . Equivalentemente, la Legge debole dei grandi numeri afferma che $P(|\bar{X}_n - \mu| \leq \epsilon) \rightarrow 1$ per $n \rightarrow +\infty$; quindi, essa mette in evidenza che, pur partendo da un esperimento aleatorio costituito da prove ripetute del quale poco si può predire ad ogni prova (le prove sono indipendenti), facendo le medie di tali prove si ottiene un esperimento il cui risultato può essere predetto con un elevato grado di certezza. In realtà vale il seguente risultato “più forte” la cui dimostrazione è più laboriosa.

Proposizione 4.11.2 *Sia X_1, X_2, \dots una successione di variabili aleatorie i.i.d. con media finita μ . Allora*

$$P(\{\omega : \lim_{n \rightarrow +\infty} \frac{S_n(\omega)}{n} = \mu\}) = 1$$

In pratica la legge forte applicata all'esempio dei lanci di una moneta dice che per “quasi tutte” le successioni di risultati X_1, X_2, \dots la frequenza relativa di testa S_n/n converge al trucco p della moneta.

Esempio 4.11.3 (Metodo di integrazione Monte Carlo) Sia h una funzione continua su $[0, 1]$. Vogliamo calcolare in modo approssimato $\int_0^1 h(x) dx$. Esistono molte formule di quadratura, ma la tecnica Monte Carlo è una delle più semplici. Inoltre, anche se può non risultare il miglior metodo per funzioni su $[0, 1]$, si estende facilmente e diventa competitiva nel caso di integrali multidimensionali. Infatti, nei metodi numerici “tradizionali”, l'errore di approssimazione dipende dalla dimensione, mentre ciò non accade nel caso del metodo Monte Carlo. I generatori di numeri casuali in ogni libreria di sistema producono valori le cui proprietà si avvicinano alle realizzazioni di variabili aleatorie i.i.d. con densità $\mathcal{U}(0, 1)$ e rendono implementabile il metodo Monte Carlo basato sul seguente corollario alla Legge forte dei grandi numeri:

Corollario 4.11.4 *Sia h una funzione su $[0, 1]$ con $\int_0^1 |h(x)| dx < +\infty$. Siano U_1, U_2, \dots variabili aleatorie i.i.d. con densità uniforme su $[0, 1]$. Allora*

$$P\left(I_{1n} := \frac{1}{n} \sum_{j=1}^n h(U_j) \rightarrow \int_0^1 h(x) dx, n \rightarrow +\infty\right) = 1$$

Dimostrazione È sufficiente osservare che le variabili aleatorie $h(U_1), h(U_2), \dots$ sono i.i.d. con media finita $\int_0^1 h(x) dx$ ed applicare la Legge forte dei grandi numeri. ■

Il metodo Monte Carlo consiste nell'approssimare $\int_0^1 h(x) dx$ con I_{1n} per n “grande”. Per ogni n fissato, la bontà dell'approssimazione può essere valutata tramite

$$\text{Var}(I_{1n}) = \text{Var}\left(\frac{1}{n} \sum_{j=1}^n h(U_j)\right) = \frac{\int_0^1 h^2(x) dx - (\int_0^1 h(x) dx)^2}{n}$$

Al fine di ridurre la varianza, il metodo delle “variabili antitetiche” approssima il valore dell'integrale mediante

$$I_{2n} := \frac{1}{2n} \sum_{i=1}^n (h(U_i) + h(1 - U_i))$$

Esercizio 4.11.5 1. Mostrare che

$$P\left(\lim_{n \rightarrow +\infty} I_{2n} = \int_0^1 h(x) dx\right) = 1$$

2. Calcolare $\text{Var}(I_{2n})$.
3. Dedurre che $\text{Var}(I_{2n}) \leq \text{Var}(I_{1n})$.

4.11.2 Teorema centrale del limite

Consideriamo n variabili aleatorie X_1, \dots, X_n i.i.d. con media μ e varianza σ^2 , entrambe finite. Abbiamo visto nella precedente sezione che per n “grande”, la media campionaria \bar{X}_n approssima in un opportuno senso la media μ :

$$\bar{X}_n \simeq \mu.$$

Se inoltre X_1, \dots, X_n sono gaussiane, allora è immediato verificare che $\bar{X}_n \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$, e quindi siamo in grado di valutare probabilisticamente la “dispersione” dei valori assunti da \bar{X}_n intorno a μ : ad esempio, osservando che

$$\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \sim \mathcal{N}(0, 1)$$

otteniamo

$$P(|\bar{X}_n - \mu| \leq \delta) = \Phi\left(\frac{\sqrt{n}}{\sigma}\delta\right) - \Phi\left(-\frac{\sqrt{n}}{\sigma}\delta\right) = 2\Phi\left(\frac{\sqrt{n}}{\sigma}\delta\right) - 1$$

che si calcola usando le tavole della ripartizione gaussiana standard.

In questa sezione presenteremo una versione semplice del *Teorema centrale del limite* (o *Teorema del limite centrale*) il cui significato euristico è il seguente: la media campionaria di un numero n , sufficientemente grande, di variabili aleatorie i.i.d., di media μ e varianza σ^2 finite ha una funzione di ripartizione che è approssimativamente gaussiana di media μ e varianza σ^2/n .

Teorema 4.11.6 Sia X_1, X_2, \dots una successione di variabili aleatorie i.i.d. con media μ e varianza σ^2 , con $0 < \sigma^2 < +\infty$. Allora per ogni $x \in \mathbb{R}$:

$$\lim_{n \rightarrow +\infty} P\left(\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \leq x\right) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} du = \Phi(x) \quad (4.11.1)$$

Il teorema può essere interpretato nel modo seguente: pur di prendere un numero elevato di variabili nella successione, la funzione di ripartizione di $\sqrt{n}(\bar{X}_n - \mu)/\sigma$, cioè della standardizzata della media campionaria \bar{X}_n , è approssimabile con quella gaussiana standard. Quindi, per quanto visto sulle standardizzate di variabili aleatorie gaussiane, approssimativamente \bar{X}_n ha funzione di ripartizione gaussiana di media μ e varianza σ^2/n . La bontà dell'approssimazione dipende dal numero di variabili aleatorie sommate e dalla forma della funzione di ripartizione delle variabili aleatorie di cui si fa la media.

Equivalentemente, l'enunciato del teorema centrale del limite può essere dato in termini di somme di variabili aleatorie i.i.d.. Infatti

$$\sqrt{n} \frac{\bar{X}_n - \mu}{\sigma} = \frac{\sqrt{n}(\sqrt{n}(\bar{X}_n - \mu))}{\sqrt{n}\sigma} = \frac{S_n - n\mu}{\sqrt{n}\sigma}$$

cioè la standardizzata di \bar{X}_n coincide con quella di S_n . Quindi, sotto le ipotesi del teorema centrale del limite:

$$\lim_{n \rightarrow +\infty} P\left(\frac{S_n - n\mu}{\sqrt{n}\sigma} \leq x\right) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} du = \Phi(x)$$

Poiché diverse variabili aleatorie di uso comune si possono rappresentare come somma di numerose variabili i.i.d., allora il teorema centrale del limite può essere usato per approssimare le vere funzioni di ripartizione di queste variabili. Ad esempio, gli errori di misura si possono rappresentare come somma di un numero elevato di singoli termini (errori elementari), ciascuno dei quali è dovuto ad una causa, non dipendente dalle altre. Quali che siano le funzioni di ripartizione degli errori elementari, le peculiarità di queste non si manifestano nella somma di un gran numero di termini e la funzione di ripartizione della somma è vicina alla funzione di ripartizione gaussiana.

Seguono alcuni esempi di applicazione del teorema centrale del limite.

Esempio 4.11.7 Nel Capitolo 3 abbiamo discusso la possibilità di approssimare la funzione di ripartizione binomiale con quella gaussiana, sulla base del Teorema 3.6.1 di De Moivre Laplace. Effettivamente, il Teorema 3.6.1 di De Moivre Laplace è un caso particolare del teorema centrale del limite. In realtà esso rappresenta una prima versione del teorema centrale del limite. Infatti, una variabile aleatoria binomiale $\mathbf{Bi}(n, p)$ ha la stessa densità della somma di n variabili aleatorie i.i.d. di Bernoulli di parametro $p \in (0, 1)$.

Rimandiamo all'Esempio 3.6.4 per la discussione sulla bontà della approssimazione.

Invece, per quanto concerne la correzione di continuità, può essere utile ricordare qui come si apporta nel caso di una somma di variabili aleatorie indipendenti a valori interi (ma non necessariamente bernoulliane).

Se X_1, \dots, X_n sono variabili aleatorie i.i.d. discrete e a valori interi con comune media μ e comune varianza $\sigma^2 > 0$, $S_n = \sum_{j=1}^n X_j$ ed n è grande, la *correzione di continuità* si apporta nel seguente modo:

$$P(S_n \leq r) \simeq \Phi\left(\frac{r + 0.5 - n\mu}{\sqrt{n\sigma^2}}\right)$$

per ogni r intero.

Esempio 4.11.8 Sia X una variabile aleatoria di Poisson di parametro $\lambda = 100$. Calcolare un valore approssimato di $P(X < 110)$.

La variabile aleatoria $X \sim \mathcal{P}(100)$ ha la stessa densità della somma di 100 variabili aleatorie $Y_1, \dots, Y_{100} \stackrel{i.i.d.}{\sim} \mathcal{P}(1)$; queste variabili aleatorie sono discrete a valori interi e hanno media e varianza pari a 1. Quindi, per il teorema centrale del limite, la fdr $\mathcal{P}(100)$ si può approssimare con la fdr $\mathcal{N}(100, 100)$. Inoltre, l'approssimazione è migliore con la correzione di continuità. In particolare:

$$\begin{aligned} P(X < 110) &= P(X \leq 109) = P\left(\sum_{j=1}^{100} Y_j \leq 109\right) = P\left(\sum_{j=1}^{100} Y_j \leq 109.5\right) \\ &= P\left(\frac{\sum_{j=1}^{100} Y_j - 100}{10} \leq \frac{109.5 - 100}{10}\right) \simeq \Phi\left(\frac{109.5 - 100}{10}\right) \simeq 0.8289 \end{aligned}$$

Senza la correzione di continuità, un valore approssimato di $P(X < 110)$ è dato da $\Phi((109 - 100)/10) \simeq 0.8159$. (Il valore esatto di $P(X < 110)$ è 0.82944.)

Esempio 4.11.9 Siano U_1, \dots, U_{147} variabili aleatorie indipendenti e uniformi sull'intervallo $(0, 2)$ e $S = \sum_{j=1}^{147} U_j$. Calcolare un valore approssimato di $P(S < 161)$.

In quanto somma di variabili aleatorie i.i.d. assolutamente continue, anche S è assolutamente continua da cui $P(S < 161) = P(S \leq 161)$. Inoltre $E(S) = 147 \times E(U_1) = 147$ e $\text{Var}(S) = 147 \text{Var}(U_1) = 147/3 = 49$. Per il teorema centrale del limite, la funzione di ripartizione di $(S - E(S))/\sqrt{\text{Var } S}$ converge a Φ . Quindi, $P(S < 161) = F_S(161) \simeq \Phi((161 - 147)/\sqrt{49}) = \Phi(2) \simeq 0.9772$. (Qui non serve la correzione di continuità perché S è già continua...)

Appendice A

Richiami di analisi matematica

La presente appendice ha il solo scopo di richiamare alcune nozioni di teoria degli insiemi, algebra lineare e analisi. Per le dimostrazioni si rimanda a [5, Volumi 1 e 2].

A.1 Richiami di teoria degli insiemi

Dato un insieme Ω siano A , B e C sottoinsiemi di Ω ; \emptyset rappresenta l'insieme vuoto.

Definizione A.1.1 A^c : *L'insieme complementare di A (rispetto a Ω) è l'insieme di tutti gli elementi che sono in Ω ma non in A ;*

Vale che $(A^c)^c = A$: il complementare del complementare di A è A ;

$A \cup B$: *L'unione di A e B è l'insieme degli elementi che appartengono o ad A o a B o ad entrambi;*

$A \cap B$: *L'intersezione di A e B è l'insieme degli elementi che appartengono sia ad A che a B ;*

$A \setminus B = A \cap B^c$: *La differenza di B da A è l'insieme degli elementi di A che non appartengono a B ;*

$A \triangle B = (A \setminus B) \cup (B \setminus A) = (A \cup B) \setminus (A \cap B)$: *La differenza simmetrica di A e B è l'insieme costituito dagli elementi di A che non appartengono a B e da quelli di B che non appartengono ad A . Cioè l'insieme degli elementi che appartengono ad A o a B ma non ad entrambi.*

Le operazioni insiemistiche di unione, intersezione e complemento godono delle proprietà elencate in Tabella A.1.

Proprietà	unione	intersezione
commutativa	$A \cup B = B \cup A$	$A \cap B = B \cap A$
associativa	$A \cup (B \cup C) = (A \cup B) \cup C$	$A \cap (B \cap C) = (A \cap B) \cap C$
distributiva	$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$	$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$
di inclusione	$A \subseteq B$ se e solo se $A \cup B = B$	$A \subseteq B$ se e solo se $A \cap B = A$
	$A \cup \Omega = \Omega$	$A \cap \Omega = A$
	$A \cup \emptyset = A$	$A \cap \emptyset = \emptyset$
	$A \cup A = A$	$A \cap A = A$
	$A \cup A^c = \Omega$	$A \cap A^c = \emptyset$
Leggi di De Morgan	$(A \cup B)^c = A^c \cap B^c$	$(A \cap B)^c = A^c \cup B^c$

Tabella A.1: Alcune proprietà di unione, intersezione e complemento

A.2 Alcuni limiti notevoli

Il numero $e = \lim_{x \rightarrow +\infty} (1 + \frac{\lambda}{x})^x = e^\lambda \quad \forall \lambda \in \mathbb{R}$

A.3 Calcolo integrale

A.3.1 Proprietà dell'integrale

- 1. Linearità dell'operatore integrale** Siano f e g due funzioni definite su $[a, b]$ ed ivi integrabili. Allora $cf + g$ è integrabile su $[a, b]$ per ogni $c \in \mathbb{R}$ e $\int (cf(x) + g(x))dx = c \int f(x)dx + \int g(x)dx$.
- 2. Monotonia** Siano f e g due funzioni definite su $[a, b]$ ed ivi integrabili. Se $f(x) \leq g(x)$ per ogni $x \in [a, b]$, allora $\int f(x)dx \leq \int g(x)dx$.
- 3.** Se $f(x)$ è una funzione pari ($f(x) = f(-x) \quad \forall x \geq 0$) integrabile su $[-a, a]$, allora $\int_{-a}^a f(x)dx = 2 \int_0^a f(x)dx$.
- 5.** Se $f(x)$ è una funzione dispari ($f(x) = -f(-x) \quad \forall x \geq 0$) integrabile su $[-a, a]$, allora $\int_{-a}^a f(x)dx = 0$;
- 6.** Se $a = b$ allora $\int_a^b f(x)dx = 0$.

A.3.2 Regole di integrazione

Integrazione per parti

$$\int_a^b f(x)g'(x)dx = f(b)g(b) - f(a)g(a) - \int_a^b f'(x)g(x)dx$$

f è detto *fattore finito* e $g'(x)dx$ *fattore differenziale*. Per brevità spesso si usa $f(x)g(x)]_a^b = f(b)g(b) - f(a)g(a)$

Integrazione per sostituzione Se $f(x)$ è una funzione continua su $[a, b]$ e $\varphi(x)$ è una funzione continua, derivabile con continuità e invertibile, allora

$$\int_c^d f(x)dx = \int_{\varphi^{-1}(c)}^{\varphi^{-1}(d)} f(\varphi(x))\varphi'(x)dx$$

per ogni $a \leq c < d \leq b$

A.3.3 Alcuni integrali immediati

Utilizzando il metodo di integrazione per parti o per sostituzione si verifichi che:

$$\int_a^b dx = b - a \quad \forall -\infty < a < b < +\infty \quad (\text{A.3.1})$$

$$\int_a^b e^{-\lambda x} dx = \frac{1}{\lambda}(e^{-\lambda a} - e^{-\lambda b}) \quad \forall a < b < +\infty \text{ e } \lambda \neq 0 \quad (\text{A.3.2})$$

in particolare, se $\lambda > 0$, allora $\int_0^{+\infty} e^{-\lambda x} dx = \frac{1}{\lambda}$

$$\int_a^b x e^{-\lambda x} dx = \frac{e^{-\lambda a}}{\lambda} \left(a + \frac{1}{\lambda}\right) - \frac{e^{-\lambda b}}{\lambda} \left(b + \frac{1}{\lambda}\right) \quad \forall \lambda > 0 \quad (\text{A.3.3})$$

in particolare $\int_0^{+\infty} x e^{-\lambda x} dx = \frac{1}{\lambda^2}$

$$\int_0^{+\infty} x^2 e^{-\lambda x} dx = \frac{2}{\lambda^3} \quad \forall \lambda > 0 \quad (\text{A.3.4})$$

$$\int_a^b \frac{1}{(1+x^2)} dx = \arctan(b) - \arctan(a) \quad (\text{A.3.5})$$

in particolare $\int_{-\infty}^{+\infty} \frac{1}{(1+x^2)} dx = \pi$

$$\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} x e^{-\frac{x^2}{2}} dx = 0 \quad (\text{NB: la funzione integranda è dispari}) \quad (\text{A.3.6})$$

$$\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} x^2 e^{-\frac{x^2}{2}} dx = 2 \int_0^{+\infty} \frac{1}{\sqrt{2\pi}} x^2 e^{-\frac{x^2}{2}} dx = 1 \quad (\text{A.3.7})$$

Dimostriamo ora che

$$\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = 1 \quad (\text{A.3.8})$$

Si osservi che è equivalente verificare che

$$\left(\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx \right)^2 = 1$$

Procediamo nel seguente modo:

$$\left(\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx \right)^2 = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \frac{1}{2\pi} e^{-\frac{x^2+y^2}{2}} dx dy$$

e in coordinate polari ($x = \rho \cos(\theta)$, $y = \rho \sin(\theta)$)

$$\begin{aligned} &= \int_0^{+\infty} \int_0^{2\pi} \frac{1}{2\pi} e^{-\frac{\rho^2}{2}} \rho d\theta d\rho \\ &= \int_0^{+\infty} e^{-\frac{\rho^2}{2}} \rho \left(\int_0^{2\pi} \frac{1}{2\pi} d\theta \right) d\rho \\ &= -e^{-\frac{\rho^2}{2}} \Big|_0^{+\infty} = 1 \end{aligned}$$

A.4 Successioni e serie

Somma dei primi numeri naturali e dei loro quadrati

$$\sum_{j=1}^n j = \frac{n(n+1)}{2} \qquad \sum_{j=1}^n j^2 = \frac{n(n+1)(2n+1)}{6}$$

Serie telescopiche, Serie di Mengoli

Serie telescopica: $\sum_{k=0}^{\infty} (a_k - a_{k+1}) = \lim_{n \rightarrow +\infty} \sum_{k=0}^n (a_k - a_{k+1}) = \lim_{n \rightarrow +\infty} (a_0 - a_{n+1})$

Serie di Mengoli: $\sum_{k=1}^{\infty} \frac{1}{k(k+1)} = \sum_{k=0}^{\infty} \left(\frac{1}{k} - \frac{1}{k+1} \right) = \lim_{n \rightarrow +\infty} \left(1 - \frac{1}{n+1} \right) = 1$

Serie geometrica Si ha

$$\sum_{j=0}^n q^j = \frac{1 - q^{n+1}}{1 - q} \quad \forall q \neq 1$$

da cui derivano per la *serie geometrica di ragione q* i seguenti risultati:

$$\sum_{j=0}^{+\infty} q^j = \begin{cases} = \frac{1}{1-q} & \text{se } |q| < 1 \\ +\infty & \text{se } q \geq 1 \\ \text{indeterminata} & \text{altrimenti} \end{cases}$$

Serie esponenziale

$$\sum_{n=0}^{+\infty} \frac{x^n}{n!} = e^x \quad \forall x \in \mathbb{R}$$

Derivazione e serie

Teorema A.4.1 Consideriamo la serie di funzioni $\sum_{n=1}^{+\infty} f_n(x)$ e supponiamo che per ogni $n \geq 1$ la funzione f_n sia derivabile sull'intervallo aperto (a, b) , con derivata f'_n . Se $\sum_{n=1}^{+\infty} f_n(x)$ converge in (a, b) e la serie $\sum_{n=1}^{+\infty} f'_n(x)$ converge uniformemente su (a, b) , allora $\sum_{n=1}^{+\infty} f_n(x)$ è derivabile su (a, b) e la derivata della serie coincide con la serie delle derivate.

Esempio A.4.2 Calcoliamo il valore delle serie $\sum_{x=1}^{+\infty} x(1-p)^{x-1}$ per $p \in (0, 1)$

$$\begin{aligned} \sum_{x=1}^{\infty} x(1-p)^{x-1} &= \sum_{x=1}^{\infty} \frac{d}{dp} (-1)(1-p)^x = \frac{d}{dp} (-1) \sum_{x=1}^{\infty} (1-p)^x \\ &= -\frac{d}{dp} \left(\frac{1}{1-(1-p)} - 1 \right) = \frac{1}{p^2} \end{aligned}$$

Appendice B

Calcolo combinatorio

B.1 Introduzione

Il calcolo combinatorio è costituito da una serie di tecniche che consentono di *contare* il numero di elementi di un dato insieme senza enumerarli esplicitamente. L'importanza che le tecniche di calcolo combinatorio hanno per il calcolo delle probabilità risiede nel fatto che nel caso di spazi equiprobabili finiti, il problema di calcolare la probabilità di un dato evento viene ridotto al conteggio dei modi in cui si può verificare l'evento.

B.2 Disposizioni e permutazioni

Sia E un insieme finito o collezione di oggetti e $|E|$ la *cardinalità* di E . Supponiamo che $|E| > 0$.

Definizione B.2.1 *Si chiamano disposizioni senza ripetizione (o semplici) di E di classe (o ordine) $r \leq |E|$ le r -uple ordinate di elementi di E senza ripetizioni. In particolare, le disposizioni senza ripetizione di ordine $|E|$ sono dette permutazioni.*

Si chiamano disposizioni con ripetizione di E di classe r , con $r \in \mathbf{N}$, le stringhe di r elementi di E .

La definizione dice che le disposizioni semplici di classe $r \leq |E|$ sono

$$\{(a_1, \dots, a_r) : a_k \in E, a_h \neq a_k \forall h \neq k, h, k = 1, \dots, r\},$$

mentre le disposizioni con ripetizione di classe $r \in \mathbf{N}$ di E sono

$$\{(a_1, \dots, a_r) : a_k \in E, k = 1, \dots, r\}.$$

Esempio B.2.2 Sia $E = \{a, b, c\}$, allora $|E| = 3$ e:

1. le disposizioni senza ripetizione di classe 2 di E sono (a, b) , (a, c) , (b, a) , (b, c) , (c, a) , (c, b) ;

2. le disposizioni con ripetizione di classe 2 di E sono $(a, a), (a, b), (a, c), (b, a), (b, b), (b, c), (c, a), (c, b), (c, c)$;
3. le permutazioni (o disposizioni senza ripetizione di classe 3) di E sono $(a, b, c), (a, c, b), (b, a, c), (b, c, a), (c, a, b)$ e (c, b, a) .

Nell'esempio B.2.2 possiamo contare direttamente quante sono le disposizioni di un dato ordine semplicemente elencandole. Le cose invece si complicano se aumenta la cardinalità dell'insieme.

Esempio B.2.3 Elencare tutte le disposizioni con o senza ripetizione di ordine 4 e tutte le permutazioni di $E = \{a, b, c, d, e\}$.

Da qui la necessità di contare senza elencare. Per le disposizioni semplici vale la seguente proposizione:

Proposizione B.2.4 *Il numero $(n)_r$ di disposizioni senza ripetizione di ordine $r \leq n$ di un insieme di n elementi è dato da*

$$(n)_r = n(n-1) \cdots (n-r+1).$$

Dimostrazione Per elencare le disposizioni semplici, possiamo procedere nel seguente modo: la prima posizione della stringa può essere occupata da uno qualsiasi degli n elementi disponibili. Per ogni scelta della prima posizione, rimangono $n-1$ elementi diversi fra cui scegliere per la seconda (perché non posso scegliere lo stesso elemento). Mentre, per il terzo elemento abbiamo $n-2$ scelte per ognuna delle $n(n-1)^1$ scelte delle prime 2 posizioni e così via. Infine, fissata una fra le $n(n-1) \cdots (n-(r-2))$ possibili scelte per le prime $r-1$ posizioni, per l' r -esimo elemento abbiamo soltanto $n-r+1$ scelte. In totale otteniamo $n(n-1)(n-2) \cdots (n-r+1)$ possibili scelte. ■

Dalla proposizione precedente (prendendo $n=r$) discende direttamente:

Corollario B.2.5 *Il numero $P(n)$ di permutazioni di un insieme di n elementi è dato da:*

$$P(n) = n(n-1) \cdots 2 \cdot 1.$$

Risulta comoda la seguente notazione:

Definizione B.2.6 *Se $n \in \mathbf{N}$ indichiamo con il simbolo $n!$, (fattoriale) il numero:*

$$n! := n(n-1) \cdots 2 \cdot 1;$$

poniamo inoltre $0! := 1$.

¹È ovvio che se il primo elemento può essere scelto in n modi ed il secondo può essere scelto in $n-1$ modi per ciascuno dei modi con il quale scelgo il primo, ottengo $n(n-1)$ modi di scegliere primo e secondo elemento.

Con la precedente definizione otteniamo:

$$(n)_r = \frac{n!}{(n-r)!} \quad \text{e} \quad P(n) = n!$$

Per quanto riguarda le disposizioni con ripetizione la cosa è ancora più semplice:

Proposizione B.2.7 *Le disposizioni con ripetizione di ordine r di un insieme di n elementi son n^r .*

Esempio B.2.8 Dimostrare la Proposizione B.2.7.

B.3 Combinazioni

Definizione B.3.1 *Sia E un insieme finito. Ogni sottoinsieme di E di cardinalità $r \leq |E|$ è detto combinazione di classe r di E .*

La definizione afferma che le combinazioni di un insieme E sono

$$\{F: F \subset E\}$$

Esempio B.3.2 Se $E = \{a, b, c\}$, allora

1. le combinazioni di E di classe 2 sono $\{a, b\}$, $\{a, c\}$, $\{b, c\}$;
2. la combinazione di E di classe 3 è $\{a, b, c\} = E$.

Per contare il numero di combinazioni di classe r di n elementi, basta osservare che ogni fissata combinazione dà luogo a $r!$ disposizioni semplici di classe r . Quindi se $C(n, r)$ indica il numero di combinazioni di classe r di un insieme di n elementi, allora

$$(n)_r = r!C(n, r) \quad \text{da cui} \quad C(n, r) = \frac{n!}{r!(n-r)!} = \binom{n}{r}$$

Il simbolo $\binom{n}{r}$ è detto *coefficiente binomiale* e si legge *n sopra r* . Abbiamo dimostrato che:

Proposizione B.3.3 *Il numero di combinazioni di classe r di n elementi è*

$$C(n, r) = \binom{n}{r}$$

Esempio B.3.4

$$\binom{8}{2} = \frac{8!}{2!6!} = \frac{8 \cdot 7}{2 \cdot 1} = 28 \quad \binom{5}{0} = \frac{5!}{0!5!} = \frac{5!}{1 \cdot 5!} = 1.$$

Esempio B.3.5 In quanti modi si possono estrarre 10 carte da un mazzo di 40?

Dato un insieme E costituito dalle 40 carte, ogni presa di 10 carte corrisponde a un sottoinsieme di cardinalità 10, quindi il numero cercato è $\binom{40}{10} = 847660528$.

B.4 Esercizi

Esercizio B.4.1 Dimostrare che

$$\sum_{k=0}^n \binom{n}{k} = 2^n$$

senza usare la formula del binomio di Newton.

Esercizio B.4.2 Verificare che

$$\binom{n}{r} = \binom{n}{n-r}.$$

Esercizio B.4.3 In quanti modi 7 persone possono disporsi

- (a) su 7 sedie allineate?
- (b) Attorno ad un tavolo circolare?

Soluzione

- (a) Sono i modi di ordinare 7 oggetti (permutazioni), cioè $7!$.
- (b) Se consideriamo i posti intorno al tavolo numerati, allora si hanno $7!$ modi di sedersi. Se però consideriamo che la posizione relativa delle persone rispetto al tavolo è ininfluente, cioè consideriamo due configurazioni equivalenti se si ottengono mediante una rotazione “rigida” attorno al tavolo, si vede che il numero di configurazioni possibili diventano: $7!/7 = 6!$. ■

Esercizio B.4.4 Quante parole di lunghezza ≤ 10 si possono formare con un alfabeto binario.

Soluzione Con un alfabeto binario si possono formare 2 parole di lunghezza 1, $2 \times 2 = 2^2$ parole di lunghezza 2, \dots , 2^n parole di lunghezza n (cfr. Proposizione B.2.7). In definitiva ci sono

$$2 + 2^2 + \dots + 2^{10} = \frac{2^{11} - 1}{2 - 1} - 1 = 2(2^{10} - 1) = 2046.$$

parole di lunghezza minore od uguale a 10. ■

Esercizio B.4.5 Le tessere del domino sono marcate con 2 numeri. Le tessere sono simmetriche (cioè le coppie non sono ordinate). Quante sono le tessere che si ottengono utilizzando i numeri $1, \dots, n$?

Soluzione Le tessere del domino con i due numeri differenti sono $\binom{n}{2}$; quelle in cui i due numeri sono uguali sono n , in totale sono $\binom{n}{2} + n$. ■

Bibliografia

- [1] Baldi, P. (1998) *Calcolo delle probabilità e statistica*, Mc Graw Hill Italia.
- [2] Baldi, P. Giuliano R., Ladelli, L. (1995) *Laboratorio di Statistica e Probabilità, problemi svolti*, Mc Graw Hill Italia.
- [3] Bramanti, M. (1998) *Calcolo delle probabilità e statistica*, Progetto Leonardo Bologna.
- [4] Feller, W. (1950) *An Introduction to Probability Theory and Its Applications*, volume 1. John Wiley & Sons.
- [5] Pagani, C.D. e Salsa, S. (1992) *Analisi Matematica Vol. 1 e 2* Masson, Milano.
- [6] Robert, C.P. e Casella, G. (1999) *Monte Carlo Statistical Methods* Springer, New York.
- [7] Ross, S.M. (2007) *Calcolo delle probabilità, Settima Ed.* Apogeo.
- [8] Ross, S.M. (2015) *Probabilità e statistica per l'ingegneria e le scienze, Terza Ed.* Editore Apogeo Education.
- [9] Roussas, G.G. (1997) *A Course in Mathematical Statistics*, Academic Press.