

Francesco Caravenna
Paolo Dai Pra

Probabilità

Un'introduzione attraverso
modelli e applicazioni



 Springer



TEXTIN

UNITEXT – La Matematica per il 3+2

Volume 67

For further volumes:
<http://www.springer.com/series/5418>

Francesco Caravenna · Paolo Dai Pra

Probabilità

**Un'introduzione attraverso
modelli e applicazioni**

 Springer

Francesco Caravenna

Dipartimento Matematica e Applicazioni
Università degli Studi di Milano-Bicocca
Milano

Paolo Dai Pra

Dipartimento Matematica Pura e Applicata
Università degli Studi di Padova
Padova

Ulteriore materiale collegato al libro può essere scaricato da <http://extras.springer.com>
Password: 978-88-470-2594-3

UNITEXT – La Matematica per il 3+2

ISSN versione cartacea: 2038-5722

ISSN versione elettronica: 2038-5757

ISBN 978-88-470-2594-3

ISBN 978-88-470-2595-0 (eBook)

DOI 10.1007/978-88-470-2595-0

Springer Milan Heidelberg New York Dordrecht London

© Springer-Verlag Italia 2013

Quest'opera è protetta dalla legge sul diritto d'autore e la sua riproduzione è ammessa solo ed esclusivamente nei limiti stabiliti dalla stessa. Le fotocopie per uso personale possono essere effettuate nei limiti del 15% di ciascun volume dietro pagamento alla SIAE del compenso previsto dall'art. 68. Le riproduzioni per uso non personale e/o oltre il limite del 15% potranno avvenire solo a seguito di specifica autorizzazione rilasciata da AIDRO, Corso di Porta Romana n. 108, Milano 20122, e-mail segreteria@aidro.org e sito web www.aidro.org.

Tutti i diritti, in particolare quelli relativi alla traduzione, alla ristampa, all'utilizzo di illustrazioni e tabelle, alla citazione orale, alla trasmissione radiofonica o televisiva, alla registrazione su microfilm o in database, o alla riproduzione in qualsiasi altra forma (stampata o elettronica) rimangono riservati anche nel caso di utilizzo parziale. La violazione delle norme comporta le sanzioni previste dalla legge.

L'utilizzo in questa pubblicazione di denominazioni generiche, nomi commerciali, marchi registrati, ecc. anche se non specificatamente identificati, non implica che tali denominazioni o marchi non siano protetti dalle relative leggi e regolamenti.

9 8 7 6 5 4 3 2 1

Lay-out di copertina: Beatrice D., Milano

Impaginazione: PTP-Berlin, Protago TeX-Production GmbH, Germany (www.ptp-berlin.eu)

Springer-Verlag Italia S.r.l., Via Decembrio 28, I-20137 Milano

Springer-Verlag fa parte di Springer Science+Business Media (www.springer.com)

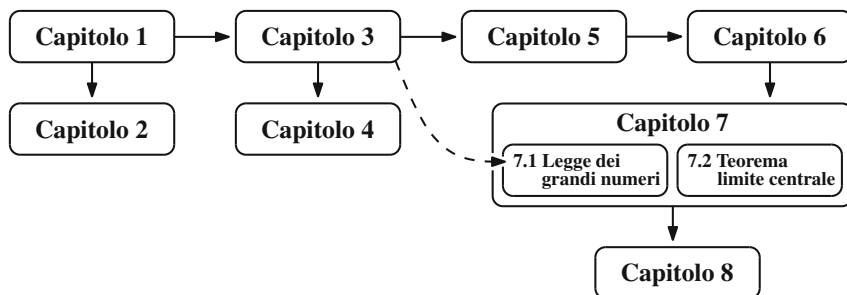
Prefazione

L'obiettivo di questo libro è di fornire un'introduzione alla teoria della probabilità e alle sue applicazioni, senza fare ricorso alla teoria della misura, per studenti dei corsi di laurea scientifici (in particolar modo di matematica, fisica e ingegneria). La scelta degli argomenti e l'approccio adottato sono il frutto di alcuni anni di esperienza con i corsi da noi tenuti per la laurea triennale in matematica, presso le università di Padova e di Milano-Bicocca.

Si è deciso di porre grande enfasi sulla probabilità *discreta*, vale a dire su spazi finiti o numerabili, a cui i primi quattro capitoli sono dedicati. La prima ragione è che, in questo contesto, sono sufficienti pochi strumenti analitici per presentare la teoria in modo completo e rigoroso (bastano sostanzialmente successioni e serie). Questo permette di introdurre il linguaggio e le nozioni basilari di probabilità senza eccessive complicazioni tecniche, concentrando l'attenzione sulle difficoltà sostanziali che gli studenti incontrano nella fase iniziale dello studio di questa disciplina. La seconda ragione è che poche nozioni di probabilità discreta sono più che sufficienti per discutere problemi e modelli estremamente interessanti, alcuni tuttora oggetto di ricerca. Una selezione di esempi in questa direzione è presentata nel Capitolo 2, mentre problemi più avanzati, che coinvolgono variabili aleatorie, sono descritti nel Capitolo 4. Riteniamo che la trattazione di uno o più di tali esempi già nella prima parte del corso costituisca un ottimo elemento formativo.

La trattazione degli spazi di probabilità generali, nel Capitolo 5, è piuttosto succinta e principalmente focalizzata alla discussione delle variabili aleatorie assolutamente continue, oggetto del Capitolo 6. In queste parti del testo diverse dimostrazioni sono omesse, ma si è cercato di dare sempre definizioni matematicamente precise, esplicitando le questioni tecniche che non possono essere risolte con gli strumenti a disposizione. I prerequisiti sono al livello di un primo corso di analisi matematica (limiti, derivate, integrale di Riemann), ad eccezione dei paragrafi conclusivi sui vettori aleatori, segnalati con un asterisco *, per i quali è richiesta la conoscenza di un po' di analisi multivariata (integrale di Riemann multidimensionale).

Vengono quindi presentati nel Capitolo 7 i teoremi limite classici del calcolo delle probabilità, ossia la legge (debole) dei grandi numeri e il teorema limite centrale. Per quest'ultimo, viene fornita una dimostrazione completa (con l'ipotesi di momen-



Schema delle dipendenze tra i capitoli. La via d'accesso più naturale al Capitolo 7 è quella che proviene dal Capitolo 6; tuttavia, volendo, la prima metà del Capitolo 7, sulla legge dei grandi numeri, è accessibile già dopo il Capitolo 3

to terzo finito) e viene discussa in dettaglio la tecnica dell'approssimazione normale. Infine, il Capitolo 8 è dedicato ad alcune applicazioni alla statistica matematica. Sopra è riportato un diagramma con i possibili ordini di lettura.

L'esposizione è arricchita da numerosi esempi, che costituiscono una parte fondamentale della presentazione, e da una vasta selezione di esercizi, per i quali viene fornita la soluzione dettagliata sulla pagina del sito Springer dedicata al volume (accessibile all'indirizzo <http://extras.springer.com> mediante il codice 978-88-470-2594-3). Alcune parti piuttosto tecniche, o che abbiamo ritenuto non essenziali, appaiono in corpo minore, oppure sono contenute nell'Appendice.

Questo libro, com'è ovvio, risente della nostra formazione, dei nostri interessi di ricerca e del nostro gusto. Siamo stati ispirati e aiutati da pareri e osservazioni di vari colleghi, che ringraziamo di cuore per i loro suggerimenti. Un grazie particolare a Wolfgang J. Runggaldier e Tiziano Vargiolu, per le numerose e utili discussioni.

Siamo inoltre debitori a numerosi autori di articoli e libri, dai quali abbiamo imparato molta della matematica che qui presentiamo. In particolare crediamo, e speriamo, di essere stati influenzati dai due splendidi testi di William Feller [23] e Patrick Billingsley [6].

Infine, siamo riconoscenti agli studenti della laurea triennale in matematica delle Università di Padova e di Milano-Bicocca che, con il loro studio, i loro commenti, le loro critiche e segnalazioni di errori, hanno contribuito alla progettazione e costruzione di questo libro.

Milano e Padova, marzo 2013

Francesco Caravenna
Paolo Dai Pra

Indice

Nozioni preliminari	1
Notazioni	1
Alcuni richiami di analisi matematica	3
Somme infinite	3
1 Spazi di probabilità discreti: teoria	7
1.1 Modelli probabilistici discreti	7
1.1.1 Considerazioni introduttive	7
1.1.2 Assiomi della probabilità	10
1.1.3 Probabilità e densità discreta	12
1.1.4 Proprietà fondamentali	15
1.2 Calcolo combinatorio	20
1.2.1 Principi basilari	20
1.2.2 Disposizioni con ripetizione	21
1.2.3 Il principio fondamentale	22
1.2.4 Disposizioni semplici e permutazioni	24
1.2.5 Combinazioni	28
1.2.6 Estrazioni di palline da un'urna	29
1.3 Probabilità condizionale e indipendenza	33
1.3.1 Probabilità condizionale	33
1.3.2 Bayes e dintorni	36
1.3.3 Indipendenza di eventi	40
1.3.4 Prove ripetute e indipendenti	47
1.3.5 Esempi e paradossi sul condizionamento	52
1.4 Esercizi di riepilogo	61
1.5 Note bibliografiche	67
2 Spazi di probabilità discreti: esempi e applicazioni	69
2.1 Permutazioni aleatorie	69
2.2 La passeggiata aleatoria semplice	77
2.3 Statistiche classiche e quantistiche	84

2.4	Il modello di Ising in meccanica statistica	90
2.5	Il modello di Hardy-Weinberg in genetica	98
2.6	Note bibliografiche	101
3	Variabili aleatorie discrete: teoria	103
3.1	Variabili aleatorie e distribuzioni	103
3.1.1	Considerazioni introduttive	103
3.1.2	Definizioni	104
3.1.3	Probabilità discrete su spazi generali	106
3.1.4	Distribuzione e densità discreta	107
3.1.5	Osservazioni ed esempi	111
3.1.6	Costruzione canonica di una variabile aleatoria	115
3.2	Indipendenza di variabili aleatorie	117
3.2.1	Distribuzioni congiunte e marginali	117
3.2.2	Indipendenza di variabili aleatorie	120
3.2.3	Rivisitazione delle prove ripetute e indipendenti	124
3.2.4	Proprietà dell'indipendenza	125
3.2.5	Costruzione di variabili aleatorie indipendenti	127
3.2.6	Dallo spazio di probabilità alle variabili aleatorie	128
3.3	Valor medio e disuguaglianze	130
3.3.1	Definizione	130
3.3.2	Proprietà del valor medio	134
3.3.3	Momenti, varianza e covarianza	139
3.3.4	Valor medio e indipendenza	146
3.3.5	Disuguaglianze	147
3.3.6	Coefficiente di correlazione	151
3.4	Lavorare con le distribuzioni	154
3.4.1	Somma di variabili aleatorie	154
3.4.2	Funzione di ripartizione	155
3.4.3	Massimo e minimo di variabili aleatorie indipendenti	158
3.4.4	Funzione generatrice dei momenti	159
3.5	Classi notevoli di variabili aleatorie discrete	164
3.5.1	Uniforme discreta	164
3.5.2	Bernoulli	166
3.5.3	Binomiale	167
3.5.4	Ipergeometrica	171
3.5.5	Poisson	172
3.5.6	Geometrica	176
3.6	Esercizi di riepilogo	182
3.7	Note bibliografiche	187
4	Variabili aleatorie discrete: esempi e applicazioni	189
4.1	Sulla legge dei piccoli numeri	189
4.2	Un'applicazione alla finanza: il modello binomiale	193
4.3	Il problema del collezionista di figurine	200

4.4	Mescolare un mazzo di carte	204
4.5	Rivisitazione delle passeggiate aleatorie	211
4.6	La condensazione di Bose-Einstein	218
4.7	Note bibliografiche	229
5	Spazi di probabilità e variabili aleatorie generali	231
5.1	σ -algebre e misure di probabilità	231
5.2	Variabili aleatorie generali	235
5.3	Indipendenza e valor medio	238
5.4	Costruzione di modelli probabilistici	241
5.5	Note bibliografiche	242
6	Variabili aleatorie assolutamente continue	243
6.1	Richiami sull'integrale di Riemann	243
6.1.1	L'integrale in senso proprio	243
6.1.2	L'integrale in senso improprio	244
6.1.3	Alcuni esempi	246
6.1.4	Approfondimenti sull'integrabilità	248
6.1.5	Proprietà dell'integrale	250
6.2	Variabili aleatorie reali assolutamente continue	251
6.2.1	Definizione e prime proprietà	252
6.2.2	Determinare la densità	254
6.2.3	Il calcolo del valor medio	257
6.2.4	Calcoli con variabili aleatorie indipendenti	259
6.3	Classi notevoli di variabili aleatorie reali assolutamente continue ...	261
6.3.1	Uniforme continua	261
6.3.2	Gamma	263
6.3.3	Esponenziale	266
6.3.4	Normale	267
6.4	Vettori aleatori assolutamente continui *	271
6.4.1	Definizione e prime proprietà *	273
6.4.2	Densità congiunta e marginali *	275
6.4.3	Calcoli con densità*	278
6.5	Esempi e applicazioni	282
6.5.1	Le variabili aleatorie chi-quadro	282
6.5.2	Statistiche d'ordine e variabili aleatorie Beta	283
6.5.3	Il processo di Poisson (parte I)	286
6.5.4	Il processo di Poisson (parte II) *	289
6.5.5	I vettori aleatori uniformi e il paradosso di Bertrand *	292
6.6	Vettori aleatori normali *	294
6.6.1	Matrice di covarianza *	294
6.6.2	Definizione e proprietà principali *	296
6.6.3	Proiezioni ortogonali di vettori normali *	300
6.7	Esercizi di riepilogo	303
6.8	Note bibliografiche	311

7	Teoremi limite	313
7.1	La legge dei grandi numeri	313
7.1.1	Enunciato, dimostrazione e discussione	314
7.1.2	Il metodo Monte Carlo per il calcolo di integrali	317
7.1.3	Il teorema di approssimazione di Weierstrass	319
7.1.4	Un esempio con variabili aleatorie correlate	321
7.2	Il teorema limite centrale	324
7.2.1	Enunciato e discussione	324
7.2.2	Il metodo dell'approssimazione normale	327
7.2.3	Dimostrazione del teorema limite centrale	333
7.2.4	Un teorema limite locale per variabili esponenziali	338
7.3	Esercizi di riepilogo	340
7.4	Note bibliografiche	343
8	Applicazioni alla statistica matematica	345
8.1	Modelli statistici parametrici	345
8.2	Intervalli di confidenza per campioni normali	350
8.3	Proprietà asintotiche	354
8.4	Stimatori di massima verosimiglianza	358
8.5	Note bibliografiche	372
	Appendice	373
A.1	Somme infinite	373
A.2	Una misura finitamente additiva (ma non σ -additiva) su \mathbb{N}	378
A.3	Il principio fondamentale del calcolo combinatorio	382
	Tavola della distribuzione normale	385
	Principali distribuzioni notevoli su \mathbb{R}	387
	Riferimenti bibliografici	389
	Indice analitico	393

Nozioni preliminari

Sommario. In questo capitolo fissiamo le notazioni più importanti che saranno usate in questo libro e richiamiamo alcuni risultati di analisi matematica che ricorreranno di frequente. Definiamo quindi la nozione di somma per una famiglia infinita di numeri reali indicizzata da un insieme generico, discutendone le proprietà.

Notazioni

Dato un insieme Ω e due suoi sottoinsiemi $A, B \subseteq \Omega$, useremo le notazioni standard

$$\begin{aligned} A \cup B &:= \{\omega \in \Omega : \omega \in A \text{ o } \omega \in B\}, \\ A \cap B &:= \{\omega \in \Omega : \omega \in A \text{ e } \omega \in B\}, \\ A^c &:= \{\omega \in \Omega : \omega \notin A\}, \\ A \setminus B &:= A \cap B^c, \\ A \triangle B &:= (A \setminus B) \cup (B \setminus A) = (A \cup B) \setminus (A \cap B), \end{aligned}$$

dove il simbolo “:=” indica una definizione. Le nozioni di unione e intersezione si estendono in modo naturale a una famiglia arbitraria $(A_i)_{i \in I}$ di sottoinsiemi di Ω :

$$\begin{aligned} \bigcup_{i \in I} A_i &:= \{\omega \in \Omega : \exists i \in I \text{ tale che } \omega \in A_i\}, \\ \bigcap_{i \in I} A_i &:= \{\omega \in \Omega : \forall i \in I \text{ si ha che } \omega \in A_i\}. \end{aligned}$$

Con un piccolo abuso di terminologia, diremo che un’unione $\bigcup_{i \in I} A_i$ è *disgiunta* se gli eventi $(A_i)_{i \in I}$ sono *a due a due disgiunti*, ossia se $A_i \cap A_j = \emptyset$ per ogni $i \neq j$ (nel seguito, per brevità, scriveremo semplicemente disgiunti).

Useremo spesso la proprietà distributiva dell’intersezione rispetto all’unione:

$$(A \cup B) \cap C = (A \cap C) \cup (B \cap C),$$

e le leggi di De Morgan:

$$(A \cup B)^c = A^c \cap B^c, \quad (A \cap B)^c = A^c \cup B^c,$$

che valgono, più in generale, per famiglie arbitrarie di insiemi:

$$\begin{aligned} \left(\bigcup_{i \in I} A_i \right) \cap C &= \bigcup_{i \in I} (A_i \cap C), \\ \left(\bigcup_{i \in I} A_i \right)^c &= \bigcap_{i \in I} A_i^c, \quad \left(\bigcap_{i \in I} A_i \right)^c = \bigcup_{i \in I} A_i^c. \end{aligned}$$

Indicheremo con $\mathbb{N} := \{1, 2, 3, \dots\}$ i numeri naturali, *zero escluso*; quando vorremo includerlo, useremo la notazione $\mathbb{N}_0 := \{0, 1, 2, \dots\}$. Adotteremo le notazioni standard per i numeri interi, razionali e reali, indicati rispettivamente con \mathbb{Z} , \mathbb{Q} e \mathbb{R} , e porremo $\mathbb{R}^+ := [0, \infty) = \{x \in \mathbb{R} : x \geq 0\}$.

Diremo che un numero $x \in \mathbb{R}$ è *positivo* se $x \geq 0$ e *strettamente positivo* se $x > 0$; analogamente, diremo che x è *negativo* se $x \leq 0$ e *strettamente negativo* se $x < 0$. Si noti che con queste convenzioni 0 è sia positivo sia negativo. La parte positiva e negativa di un numero $x \in \mathbb{R}$ sono definite rispettivamente da

$$x^+ := \max\{x, 0\}, \quad x^- := -\min\{x, 0\} = \max\{-x, 0\},$$

così che $x^+, x^- \geq 0$ e $x = x^+ - x^-$. Il valore assoluto di x è dato da $|x| = x^+ + x^-$.

Utilizzeremo gli aggettivi “crescente” e “decrescente” in senso debole: una funzione $f : \mathbb{R} \rightarrow \mathbb{R}$ è crescente (risp. decrescente) se per ogni $x \geq y$ si ha $f(x) \geq f(y)$ (risp. $f(x) \leq f(y)$). Una funzione costante è dunque sia crescente sia decrescente.

Il prodotto cartesiano $A \times B$ di due insiemi A e B è l'insieme delle coppie ordinate (a, b) con $a \in A$ e $b \in B$. Si definisce analogamente il prodotto di più di due insiemi. Indicheremo con A^n il prodotto $A \times \dots \times A$ (n volte), per $n \in \mathbb{N}$.

Ricordiamo che un insieme A si dice numerabile se è in corrispondenza biunivoca con \mathbb{N} , cioè se esiste una applicazione $f : A \rightarrow \mathbb{N}$ iniettiva e suriettiva.

Dato un insieme *finito* A , ne denoteremo col simbolo $|A|$ la cardinalità, cioè il numero dei suoi elementi, e scriveremo $|A| < \infty$ per indicare che l'insieme è finito.

Dato un insieme generico Ω , per ogni sottoinsieme $A \subseteq \Omega$ è definita la funzione $\mathbb{1}_A : \Omega \rightarrow \mathbb{R}$, detta *indicatrice di A*, che vale 1 su A e 0 su A^c :

$$\mathbb{1}_A(x) := \begin{cases} 1 & \text{se } x \in A \\ 0 & \text{se } x \notin A \end{cases}. \quad (0.1)$$

In altri contesti, questa funzione è detta caratteristica e indicata con χ_A . In probabilità è prassi indicare con “funzione caratteristica” un oggetto diverso (che non descriveremo in questo libro), pertanto ci atterremo alla terminologia probabilistica “funzione indicatrice” e alla notazione $\mathbb{1}_A$.

Alcuni richiami di analisi matematica

Data una successione di numeri reali $(a_n)_{n \in \mathbb{N}}$, la *serie* $\sum_{n=1}^{\infty} a_n$ è definita come il limite della successione delle somme parziali $(s_N := \sum_{n=1}^N a_n)_{N \in \mathbb{N}}$, quando tale limite esiste in $\mathbb{R} \cup \{\pm\infty\}$. Ricordiamo che per ogni $a \in \mathbb{R}$ si ha

$$\sum_{n=1}^{\infty} \frac{1}{n^a} \begin{cases} < +\infty & \text{se } a > 1 \\ = +\infty & \text{se } a \leq 1 \end{cases}, \quad (0.2)$$

come si mostra, ad esempio, confrontando la serie con un opportuno integrale.

Le somme parziali della *serie geometrica* (a partire da 0 o da 1) sono note:

$$\sum_{n=0}^N x^n = \frac{1-x^{N+1}}{1-x}, \quad \sum_{n=1}^N x^n = \frac{x(1-x^N)}{1-x}, \quad \forall x \in \mathbb{R} \setminus \{1\}, \quad (0.3)$$

come si dimostra facilmente per induzione. Segue in particolare che

$$\sum_{n=0}^{\infty} x^n = \frac{1}{1-x}, \quad \sum_{n=1}^{\infty} x^n = \frac{x}{1-x}, \quad \forall x \in \mathbb{R} \text{ con } |x| < 1. \quad (0.4)$$

Entrambe queste serie valgono $+\infty$ se $x \geq 1$, mentre esse non sono definite se $x \leq -1$ (la successione delle somme parziali non ha limite).

Ricordiamo le serie di Taylor delle funzioni esponenziale e logaritmo:

$$e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!} \quad \forall x \in \mathbb{R}, \quad (0.5)$$

$$\log(1+x) = \sum_{n=1}^{\infty} (-1)^{n+1} \frac{x^n}{n} \quad \forall x \in (-1, 1], \quad (0.6)$$

dove il simbolo $n!$ (“ n fattoriale”) è richiamato nel Paragrafo 1.2.4.

Ricordiamo inoltre le seguenti formule esplicite per la somma dei numeri interi da 1 a n e dei loro quadrati, che si dimostrano facilmente per induzione:

$$\sum_{k=1}^n k = \frac{n(n+1)}{2}, \quad \sum_{k=1}^n k^2 = \frac{n(n+1)(2n+1)}{6}, \quad \forall n \in \mathbb{N}. \quad (0.7)$$

Somme infinite

Per definire la somma di una famiglia di numeri reali $(a_i)_{i \in I}$ indicizzata da un insieme infinito I generico, anche più che numerabile e senza un ordinamento, si procede nel modo seguente. Nel caso speciale di una famiglia $(a_i)_{i \in I}$ a termini positivi, ossia $a_i \geq 0$ per ogni $i \in I$, la somma $\sum_{i \in I} a_i$ è sempre ben definita come l'estremo superiore di tutte le possibili somme finite degli elementi a_i :

$$\sum_{i \in I} a_i := \sup_{A \subseteq I, |A| < \infty} \sum_{j \in A} a_j \in [0, +\infty].$$

Si noti che $\sum_{i \in I} a_i$ può assumere il valore $+\infty$.

Nel caso generale di una famiglia $(a_i)_{i \in I}$ di numeri reali, non necessariamente positivi, diremo che tale famiglia *ammette somma* se almeno una delle due somme $\sum_{i \in I} a_i^+$, $\sum_{i \in I} a_i^-$ è finita (si osservi che sono entrambe somme a termini positivi, dunque sono sempre ben definite). In tal caso, la somma $\sum_{i \in I} a_i$ è definita da

$$\sum_{i \in I} a_i := \sum_{i \in I} a_i^+ - \sum_{i \in I} a_i^- \in [-\infty, +\infty]. \quad (0.8)$$

Si noti che ogni famiglia con tutti i termini positivi, o tutti negativi, ammette somma.

Se una famiglia $(a_i)_{i \in I}$ ammette somma e se $\sum_{i \in I} a_i \in (-\infty, +\infty)$, diremo che la famiglia *ammette somma finita*. Chiaramente ciò accade se e solo se *entrambe* le somme $\sum_{i \in I} a_i^+$ e $\sum_{i \in I} a_i^-$ sono finite, il che è equivalente a $\sum_{i \in I} |a_i| < \infty$, per la proprietà di linearità richiamata più sotto.

Sottolineiamo che, per una famiglia $(a_n)_{n \in \mathbb{N}}$ indicizzata dai numeri naturali, ossia per una successione, la nozione di *ammettere somma finita* corrisponde alla *convergenza assoluta* della serie corrispondente, ossia $\sum_{n=1}^{\infty} |a_n| < \infty$.

Osservazione 0.1. La definizione data di somma infinita $\sum_{i \in I} a_i$ si può estendere senza alcuna modifica al caso in cui gli elementi della famiglia $(a_i)_{i \in I}$ appartengano alla *retta reale estesa* $\mathbb{R} := [-\infty, +\infty]$, con l'ovvia convenzione che una somma di elementi di $[0, +\infty]$ vale $+\infty$ non appena almeno uno degli addendi vale $+\infty$. \square

Osservazione 0.2. L'insieme degli indici I può anche essere infinito più che numerabile, ma se la famiglia $(a_i)_{i \in I}$ ammette somma finita, ossia se $\sum_{i \in I} |a_i| < \infty$, gli indici dei termini non nulli sono necessariamente un insieme finito o numerabile. Infatti, per ogni $\varepsilon > 0$ fissato, i termini tali che $|a_i| > \varepsilon$ sono in numero finito (in caso contrario si mostra facilmente che $\sum_{i \in I} |a_i| = \infty$). Scrivendo quindi

$$\{i \in I : a_i \neq 0\} = \bigcup_{n \in \mathbb{N}} \{i \in I : |a_i| > \frac{1}{n}\}, \quad (0.9)$$

segue che l'insieme $\{i \in I : a_i \neq 0\}$ è unione numerabile di insiemi finiti, dunque è finito oppure numerabile. \square

Osservazione 0.3. Supponiamo che una famiglia $(a_i)_{i \in I}$ abbia un insieme numerabile di termini non nulli (sarà il caso tipico che incontreremo). Essi possono allora essere elencati in una successione $(a_{i_n})_{n \in \mathbb{N}}$, ma la scelta di tale successione (ossia dell'ordine in cui elencare i termini non nulli) è arbitraria.

Se la famiglia $(a_i)_{i \in I}$ ammette somma (finita o infinita), in particolare se i termini sono tutti positivi o tutti negativi, la somma $\sum_{i \in I} a_i$ coincide con la serie $\sum_{n=1}^{\infty} a_{i_n} = \lim_{N \rightarrow \infty} \sum_{n=1}^N a_{i_n}$, la quale dunque *non dipende* dal modo di scegliere la successione $(a_{i_n})_{n \in \mathbb{N}}$ (e fornisce un modo operativo di calcolare la somma $\sum_{i \in I} a_i$).

Questo non è più vero se la famiglia $(a_i)_{i \in I}$ non ammette somma: ordinamenti diversi della successione $(a_{i_n})_{n \in \mathbb{N}}$ possono produrre valori diversi della serie $\sum_{n=1}^{\infty} a_{i_n}$, o anche serie che non convergono. Questa è la ragione per cui occorre cautela nel definire la somma di una famiglia infinita $(a_i)_{i \in I}$ quando non c'è un ordinamento canonico dell'insieme degli indici I (o, in presenza di un ordinamento, quando si vuole che la somma non dipenda dall'ordine in cui vengono elencati i termini).

Un esempio classico è dato dalla famiglia $(a_i := (-1)^i \frac{1}{i})_{i \in \mathbb{N}}$. La serie corrispondente è convergente: infatti, per la relazione (0.6),

$$\sum_{i=1}^{\infty} \frac{(-1)^i}{i} := \lim_{N \rightarrow \infty} \sum_{i=1}^N \frac{(-1)^i}{i} = -\log 2.$$

Ciononostante, la famiglia *non ammette somma*, secondo le definizioni date: infatti

$$\sum_{i \in \mathbb{N}} a_i^+ = \sum_{i \in 2\mathbb{N}} \frac{1}{i} = +\infty, \quad \sum_{i \in \mathbb{N}} a_i^- = \sum_{i \in 2\mathbb{N}_0+1} \frac{1}{i} = +\infty,$$

avendo indicato con $2\mathbb{N}$ e $2\mathbb{N}_0 + 1$ rispettivamente i numeri pari e dispari. In effetti, in questo caso è possibile mostrare che, scegliendo un opportuno riordinamento $(a_{i_n})_{n \in \mathbb{N}}$ della famiglia, si può far convergere la serie corrispondente $\sum_{n=1}^{\infty} a_{i_n}$ verso un arbitrario numero reale prefissato! \square

Enunciamo alcune proprietà importanti delle somme infinite, di cui faremo uso frequente. Le dimostrazioni non sono difficili, ma i dettagli sono piuttosto noiosi: il lettore interessato li può trovare nell'Appendice A.1.

Cominciamo con la *linearità*: se le famiglie $(a_i)_{i \in I}$ e $(b_i)_{i \in I}$ ammettono somma finita, per ogni $\alpha, \beta \in \mathbb{R}$ la famiglia $(\alpha a_i + \beta b_i)_{i \in I}$ ammette somma finita e si ha

$$\sum_{i \in I} (\alpha a_i + \beta b_i) = \alpha \left(\sum_{i \in I} a_i \right) + \beta \left(\sum_{i \in I} b_i \right).$$

Questa relazione vale anche se le famiglie sono positive, senza richiedere che ammettano somma finita, purché $\alpha, \beta \geq 0$.

Quindi la *monotonia*: se le due famiglie $(a_i)_{i \in I}$ e $(b_i)_{i \in I}$ ammettono somma (finita o infinita) e sono tali che $a_i \leq b_i$ per ogni $i \in I$, allora

$$\sum_{i \in I} a_i \leq \sum_{i \in I} b_i.$$

Descriviamo infine una proprietà, nota come *somma a blocchi*, che corrisponde a una generalizzazione dell'associatività di somme finite. Supponiamo che la famiglia $(a_i)_{i \in I}$ ammetta somma e sia $(I_j)_{j \in J}$ una *partizione* di I , ossia

$$I = \bigcup_{j \in J} I_j, \quad I_j \cap I_k = \emptyset \quad \text{per } j \neq k.$$

Allora, per ogni $j \in J$ fissato, la sottofamiglia $(a_i)_{i \in I_j}$ ammette somma; ponendo $s_j := \sum_{i \in I_j} a_i$, anche la famiglia $(s_j)_{j \in J}$ ammette somma (come famiglia in $[-\infty, +\infty]$, si veda l'Osservazione 0.1) e si ha

$$\sum_{i \in I} a_i = \sum_{j \in J} s_j = \sum_{j \in J} \left(\sum_{i \in I_j} a_i \right). \quad (0.10)$$

Come caso particolare, un insieme prodotto $I = E \times F$ può essere partizionato in modo naturale come $I = \bigcup_{x \in E} \{x\} \times F$, oppure come $I = \bigcup_{y \in F} E \times \{y\}$. Di conseguenza, se la famiglia $(a_i)_{i \in I} = (a_{x,y})_{(x,y) \in E \times F}$ ammette somma, vale la seguente versione discreta del teorema di Fubini-Tonelli:

$$\sum_{(x,y) \in E \times F} a_{x,y} = \sum_{x \in E} \left(\sum_{y \in F} a_{x,y} \right) = \sum_{y \in F} \left(\sum_{x \in E} a_{x,y} \right). \quad (0.11)$$

Questa relazione si estende al caso in cui $I = E_1 \times \dots \times E_n$ sia il prodotto cartesiano di un numero finito di insiemi.

Spazi di probabilità discreti: teoria

Sommario. In questo capitolo introduciamo la nozione fondamentale di spazio di probabilità discreto, per descrivere un esperimento aleatorio in cui l'insieme degli esiti sia finito o numerabile, e ne studiamo le principali proprietà. Sviluppiamo quindi le tecniche classiche di conteggio per insiemi finiti, note come calcolo combinatorio. Arrivati a questo punto, è già possibile, e forse persino consigliabile, affrontare l'analisi di uno o più dei modelli probabilistici presentati nel Capitolo 2. Infine, discutiamo le nozioni cruciali di condizionamento e indipendenza di eventi, illustrandole con una selezione di esempi e paradossi.

1.1 Modelli probabilistici discreti

1.1.1 Considerazioni introduttive

In questo libro, con la dicitura *esperimento aleatorio* indicheremo un'osservazione relativa ad un qualunque fenomeno (fisico, economico, sociale, ...) il cui esito non sia determinabile con certezza a priori. Il nostro obiettivo è di fornire una descrizione matematica di un esperimento aleatorio, definendo un *modello probabilistico*.

Il primo passo consiste nell'identificare un insieme Ω , detto *spazio campionario*, che contiene tutti gli esiti possibili dell'esperimento.

Esempio 1.1. (i) Per il lancio di un dado ordinario a sei facce, lo spazio campionario naturale è $\Omega = \{1, 2, 3, 4, 5, 6\}$.

(ii) Per il numero di voti ottenuti dal politico Cetto alle elezioni del prossimo anno, scelte possibili per lo spazio campionario sono $\Omega = \{0, 1, \dots, M\}$, dove $M \in \mathbb{N}$ è il numero di individui aventi diritto di voto, oppure $\Omega = \mathbb{N}_0$.

(iii) Per la misurazione del tempo (diciamo espresso in secondi e frazioni di secondo) in cui una particella radioattiva viene emessa da un determinato atomo, una scelta naturale di spazio campionario è data da $\Omega = [0, +\infty)$. \square

Il secondo elemento di un modello probabilistico è costituito dagli *eventi*, intesi informalmente come *affermazioni sull'esito di un esperimento aleatorio*. Matematicamente, gli eventi sono descritti da opportuni *sottoinsiemi dello spazio campionario* Ω : ogni affermazione è identificata con il sottoinsieme di Ω costituito da tutti e soli gli esiti dell'esperimento per cui l'affermazione si verifica. Quando non c'è nessun esito che verifica l'affermazione, l'evento è detto impossibile ed è identificato con l'insieme vuoto \emptyset . All'estremo opposto, un'affermazione che è vera qualunque sia l'esito dell'esperimento corrisponde all'evento certo, dato dall'intero insieme Ω .

Esempio 1.2. Facciamo riferimento all'esempio precedente.

(i) Lanciando un dado a sei facce, sono esempi di eventi:

$$A := \text{"Esce un numero pari"} = \{2, 4, 6\},$$

$$B := \text{"Esce un numero multiplo di 3"} = \{3, 6\},$$

$$C := \text{"Esce un numero pari multiplo di 3"} = \{6\} = A \cap B.$$

(ii) L'evento $D = \text{"Cetto non ottiene alcun voto alle elezioni del prossimo anno"}$ corrisponde al sottoinsieme $D = \{0\} \subseteq \Omega$. Un esempio di evento impossibile è dato da $E := \{\text{Cetto ottiene } M + 1 \text{ voti}\} = \emptyset$, essendo M il numero degli aventi diritto al voto.

(iii) L'evento $F = \text{"La particella viene emessa dopo più di un anno"}$ corrisponde al sottoinsieme $F = (60 \cdot 60 \cdot 24 \cdot 365, \infty) = (31\,536\,000, \infty) \subseteq \Omega$. \square

Si noti che le operazioni logiche di *congiunzione*, *disgiunzione* e *negazione* di affermazioni corrispondono all'*intersezione*, *unione* e *complementare* di insiemi:

$$\text{"si verificano sia } A \text{ sia } B" \longrightarrow A \cap B,$$

$$\text{"si verifica almeno uno tra } A \text{ e } B" \longrightarrow A \cup B, \quad (1.1)$$

$$\text{"non si verifica } A" \longrightarrow A^c.$$

Osservazione 1.3. Una questione delicata è se sia opportuno considerare eventi *tutti* i sottoinsiemi di uno spazio campionario Ω . Questo è senz'altro possibile quando Ω è finito o numerabile, mentre se Ω è infinito non numerabile (come nell'esempio (iii) visto sopra) può risultare necessario considerare eventi solo una classe ristretta di sottoinsiemi di Ω , per ragioni che saranno chiare nel seguito. In questo caso l'analisi diventa più tecnica e complicata, e per questa ragione nella prima parte di questo libro ci concentreremo sugli spazi campionari finiti o numerabili. \square

Il terzo e ultimo ingrediente di un modello probabilistico, il più importante, è l'assegnazione di un "grado di fiducia", o *probabilità*, agli eventi di un esperimento aleatorio. Questo permette di formalizzare espressioni quali "La probabilità che esca un numero pari (lanciando un dado a sei facce) vale $1/2$ ". Matematicamente, una probabilità viene descritta da una applicazione P che a ogni evento $A \subseteq \Omega$ assegna un numero $P(A) \in [0, 1]$. Talvolta la probabilità di un evento viene espressa come percentuale, scrivendo 50% invece di $1/2$, 10% invece di 0.1, e così via.

Che significato concreto dare al “grado di fiducia” $P(A)$? Se l’esperimento aleatorio può essere ripetuto un numero elevato $N \gg 1$ di volte in condizioni “analoghe e indipendenti”, è possibile (almeno in linea di principio) contare il numero di volte $\mathcal{S}_N(A) \in \{0, \dots, N\}$ in cui l’evento A si verifica. In questo caso è naturale interpretare $P(A)$ come *la frazione di volte in cui l’evento A si verifica*, ossia $P(A) \simeq \mathcal{S}_N(A)/N$ (idealmente $P(A) = \lim_{N \rightarrow +\infty} \mathcal{S}_N(A)/N$). Questa è la celebre *interpretazione frequentista* della probabilità (che riceve una sorta di giustificazione a posteriori dalla *legge dei grandi numeri*, un teorema che studieremo nel Capitolo 7).

Sottolineiamo tuttavia che non si tratta di una definizione operativa: innanzitutto, non tutti gli esperimenti aleatori possono essere ripetuti in condizioni “analoghe e indipendenti” (si pensi al caso di un’elezione, come nell’Esempio 1.1 (ii)). Inoltre, anche quando è possibile farlo, niente garantisce a priori che il rapporto $\mathcal{S}_N(A)/N$ converga verso un limite e, se anche ciò avvenisse, non è chiaro quanto N debba essere grande perché l’approssimazione $P(A) \simeq \mathcal{S}_N(A)/N$ sia buona. Tuttavia, pur con le dovute cautele, suggeriamo di tenere sempre a mente l’interpretazione frequentista per dare contenuto intuitivo ai risultati che incontreremo.

La scelta della probabilità è in effetti un punto delicato. In alcuni casi esiste una scelta “naturale”, sulla base di considerazioni sulla natura dell’esperimento aleatorio in esame, in particolare di simmetrie. Molto spesso, però, non è così e *la scelta può dipendere da valutazioni soggettive*. L’osservazione cruciale è che, comunque venga scelta, una probabilità P dovrà soddisfare alcune proprietà, ad esempio:

$$P(\Omega) = 1, \quad P(A \cup B) = P(A) + P(B) \quad \text{se} \quad A \cap B = \emptyset. \quad (1.2)$$

Queste relazioni sono del tutto naturali se si pensa all’interpretazione frequentista $P(A) \simeq \mathcal{S}_N(A)/N$, dal momento che $\mathcal{S}_N(\Omega) = N$ (si ricordi che Ω contiene *tutti* gli esiti possibili dell’esperimento) e inoltre $\mathcal{S}_N(A \cup B) = \mathcal{S}_N(A) + \mathcal{S}_N(B)$ se $A \cap B = \emptyset$.

Il punto di vista moderno consiste nel definire probabilità una qualunque funzione P che soddisfi una versione rafforzata delle proprietà (1.2). Questo approccio ha conseguenze ricche e profonde, come avremo modo di apprezzare.

Osservazione 1.4. Se la *scelta* del modello probabilistico per un esperimento aleatorio può dipendere da considerazioni extra-matematiche, sottolineiamo che, una volta scelto il modello, il suo *studio* è un problema genuinamente matematico.

Osserviamo inoltre che, ove possibile, il modello scelto va sottoposto a verifica, sulla base di dati sperimentali. Questo costituisce uno degli obiettivi principali della *statistica matematica*, a cui accenneremo nel Capitolo 8. \square

1.1.2 Assiomi della probabilità

Motivati dalle precedenti considerazioni, diamo alcune definizioni fondamentali.

Definizione 1.5 (Assiomi della probabilità). Sia Ω un insieme (non vuoto) *finito o numerabile* e sia $\mathcal{P}(\Omega)$ la famiglia di tutti i sottoinsiemi di Ω . Una funzione $P: \mathcal{P}(\Omega) \rightarrow [0, 1]$ si dice *probabilità* se valgono le seguenti proprietà:

(P1) $P(\Omega) = 1$.

(P2) (σ -additività) Per ogni successione $(A_n)_{n \in \mathbb{N}}$ di sottoinsiemi di Ω disgiunti (ossia $A_n \cap A_m = \emptyset$ per $n \neq m$) si ha

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

La coppia (Ω, P) è detta *spazio di probabilità discreto*. L'insieme Ω è detto *spazio campionario* e i suoi sottoinsiemi sono chiamati *eventi*.

L'interpretazione di uno spazio di probabilità discreto (Ω, P) è dunque la seguente: l'insieme Ω contiene tutti i possibili esiti di un esperimento aleatorio e, per ogni sottoinsieme $A \subseteq \Omega$, il numero $P(A) \in [0, 1]$ esprime il “grado di fiducia” che si attribuisce all'eventualità che l'esito dell'esperimento sia un elemento di A .

La proprietà (P1) esprime il fatto che l'intero spazio campionario è un evento *certo*, mentre la proprietà (P2) richiede una discussione più accurata. Iniziamo col dedurne alcune conseguenze.

Proposizione 1.6. Sia (Ω, P) uno spazio di probabilità discreto. Valgono le seguenti proprietà:

(i) $P(\emptyset) = 0$.

(ii) (Additività finita) Se A_1, A_2, \dots, A_k è una famiglia finita di eventi disgiunti (ossia $A_i \cap A_j = \emptyset$ per $i \neq j$), allora

$$P(A_1 \cup A_2 \cup \dots \cup A_k) = \sum_{j=1}^k P(A_j). \quad (1.3)$$

Dimostrazione. Cominciamo con il punto (i). Sia $x := P(\emptyset) \in [0, 1]$ e si definisca $A_n = \emptyset$ per ogni $n \in \mathbb{N}$. Chiaramente $(A_n)_{n \in \mathbb{N}}$ è una successione di sottoinsiemi disgiunti di Ω . Allora, per l'assioma (P2) e il fatto che $\bigcup_{n=1}^{+\infty} A_n = \emptyset$, si ha

$$x = P(\emptyset) = P\left(\bigcup_{n=1}^{+\infty} A_n\right) = \sum_{n=1}^{+\infty} P(A_n) = \sum_{n=1}^{+\infty} x.$$

Tale identità è possibile se e solo se $x = 0$ (perché $\sum_{n=1}^{+\infty} x = \infty$ se $x > 0$).

Veniamo dunque al punto (ii). Prolunghiamo la famiglia di eventi disgiunti A_1, A_2, \dots, A_k ad una successione infinita di eventi disgiunti, ponendo $A_n := \emptyset$ per $n > k$. Allora $\bigcup_{j=1}^k A_j = \bigcup_{j=1}^{+\infty} A_j$ e per l'assioma (P2)

$$P(A_1 \cup A_2 \cup \dots \cup A_k) = P\left(\bigcup_{j=1}^{+\infty} A_j\right) = \sum_{j=1}^{+\infty} P(A_j) = \sum_{j=1}^k P(A_j),$$

che è quanto dovevamo mostrare. \square

Osservazione 1.7. Riscriviamo l'equazione (1.3) nel caso speciale $k = 2$:

$$P(A \cup B) = P(A) + P(B), \quad \forall A, B \subseteq \Omega \text{ tali che } A \cap B = \emptyset. \quad (1.4)$$

È interessante notare che la relazione (1.3) nel caso generale ($k \geq 2$) segue da (1.4) attraverso una semplice dimostrazione per induzione (esercizio). \square

La proprietà (1.3) (o equivalentemente (1.4)), detta *addittività finita* o semplicemente *addittività*, è una condizione “naturale”, che corrisponde ad un'idea intuitiva di probabilità, come abbiamo già osservato in (1.2). È pertanto significativo domandarsi se le coppie di assiomi $\{(P1), (P2)\}$ e $\{(P1), (1.3)\}$ siano equivalenti, cioè se da ciascuna si possa dedurre l'altra. La risposta è affermativa nel caso in cui Ω sia un insieme finito, dal momento che non esistono successioni infinite di eventi disgiunti e non vuoti (infatti $\mathcal{P}(\Omega)$ ha un numero finito di elementi, pari a $2^{|\Omega|}$).

Se invece Ω è infinito, gli assiomi $\{(P1), (P2)\}$ sono strettamente più forti di $\{(P1), (1.3)\}$, cioè esistono funzioni $P : \Omega \rightarrow [0, 1]$ che soddisfano (1.3) ma non (P2). Un esempio con $\Omega = \mathbb{N}$ è descritto nell'Appendice A.2 (la cui lettura può essere omessa, essendo piuttosto sofisticati gli argomenti usati).

Dunque, la σ -addittività *non* è una conseguenza dell'addittività finita. Benché la teoria della probabilità finitamente additiva sia sviluppata in una parte della letteratura matematica, in questo testo richiederemo sempre la σ -addittività, che si adatta assai bene alla maggior parte delle applicazioni e che viene adottata dalla grande maggioranza degli autori. Le ragioni per cui l'assioma (P2) è rilevante rispetto al più debole (1.3) sono diverse, in parte non comprensibili in questa fase iniziale della presentazione della teoria. Tuttavia, una implicazione rilevante della σ -addittività è già descritta nella Proposizione 1.19 più sotto.

Osservazione 1.8. Vale la pena di notare che la proprietà di addittività (1.4) è soddisfatta da nozioni a prima vista distanti dalla probabilità. Si pensi ad esempio all'*area* di una figura piana: se si suddivide la figura in due parti disgiunte, l'area totale coincide con la somma delle aree delle due parti. Un discorso analogo vale per il *volume* di un solido, o per la *quantità di massa* contenuta in una regione dello spazio. Queste (e altre) nozioni condividono un'analogia struttura matematica, la *teoria della misura*, a cui accenneremo nel Capitolo 5.

Per il momento, osserviamo che l'analogia con la nozione di area (o di volume) permette di interpretare “geometricamente” molte proprietà della probabilità, come ad esempio quelle che descriveremo a breve, nel Paragrafo 1.1.4. \square

1.1.3 Probabilità e densità discreta

Una probabilità P su uno spazio campionario Ω è una funzione il cui dominio è l'insieme delle parti $\mathcal{P}(\Omega)$. Mostriamo in questo paragrafo che è possibile *costruire* una probabilità a partire da un oggetto più semplice, detto *densità discreta*, che è una funzione il cui dominio è Ω .

Definizione 1.9 (Densità discreta). Sia Ω un insieme non vuoto. Si dice *densità discreta* su Ω ogni funzione $p : \Omega \rightarrow \mathbb{R}$ che soddisfa le seguenti relazioni:

$$p(\omega) \geq 0 \quad \forall \omega \in \Omega, \quad \sum_{\omega \in \Omega} p(\omega) = 1. \quad (1.5)$$

L'aggettivo “discreta” verrà talvolta omissso. Osserviamo che la somma in (1.5) è sempre ben definita, qualunque sia l'insieme Ω , per quanto visto nel capitolo introduttivo “Nozioni preliminari”, perché la famiglia di numeri reali $(p(\omega))_{\omega \in \Omega}$ è positiva.

Proposizione 1.10. Sia p una densità discreta su un insieme arbitrario Ω . La funzione $P : \mathcal{P}(\Omega) \rightarrow \mathbb{R}$ definita da

$$P(A) := \sum_{\omega \in A} p(\omega), \quad \forall A \subseteq \Omega, \quad (1.6)$$

soddisfa gli assiomi (P1) e (P2) della Definizione 1.5.

Dimostrazione. Chiaramente $P(\Omega) = 1$, grazie alla seconda relazione in (1.5), dunque l'assioma (P1) è verificato. Per quanto riguarda l'assioma (P2), dato che la famiglia di numeri reali $(p(\omega))_{\omega \in \Omega}$ è positiva, possiamo applicare la *somma a blocchi* (0.10): se $(A_k)_{k \in \mathbb{N}}$ è una successione di sottoinsiemi di Ω disgiunti (ossia $A_i \cap A_j = \emptyset$ per $i \neq j$), posto $A := \bigcup_{k \in \mathbb{N}} A_k$, si ha

$$P(A) = \sum_{\omega \in A} p(\omega) = \sum_{k \in \mathbb{N}} \left(\sum_{\omega \in A_k} p(\omega) \right) = \sum_{k \in \mathbb{N}} P(A_k),$$

e la dimostrazione è completata. \square

Quando Ω è finito o numerabile, la Proposizione 1.10 mostra che ogni densità discreta p su Ω determina una *probabilità* P su Ω , mediante la formula (1.6). La cosa interessante è che vale anche il viceversa.

Proposizione 1.11. Sia Ω un insieme (non vuoto) finito o numerabile. C'è una corrispondenza biunivoca tra le probabilità P su Ω e le densità discrete p su Ω , data dalle seguenti relazioni:

$$p(\omega) = P(\{\omega\}), \quad \forall \omega \in \Omega, \quad (1.7)$$

$$P(A) = \sum_{\omega \in A} p(\omega), \quad \forall A \subseteq \Omega. \quad (1.8)$$

Dimostrazione. Se p è una densità discreta su Ω , abbiamo già mostrato nella Proposizione 1.11 che la relazione (1.8) definisce una probabilità P su Ω .

Viceversa, sia P una probabilità su Ω . Per ogni $A \in \mathcal{P}(\Omega)$ possiamo scrivere $A = \bigcup_{\omega \in A} \{\omega\}$, dove l'unione è disgiunta e *al più numerabile*, dato che Ω è finito o numerabile. Usando l'assioma (P2) otteniamo dunque

$$P(A) = \sum_{\omega \in A} P(\{\omega\}) = \sum_{\omega \in A} p(\omega),$$

dove la funzione $p : \Omega \rightarrow \mathbb{R}$ è *definita* da (1.7). Dunque la relazione (1.8) è verificata. Resta solo da mostrare che p è una densità discreta. Chiaramente $p(\omega) = P(\{\omega\}) \geq 0$ per ogni ω , essendo positiva la probabilità di qualunque evento. Scegliendo $A = \Omega$ in (1.8) si vede che anche la seconda relazione in (1.5) è soddisfatta. \square

Osservazione 1.12. La Proposizione 1.11 mostra che in uno spazio di probabilità discreto la probabilità è determinata dal suo valore sui *singoletti* $\{\omega\}$, ossia sugli eventi costituiti da un solo elemento di Ω . Questo non è più vero per gli spazi di probabilità generali, come vedremo nel Capitolo 5. \square

Mostriamo ora alcuni esempi di spazi di probabilità discreti.

Esempio 1.13 (Probabilità uniforme). Sia Ω un insieme *finito*. Definiamo per $A \subseteq \Omega$

$$P(A) := \frac{|A|}{|\Omega|},$$

dove $|\cdot|$ indica il numero di elementi di un insieme. Si vede facilmente che P soddisfa gli assiomi (P1) e (P2) della Definizione 1.5 (invece di (P2) è sufficiente verificare (1.4), essendo Ω finito). Pertanto P è una probabilità, detta *uniforme su Ω* . La densità associata, grazie a (1.7), è data da

$$p(\omega) = \frac{1}{|\Omega|},$$

e dunque *non dipende da ω* . È interessante notare che vale anche il viceversa: se $p(\omega) = P(\{\omega\}) = c$ non dipende da $\omega \in \Omega$, allora necessariamente $c = 1/|\Omega|$ e dunque P è la probabilità uniforme su Ω (esercizio).

Lo spazio (Ω, P) così definito si dice *spazio di probabilità uniforme*. Esso è il modello probabilistico adeguato a descrivere gli esperimenti aleatori in cui tutti gli esiti si possono ritenere equiprobabili (per esempio, per ragioni di simmetria). Casi tipici sono il lancio di un dado regolare, l'estrazione di un numero in una ruota del lotto, la successione delle carte in un mazzo accuratamente mescolato... Ritourneremo su questi esempi nel Paragrafo 1.2. \square

Esempio 1.14 (Misure di Gibbs). Sia Ω un insieme *finito*, e sia $H : \Omega \rightarrow \mathbb{R}$ una funzione arbitraria fissata. Per ogni parametro reale $\beta \geq 0$ fissato, definiamo la funzione $p_\beta : \Omega \rightarrow [0, 1]$ ponendo

$$p_\beta(\omega) := \frac{1}{Z_\beta} e^{-\beta H(\omega)}, \quad (1.9)$$

dove

$$Z_\beta := \sum_{\omega \in \Omega} e^{-\beta H(\omega)}. \quad (1.10)$$

Le relazioni in (1.5) sono verificate, pertanto la funzione p_β determina, mediante (1.8), una probabilità che indicheremo con P_β .

Prendendo a prestito la terminologia della meccanica statistica, la probabilità P_β è detta *misura di Gibbs* relativa all'energia H e alla temperatura inversa β . L'interpretazione è che gli elementi $\omega \in \Omega$ rappresentano gli stati di un sistema fisico, a cui è associata una energia $H(\omega)$; quando il sistema è in equilibrio termico alla temperatura assoluta T , ponendo $\beta = \frac{1}{k_B T}$ (dove k_B è la costante di Boltzmann), la probabilità di osservare il sistema in uno stato ω è data da $p_\beta(\omega) = P_\beta(\{\omega\})$.

Si noti che, nel caso $\beta = 0$ (temperatura infinita), $p_0(\omega) = 1/Z_0$ non dipende da ω , pertanto P_0 non è altro che la probabilità uniforme su Ω . Consideriamo invece il limite $\beta \rightarrow +\infty$, che corrisponde al limite di temperatura zero (assoluto). Indichiamo con $m := \min\{H(\omega) : \omega \in \Omega\}$ il minimo assoluto della Hamiltoniana, e introduciamo l'insieme (non vuoto)

$$M := \{\omega \in \Omega : H(\omega) = m\},$$

costituito dagli elementi di Ω con energia minima. Mostriamo ora che

$$\lim_{\beta \rightarrow +\infty} P_\beta(M) = 1.$$

In altre parole, nel limite $\beta \rightarrow +\infty$, la probabilità P_β “si concentra” sugli elementi di minima energia. A tal fine, è sufficiente (perché?) mostrare che, per ogni $\omega \notin M$,

$$\lim_{\beta \rightarrow +\infty} P_\beta(\{\omega\}) = 0. \quad (1.11)$$

Grazie alla relazione (1.10) possiamo scrivere

$$Z_\beta \geq \sum_{\omega \in M} e^{-\beta H(\omega)} = |M| e^{-\beta m} \geq e^{-\beta m},$$

pertanto

$$P_\beta(\{\omega\}) = p_\beta(\omega) = \frac{e^{-\beta H(\omega)}}{Z_\beta} \leq \frac{e^{-\beta H(\omega)}}{e^{-\beta m}} = e^{-\beta[H(\omega)-m]}. \quad (1.12)$$

Per ogni $\omega \notin A$ si ha $H(\omega) > m$, quindi prendendo il limite $\beta \rightarrow +\infty$ in (1.12) si ottiene (1.11). (Si veda l'Esercizio 1.5 per un rafforzamento.) \square

Esempio 1.15 (Distribuzione di Poisson). Sia $\Omega = \mathbb{N}_0 = \{0, 1, 2, \dots\}$ e poniamo

$$p(n) := e^{-\lambda} \frac{\lambda^n}{n!},$$

dove $\lambda > 0$ è un parametro reale fissato. Ricordando la serie esponenziale (0.5), la relazione (1.5) è verificata e dunque è possibile definire una probabilità P tramite (1.8). Tale probabilità, detta di Poisson, compare in modo naturale in molti modelli probabilistici. Ne studieremo le proprietà nel Paragrafo 3.5.5. \square

1.1.4 Proprietà fondamentali

Iniziamo con l'esporre alcune conseguenze semplici, ma molto importanti, degli assiomi (P1) e (P2). Può essere utile tenere a mente l'analogia formale tra la probabilità e la nozione di area, a cui abbiamo accennato nell'Osservazione 1.8.

Proposizione 1.16. *Sia (Ω, P) uno spazio di probabilità discreto.*

(i) *Per ogni $A, B \subseteq \Omega$ tali che $A \subseteq B$*

$$P(B \setminus A) = P(B) - P(A),$$

di conseguenza

$$P(A) \leq P(B).$$

In particolare, per ogni $A \subseteq \Omega$

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

(ii) *Per ogni $A, B \subseteq \Omega$*

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

di conseguenza

$$P(A \cup B) \leq P(A) + P(B).$$

Dimostrazione. Cominciamo dal punto (i). Se $A \subseteq B$ si può scrivere $B = A \cup (B \setminus A)$ e l'unione è disgiunta, quindi per l'additività di P

$$P(B) = P(A \cup (B \setminus A)) = P(A) + P(B \setminus A),$$

ossia $P(B \setminus A) = P(B) - P(A)$. Dato che la probabilità di qualsiasi evento è un numero positivo, segue che $P(A) \leq P(B)$.

Veniamo dunque al punto (ii). Per ogni scelta di $A, B \subseteq \Omega$ si ha $B \setminus A = B \setminus (A \cap B)$. Dato che $(A \cap B) \subseteq B$, per il punto precedente $P(B \setminus A) = P(B) - P(A \cap B)$. Infine, notando che $A \cup B = A \cup (B \setminus A)$ e l'unione è disgiunta, per l'additività di P si ottiene

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

e la dimostrazione è conclusa. \square

Osservazione 1.17 (Principio dell'“almeno uno”). La formula

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

si rivela spesso molto utile. A titolo di esempio, calcoliamo la probabilità di ottenere almeno un 6 lanciando 8 dadi. Una scelta naturale di spazio campionario è

$$\Omega = \{1, 2, 3, 4, 5, 6\}^8,$$

ossia l'insieme dei vettori $\omega = (\omega_1, \dots, \omega_8)$ con componenti $\omega_i \in \{1, 2, 3, 4, 5, 6\}$. L'evento che ci interessa è dato dal sottoinsieme

$$A := \{\omega = (\omega_1, \dots, \omega_8) \in \Omega : \omega_i = 6 \text{ per qualche } 1 \leq i \leq 8\}.$$

Se i dadi sono regolari, è naturale munire Ω della probabilità uniforme P , pertanto $P(A) = |A|/|\Omega|$. Dovrebbe essere intuitivamente chiaro che $|\Omega| = 6^8$, perché ogni $\omega \in \Omega$ è determinato da 8 componenti ω_i , ciascuna delle quali può assumere 6 valori (approfondiremo i problemi di conteggio nel Paragrafo 1.2), ma non è ovvio come calcolare la cardinalità di A . Il problema si risolve facilmente notando che

$$A^c = \{\omega = (\omega_1, \dots, \omega_8) \in \Omega : \omega_i \neq 6 \text{ per ogni } 1 \leq i \leq 8\} = \{1, 2, 3, 4, 5\}^8,$$

da cui segue che $|A^c| = 5^8$. Pertanto

$$P(A) = 1 - P(A^c) = 1 - \frac{|A^c|}{|\Omega|} = 1 - \left(\frac{5}{6}\right)^8 \simeq 0.77.$$

(Vedremo un'altra applicazione, più interessante, nell'Esempio 1.33.) \square

L'identità della parte (ii) della Proposizione 1.16 può essere generalizzata all'unione di più di due eventi. Ad esempio, supponiamo di voler calcolare $P(A \cup B \cup C)$ per tre eventi A, B, C . Usando due volte l'identità appena citata si ottiene

$$\begin{aligned} P(A \cup B \cup C) &= P((A \cup B) \cup C) = P(A \cup B) + P(C) - P((A \cup B) \cap C) \\ &= P(A) + P(B) - P(A \cap B) + P(C) - P((A \cap C) \cup (B \cap C)) \\ &= P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C). \end{aligned}$$

Non è difficile, a questo punto, “indovinare” la formula generale per l'unione di un numero finito arbitrario di eventi. Il seguente risultato è chiamato *formula di inclusione-esclusione*. Ne vedremo un'applicazione interessante nel prossimo capitolo, nel Problema 2.4 del Paragrafo 2.1.

Proposizione 1.18 (Formula di inclusione-esclusione). *Sia (Ω, P) uno spazio di probabilità discreto. Per ogni scelta degli eventi $A_1, A_2, \dots, A_n \subseteq \Omega$*

$$P(A_1 \cup A_2 \cup \dots \cup A_n) = \sum_{k=1}^n \sum_{\substack{J \subseteq \{1, 2, \dots, n\} \\ \text{tali che } |J|=k}} (-1)^{k+1} P\left(\bigcap_{i \in J} A_i\right). \quad (1.13)$$

Dimostrazione. Dimostriamo per induzione su n che la relazione (1.13) è vera per ogni n -pla di eventi A_1, A_2, \dots, A_n . Per $n = 1$ la formula (1.13) si riduce a $P(A_1) = P(A_1)$, dunque non c'è nulla da dimostrare, mentre per $n = 2$ tale formula coincide con $P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2)$, che è già stata dimostrata nella Proposizione 1.16. Procediamo dunque per induzione: per $n \geq 3$, supponiamo che l'asserto sia vero per ogni $k \leq n$ e mostriamo che è vero per $n + 1$. Siano $A_1, A_2, \dots, A_n, A_{n+1}$ eventi. Per ipotesi induttiva, (1.13) vale per $n = 2$,

dunque

$$\begin{aligned}
 P(A_1 \cup A_2 \cup \dots \cup A_n \cup A_{n+1}) \\
 &= P(A_1 \cup A_2 \cup \dots \cup A_n) + P(A_{n+1}) - P((A_1 \cup A_2 \cup \dots \cup A_n) \cap A_{n+1}) \\
 &= P(A_1 \cup A_2 \cup \dots \cup A_n) + P(A_{n+1}) - P(B_1 \cup B_2 \cup \dots \cup B_n),
 \end{aligned} \tag{1.14}$$

dove abbiamo posto per comodità $B_i = A_i \cap A_{n+1}$, per $i = 1, 2, \dots, n$. Usando nuovamente l'ipotesi induttiva, questa volta per n eventi, otteniamo

$$\begin{aligned}
 P(A_1 \cup A_2 \cup \dots \cup A_n) &= \sum_{k=1}^n \sum_{\substack{J \subseteq \{1, 2, \dots, n\} \\ \text{tale che } |J|=k}} (-1)^{k+1} P\left(\bigcap_{i \in J} A_i\right) \\
 &= \sum_{k=1}^n \sum_{\substack{J \subseteq \{1, 2, \dots, n+1\} \\ \text{tale che } |J|=k \text{ e } n+1 \notin J}} (-1)^{k+1} P\left(\bigcap_{i \in J} A_i\right),
 \end{aligned} \tag{1.15}$$

e analogamente

$$\begin{aligned}
 P(B_1 \cup B_2 \cup \dots \cup B_n) &= \sum_{k=1}^n \sum_{\substack{J \subseteq \{1, 2, \dots, n\} \\ \text{tale che } |J|=k}} (-1)^{k+1} P\left(\bigcap_{i \in J} B_i\right) \\
 &= \sum_{k=1}^n \sum_{\substack{J \subseteq \{1, 2, \dots, n\} \\ \text{tale che } |J|=k}} (-1)^{k+1} P\left(A_{n+1} \cap \left(\bigcap_{i \in J} A_i\right)\right) \\
 &= - \sum_{k=2}^{n+1} \sum_{\substack{J \subseteq \{1, 2, \dots, n+1\} \\ \text{tale che } |J|=k \text{ e } n+1 \in J}} (-1)^{k+1} P\left(\bigcap_{i \in J} A_i\right).
 \end{aligned} \tag{1.16}$$

Sostituendo (1.15) e (1.16) nell'ultimo membro di (1.14), si ottiene

$$P(A_1 \cup A_2 \cup \dots \cup A_n \cup A_{n+1}) = \sum_{k=1}^{n+1} \sum_{\substack{J \subseteq \{1, 2, \dots, n+1\} \\ \text{tale che } |J|=k}} (-1)^{k+1} P\left(\bigcap_{i \in J} A_i\right),$$

che è quanto si voleva dimostrare. \square

Va notato come le dimostrazioni dei risultati delle Proposizioni 1.16 e 1.18 usino solo l'additività finita (1.3) e non la σ -additività (P2), che invece gioca un ruolo fondamentale nella seguente.

Proposizione 1.19 (Continuità dal basso e dall'alto della probabilità). *Sia Ω un insieme e sia $P : \mathcal{P}(\Omega) \rightarrow [0, 1]$ una funzione che soddisfa (P1) e l'additività in (1.3). Le seguenti proprietà sono equivalenti:*

- (i) σ -additività: P soddisfa la proprietà (P2) della Definizione 1.5, e dunque (Ω, P) è uno spazio di probabilità discreto.

(ii) *Continuità dal basso*: se $(A_n)_{n \in \mathbb{N}}$ è una successione crescente di eventi, cioè $A_n \subseteq A_{n+1} \subseteq \Omega$ per ogni $n \in \mathbb{N}$, allora

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = \lim_{n \rightarrow +\infty} P(A_n).$$

(iii) *Continuità dall'alto*: se $(A_n)_{n \in \mathbb{N}}$ è una successione decrescente di eventi, cioè $A_{n+1} \subseteq A_n \subseteq \Omega$ per ogni $n \in \mathbb{N}$, allora

$$P\left(\bigcap_{n=1}^{\infty} A_n\right) = \lim_{n \rightarrow +\infty} P(A_n).$$

Dimostrazione. (i) \Rightarrow (ii). Per una data successione crescente $(A_n)_{n \in \mathbb{N}}$ di eventi, definiamo un'altra successione $(B_n)_{n \in \mathbb{N}}$ tramite $B_1 := A_1$, e $B_n := A_n \setminus A_{n-1}$ per $n \geq 2$. Per costruzione, gli eventi B_n sono disgiunti e, per ogni $n \in \mathbb{N}$,

$$\bigcup_{k=1}^n B_k = A_n.$$

Inoltre

$$\bigcup_{n=1}^{\infty} B_n = \bigcup_{n=1}^{\infty} A_n.$$

Allora, per la σ -additività,

$$\begin{aligned} P\left(\bigcup_{n=1}^{\infty} A_n\right) &= P\left(\bigcup_{n=1}^{\infty} B_n\right) = \sum_{n=1}^{\infty} P(B_n) = \lim_{n \rightarrow +\infty} \sum_{k=1}^n P(B_k) \\ &= \lim_{n \rightarrow +\infty} P\left(\bigcup_{k=1}^n B_k\right) = \lim_{n \rightarrow +\infty} P(A_n). \end{aligned}$$

(ii) \Rightarrow (i). Sia (A_n) una successione di eventi disgiunti. La successione $(B_n)_{n \in \mathbb{N}}$ definita da $B_n := \bigcup_{k=1}^n A_k$ è crescente, pertanto usando l'additività finita e (ii) si ha:

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = P\left(\bigcup_{n=1}^{\infty} B_n\right) = \lim_{n \rightarrow +\infty} P(B_n) = \lim_{n \rightarrow +\infty} \sum_{k=1}^n P(A_k) = \sum_{n=1}^{+\infty} P(A_n).$$

(ii) \Rightarrow (iii). Sia $(A_n)_{n \in \mathbb{N}}$ una successione decrescente di eventi. Posto $B_n := A_n^c$, $(B_n)_{n \in \mathbb{N}}$ è una successione crescente di eventi. Allora, usando (ii), si ha

$$\begin{aligned} P\left(\bigcap_{n=1}^{\infty} A_n\right) &= P\left(\left(\bigcup_{n=1}^{\infty} B_n\right)^c\right) = 1 - P\left(\bigcup_{n=1}^{\infty} B_n\right) \\ &= 1 - \lim_{n \rightarrow +\infty} P(B_n) = \lim_{n \rightarrow +\infty} P(A_n). \end{aligned}$$

(iii) \Rightarrow (ii). Simile all'implicazione precedente. Si lasciano i dettagli al lettore. \square

Un utile corollario è il seguente.

Corollario 1.20 (Subadditività). Sia $(A_n)_{n \in \mathbb{N}}$ una successione di eventi. Allora

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) \leq \sum_{n=1}^{\infty} P(A_n).$$

Dimostrazione. Sia $B_n := \bigcup_{k=1}^n A_k$. Evidentemente $(B_n)_{n \in \mathbb{N}}$ è una successione crescente di eventi. Inoltre $\bigcup_{n \in \mathbb{N}} B_n = \bigcup_{n \in \mathbb{N}} A_n$. Per la parte (ii) della Proposizione 1.16, sappiamo che $P(A_1 \cup A_2) \leq P(A_1) + P(A_2)$. Con una facile dimostrazione per induzione, la precedente disuguaglianza si estende a:

$$P\left(\bigcup_{k=1}^n A_k\right) \leq \sum_{k=1}^n P(A_k).$$

Ma allora, usando anche la Proposizione 1.19, si ha

$$\begin{aligned} P\left(\bigcup_{n=1}^{\infty} A_n\right) &= P\left(\bigcup_{n=1}^{\infty} B_n\right) = \lim_{n \rightarrow +\infty} P(B_n) = \lim_{n \rightarrow +\infty} P\left(\bigcup_{k=1}^n A_k\right) \\ &\leq \lim_{n \rightarrow +\infty} \sum_{k=1}^n P(A_k) = \sum_{n=1}^{+\infty} P(A_n), \end{aligned}$$

che è la relazione cercata. \square

Esercizi

Esercizio 1.1. Sia (Ω, P) uno spazio di probabilità discreto e siano $A, B \subseteq \Omega$ eventi.

- (i) Si mostri che se $P(A) = P(B) = 0$ allora $P(A \cup B) = 0$.
- (ii) Si mostri che se $P(A) = P(B) = 1$ allora $P(A \cap B) = 1$.

Esercizio 1.2. Rafforziamo l'esercizio precedente. Sia (Ω, P) uno spazio di probabilità discreto e sia $(A_n)_{n \in \mathbb{N}}$ una famiglia *numerabile* di eventi.

- (i) Si mostri che se $P(A_n) = 0$ per ogni $n \in \mathbb{N}$, allora $P(\bigcup_{n \in \mathbb{N}} A_n) = 0$.
- (ii) Si mostri che se $P(A_n) = 1$ per ogni $n \in \mathbb{N}$, allora $P(\bigcap_{n \in \mathbb{N}} A_n) = 1$.

Esercizio 1.3. Sia (Ω, P) uno spazio di probabilità discreto e sia $C \subseteq \Omega$ un evento.

- (i) Si mostri che, se $P(C) = 1$, allora $P(A \cap C) = P(A)$ per ogni $A \subseteq \Omega$.
- (ii) Si mostri che, se $P(C) = 0$, allora $P(A \cup C) = P(A)$ per ogni $A \subseteq \Omega$.

Esercizio 1.4 (Disuguaglianza di Bonferroni). Siano A_1, \dots, A_n eventi di uno spazio di probabilità discreto (Ω, P) . Si mostri per induzione su $n \in \mathbb{N}$ che

$$P\left(\bigcup_{k=1}^n A_k\right) \geq \sum_{k=1}^n P(A_k) - \sum_{1 \leq i < j \leq n} P(A_i \cap A_j).$$

In particolare, se $P(A_i \cap A_j) = 0$ per $i \neq j$, si deduca che $P(\bigcup_{k=1}^n A_k) = \sum_{k=1}^n P(A_k)$.

Esercizio 1.5. Con le stesse notazioni dell'Esempio 1.14, si mostri che

$$\lim_{\beta \rightarrow +\infty} P_\beta(\{\omega\}) = \frac{1}{|M|}, \quad \forall \omega \in M. \quad (1.17)$$

Possiamo dire che, per $\beta \rightarrow +\infty$, la probabilità P_β “converge” (nel senso della relazione (1.17)) verso la probabilità uniforme sull'insieme M .

1.2 Calcolo combinatorio

Ricordiamo dall'Esempio 1.13 che uno spazio di probabilità discreto (Ω, P) si dice *uniforme* se Ω è un insieme finito e si ha $P(A) = \frac{|A|}{|\Omega|}$, per ogni $A \subseteq \Omega$. Pertanto, il calcolo della probabilità di un evento in uno spazio uniforme si riduce a contarne il numero di elementi. I problemi di conteggio sono tipicamente non banali e vanno affrontati con attenzione. Le tecniche rilevanti in questo contesto, che ora descriviamo, costituiscono quello che è detto *calcolo combinatorio*.

1.2.1 Principi basilari

Dati due insiemi A, B , si dice che A è in *corrispondenza biunivoca* con B se esiste un'applicazione $f: A \rightarrow B$ biunivoca, cioè iniettiva e suriettiva. Chiaramente A è in corrispondenza biunivoca con B se e soltanto se B è in corrispondenza biunivoca con A : si scrive “ A e B sono in corrispondenza biunivoca”, che rende palese la simmetria della relazione (si tratta in effetti di una relazione di equivalenza). Dato $n \in \mathbb{N}$, si dice che un insieme A ha *cardinalità* n e si scrive $|A| = n$ se A è in corrispondenza biunivoca con l'insieme $\{1, 2, \dots, n\}$ – ossia, intuitivamente, se “ A ha n elementi”. In questo paragrafo considereremo solo insiemi finiti, cioè insiemi che hanno cardinalità n per un opportuno $n \in \mathbb{N}$.

Per determinare la cardinalità di un insieme, la strategia tipica consiste nel ricondurre il calcolo all'applicazione combinata (talvolta non banale) di alcuni principi o osservazioni basilari. Una prima osservazione, elementare ma molto utile, è che se un insieme A è in corrispondenza biunivoca con un insieme B , allora $|A| = |B|$. Un'altra osservazione, anch'essa molto intuitiva, è la seguente: se A, B sono due sottoinsiemi (di uno stesso spazio) *disgiunti*, cioè tali che $A \cap B = \emptyset$, allora $|A \cup B| = |A| + |B|$. Più in generale, se A_1, \dots, A_k sono sottoinsiemi disgiunti, tali cioè che $A_i \cap A_j = \emptyset$ per $i \neq j$, allora $|\bigcup_{i=1}^k A_i| = \sum_{i=1}^k |A_i|$. La dimostrazione di queste osservazioni è semplice ed è lasciata per esercizio.

Un principio leggermente meno elementare riguarda la cardinalità degli *insiemi prodotto*. Ricordiamo che, dati due insiemi A, B , il loro prodotto cartesiano $A \times B$ è definito come l'insieme delle coppie ordinate (a, b) , con $a \in A$ e $b \in B$. Vale allora la relazione $|A \times B| = |A| \cdot |B|$. Il modo più semplice per convincersi della validità di questa formula consiste nel disporre gli elementi di $A \times B$ in una tabella rettangolare, dopo aver numerato gli elementi dei due insiemi. Più precisamente, se $A =$

$\{a_1, a_2, \dots, a_m\}$, $B = \{b_1, b_2, \dots, b_k\}$, possiamo elencare gli elementi dell'insieme $A \times B$ nel modo seguente:

$$\begin{array}{ccccccc} (a_1, b_1) & (a_1, b_2) & \cdots & (a_1, b_{k-1}) & (a_1, b_k) & & \\ (a_2, b_1) & (a_2, b_2) & \cdots & (a_2, b_{k-1}) & (a_2, b_k) & & \\ \vdots & \vdots & \ddots & \vdots & \vdots & & \\ (a_m, b_1) & (a_m, b_2) & \cdots & (a_m, b_{k-1}) & (a_m, b_k) & & \end{array},$$

da cui è chiaro che $|A \times B| = m \cdot k = |A| \cdot |B|$.

Diamo ora una dimostrazione più formale. Per $x \in A$ indichiamo con $\{x\} \times B$ il sottoinsieme di $A \times B$ costituito dagli elementi che hanno x come prima componente, cioè $\{x\} \times B := \{(x, b) : b \in B\}$. Possiamo quindi scrivere $A \times B = \bigcup_{x \in A} (\{x\} \times B)$, e si noti che questa unione è disgiunta, ossia $(\{x_1\} \times B) \cap (\{x_2\} \times B) = \emptyset$ se $x_1 \neq x_2$. Per l'osservazione enunciata sopra si ha dunque $|A \times B| = \sum_{x \in A} |\{x\} \times B|$. Dato che l'insieme $\{x\} \times B$ è in corrispondenza biunivoca con B , mediante l'applicazione $(x, b) \mapsto b$, segue che $|\{x\} \times B| = |B|$ e dunque $|A \times B| = \sum_{x \in A} |B| = |A| |B|$.

Per induzione, si estende facilmente la formula al caso di più di due fattori: dati gli insiemi A_1, \dots, A_k , con $k \in \mathbb{N}$, l'insieme prodotto $A_1 \times \cdots \times A_k$ è definito come l'insieme delle k -uple (a_1, \dots, a_k) , con $a_i \in A_i$, e ha cardinalità

$$|A_1 \times \cdots \times A_k| = |A_1| \cdots |A_k| = \prod_{i=1}^k |A_i|. \quad (1.18)$$

Un'estensione di questa formula, elementare ma non banale, conduce a quello che è noto come il principio fondamentale del calcolo combinatorio. Prima di vedere di che cosa si tratta, discutiamo qualche applicazione dei principi appena visti.

1.2.2 Disposizioni con ripetizione

Definizione 1.21. Dati $k \in \mathbb{N}$ e un insieme finito A , si dicono *disposizioni con ripetizione di k elementi estratti da A* le funzioni $f : \{1, \dots, k\} \rightarrow A$.

Le disposizioni con ripetizione sono in corrispondenza biunivoca naturale con gli elementi dell'insieme prodotto $A^k := A \times \cdots \times A$ (k volte).[†] Quindi una disposizione con ripetizione può essere vista come una *sequenza ordinata* (x_1, \dots, x_k) di elementi di A , *non necessariamente distinti*: si può cioè avere $x_i = x_j$ per $i \neq j$. Sottolineiamo che l'ordine in cui compaiono gli elementi è importante: per esempio, $(1, 3)$ e $(3, 1)$ sono due disposizioni differenti.

Per la formula (1.18) sulla cardinalità degli insiemi prodotto, si ha che $|A^k| = |A|^k = n^k$. Abbiamo pertanto il seguente risultato.

[†] Basta associare a $(x_1, \dots, x_k) \in A^k$ la funzione $f : \{1, \dots, k\} \rightarrow A$ definita da $f(i) := x_i$.

Proposizione 1.22. *Le disposizioni con ripetizione di k elementi estratti da un insieme di n elementi sono in numero n^k .*

- Esempio 1.23.* (i) I compleanni di un gruppo ordinato di 4 persone costituiscono una disposizione con ripetizione di 4 elementi estratti dall'insieme dei giorni dell'anno, che ha cardinalità 366 (contando il 29 febbraio). Sono dunque possibili $366^4 \approx 1.8 \cdot 10^{10}$ sequenze distinte di compleanni.
- (ii) Per compilare una colonna di una schedina del Totocalcio occorre scegliere, per ciascuna delle 13 partite in esame, tra la vittoria della squadra di casa (1), il pareggio (x) o la vittoria della squadra in trasferta (2). Una colonna compilata è dunque una disposizione con ripetizione di 13 elementi estratti dall'insieme $\{1, x, 2\}$ e di conseguenza ci sono $3^{13} \approx 1.6 \cdot 10^6$ modi possibili di compilare una colonna.
- (iii) Le possibili “parole” (anche prive di significato) costituite da 6 lettere dell'alfabeto inglese possono essere identificate con le disposizioni con ripetizione di 6 elementi estratti da un insieme – le lettere dell'alfabeto – che ne contiene 26: il loro numero è dunque pari a $26^6 \approx 3.1 \cdot 10^8$. Le parole che effettivamente hanno un significato (per esempio nella lingua inglese) sono naturalmente molte meno: anche includendo i termini tecnici, non si arriva a trecentomila parole. Di conseguenza, la probabilità che digitando una sequenza di sei lettere a caso si ottenga una parola di senso compiuto è certamente (molto!) minore di $3 \cdot 10^5 / (3.1 \cdot 10^8) < 10^{-3}$, ossia una probabilità su mille. \square

Osservazione 1.24. Dati due insiemi A, B , indichiamo con A^B l'insieme di tutte le funzioni $f: B \rightarrow A$. Se l'insieme $B = \{b_1, \dots, b_k\}$ ha cardinalità $k \in \mathbb{N}$, è facile vedere che A^B è in corrispondenza biunivoca con A^k (o, equivalentemente, con le disposizioni con ripetizione di k elementi estratti da A). Una corrispondenza biunivoca tra A^k e A^B è per esempio quella che a $(x_1, \dots, x_k) \in A^k$ associa la funzione $f \in A^B$ definita da $f(b_i) := x_i$. Segue dunque che $|A^B| = |A|^k$, cioè $|A^B| = |A|^{|B|}$. \square

1.2.3 Il principio fondamentale

Un esempio molto ricorrente nelle applicazioni è quello in cui gli elementi di un insieme possano essere determinati attraverso *scelte successive*.

Esempio 1.25. Sia E l'insieme delle funzioni *iniettive* da $\{1, \dots, k\}$ in un insieme A con $|A| = n$ (si noti che necessariamente $k \leq n$). Possiamo determinare ogni funzione $f \in E$ scegliendo innanzitutto la prima componente $f(1)$ come un elemento qualunque di A , quindi scegliendo la seconda componente $f(2)$ come un elemento qualunque di $A \setminus \{f(1)\}$, e così via. Abbiamo n esiti possibili per la scelta di $f(1)$, $(n-1)$ esiti possibili per la scelta di $f(2)$, \dots , $(n-(k-1)) = (n-k+1)$ esiti possibili per la scelta di $f(k)$. Per analogia con gli insiemi prodotto, dovrebbe essere chiaro che $|E| = n \cdot (n-1) \cdots (n-k+1)$. \square

Nell'esempio appena descritto, l'*insieme degli esiti* della scelta i -esima dipende dagli esiti delle scelte precedenti, tuttavia *il numero di esiti possibili* è sempre lo stesso, pari a $n - i + 1$. Generalizzando queste considerazioni, giungiamo al *principio fondamentale del calcolo combinatorio*, che possiamo formulare come segue.

Teorema 1.26 (Principio fondamentale del calcolo combinatorio). *Supponiamo che gli elementi di un insieme E possano essere determinati mediante k scelte successive, in cui ogni scelta abbia un numero fissato di esiti possibili: la prima scelta ha n_1 esiti possibili, la seconda scelta ne ha n_2 , ..., la k -esima scelta ne ha n_k , dove $n_1, \dots, n_k \in \mathbb{N}$. Se sequenze distinte di esiti delle scelte determinano elementi distinti di E (ossia, se non si può ottenere uno stesso elemento di E con due sequenze diverse di esiti delle scelte), allora $|E| = n_1 \cdot n_2 \cdots n_k$.*

Così enunciato, questo principio è un po' vago (per esempio, il concetto di "scelta" non è stato definito precisamente). Una riformulazione matematicamente precisa del Teorema 1.26, con la relativa dimostrazione, è fornita nell'Appendice A.3. Questa comporta tuttavia notazioni piuttosto pesanti e risulta di poco aiuto per l'applicazione del principio a casi concreti.

Nella pratica, si fa tipicamente riferimento all'enunciato del Teorema 1.26. L'idea cruciale è che gli elementi dell'insieme E possono essere messi in *corrispondenza biunivoca* con le sequenze di esiti delle scelte, che hanno una struttura di spazio prodotto, da cui segue la formula per la cardinalità. La condizione che sequenze distinte di esiti determinino elementi distinti di E serve proprio a garantire che la corrispondenza sia biunivoca: la mancata verifica di questa condizione è la principale fonte di errori nell'applicazione del principio. Qualche esempio chiarirà la situazione.

Esempio 1.27. (i) Un mazzo di carte da poker è costituito da 52 carte, identificate dal *seme* (cuori ♡, quadri ♦, fiori ♣, picche ♠) e dal *tipo* (un numero da 1 a 10 oppure J, Q, K). Indichiamo con E l'insieme delle carte di numero pari (figure escluse) e di colore rosso (cioè di cuori o di quadri). Ogni elemento di E può essere determinato attraverso due scelte successive: la scelta del seme, che ha 2 esiti possibili (cuori e quadri), e la scelta del tipo, che ne ha 5 (cioè 2, 4, 6, 8, 10). Segue dunque che $|E| = 2 \cdot 5 = 10$.

(ii) Dato un mazzo di carte da poker, si chiama *full* un sottoinsieme di 5 carte costituito dall'unione di un *tris* (un sottoinsieme di 3 carte dello stesso tipo) e di una *coppia* (un sottoinsieme di 2 carte dello stesso tipo). Indichiamo con E l'insieme dei possibili full. Sottolineiamo che gli elementi di E sono *sottoinsiemi* di 5 carte, non disposizioni: in particolare, le carte non sono ordinate.

Gli elementi di E possono essere determinati univocamente attraverso 4 scelte successive: 1) il tipo del tris; 2) il tipo della coppia; 3) i semi delle carte che compaiono nel tris; 4) i semi delle carte che compaiono nella coppia. Per la prima scelta ci sono 13 esiti possibili, per la seconda scelta, qualunque sia l'esito della prima scelta, ci sono 12 esiti possibili (chiaramente i due tipi devono essere differenti, perché non esistono cinque carte dello stesso tipo). Per la terza

scelta, occorre scegliere tre semi nell'insieme $\{\heartsuit, \diamondsuit, \clubsuit, \spadesuit\}$: per enumerazione diretta, è facile vedere che ci sono 4 esiti possibili; analogamente, per la quarta scelta occorre scegliere due semi e per questo ci sono 6 esiti possibili (ritorneremo nell'Esempio 1.37 sul modo di contare i sottoinsiemi). Applicando il Teorema 1.26 si ottiene dunque che $|E| = 13 \cdot 12 \cdot 4 \cdot 6 = 3744$.

- (iii) Dato un mazzo di carte da poker, indichiamo con E l'insieme delle *doppie coppie*, cioè i sottoinsiemi di 5 carte costituiti dall'unione di due *coppie* di tipi diversi, più una quinta carta di tipo diverso dai tipi delle due coppie.

Per determinare $|E|$ si potrebbe essere tentati di procedere analogamente al caso dei full, attraverso sei scelte successive: 1) il tipo della prima coppia; 2) il tipo della seconda coppia; 3) il tipo della “quinta carta”; 4) i semi delle carte che compaiono nella prima coppia; 5) i semi delle carte che compaiono nella seconda coppia; 6) il seme della “quinta carta”. Ci sono 13 esiti possibili per la prima scelta, 12 per la seconda scelta, 11 per la terza, 6 per la quarta, 6 per la quinta, 4 per la sesta: si otterrebbe dunque $|E| = 13 \cdot 12 \cdot 11 \cdot 6^2 \cdot 4 = 247\,104$. Tuttavia questo risultato è errato.

La ragione è che le scelte 1) e 2) sono ambigue, dal momento che non esiste una “prima” e una “seconda” coppia. In effetti, sequenze distinte di esiti delle sei scelte sopra elencate non conducono a elementi distinti di E : ciascun elemento di E , cioè ciascuna doppia coppia, viene infatti selezionata *esattamente due volte*. Per esempio, la doppia coppia $\{5\heartsuit, 5\diamondsuit, 6\heartsuit, 6\clubsuit, 7\spadesuit\}$ viene determinata sia con l'esito “5” della scelta 1) e l'esito “6” della scelta 2), sia viceversa. Per tale ragione, il risultato corretto è $|E| = 247\,104/2 = 123\,552$, cioè la metà di quanto ottenuto in precedenza.

Un modo alternativo di ottenere il risultato corretto è di riunire le scelte 1) e 2) nell'unica scelta 1bis) “i tipi delle due coppie”, che ha $13 \cdot 12/2 = 78$ esiti possibili (anche su questo torneremo nell'Esempio 1.37). Le scelte 1bis), 3), 4), 5) e 6) permettono di applicare correttamente il Teorema 1.26, ottenendo $|E| = 78 \cdot 11 \cdot 6^2 \cdot 4 = 123\,552$. \square

1.2.4 Disposizioni semplici e permutazioni

Definizione 1.28. Dati $k \in \mathbb{N}$ e un insieme finito A , si dicono *disposizioni semplici* (o *disposizioni senza ripetizione*) di k elementi estratti da A le funzioni $f: \{1, \dots, k\} \rightarrow A$ iniettive.

Le disposizioni semplici possono essere identificate con quegli elementi (x_1, \dots, x_k) dell'insieme prodotto $A^k := A \times \dots \times A$ (k volte) tali che $x_i \neq x_j$ se $i \neq j$, cioè con le *sequenze ordinate di elementi distinti di A* .

Per quanto visto nell'Esempio 1.25, grazie al principio fondamentale del calcolo combinatorio, il numero di disposizioni semplici di k elementi estratti da un insieme A che ne contiene n è pari a $n(n-1) \cdots (n-k+1)$. Introduciamo per $n \in \mathbb{N}$ il simbolo “ $n!$ ”, detto “ n fattoriale”, definito come il prodotto degli interi da 1 a n , e poniamo

per convenzione $0! := 1$:

$$n! := n(n-1) \cdots 1 = \prod_{i=1}^n i \quad \text{per } n \in \mathbb{N}, \quad 0! := 1. \quad (1.19)$$

Possiamo esprimere riscrivere la formula $n(n-1) \cdots (n-k+1)$ nel modo più compatto $n!/(n-k)!$. Abbiamo dunque mostrato il seguente risultato.

Proposizione 1.29. *Le disposizioni semplici di k elementi estratti da un insieme che ne contiene n , dove $1 \leq k \leq n$, sono in numero $D_{n,k} = n!/(n-k)!$.*

Osservazione 1.30. In completa analogia con l'Osservazione 1.24, per ogni insieme finito $B = \{b_1, \dots, b_k\}$ con $|B| = k$ il sottoinsieme di A^B costituito dalle funzioni *iniettive* da B in A è in corrispondenza biunivoca con le disposizioni semplici di k elementi estratti da A . La sua cardinalità è pertanto $n!/(n-k)!$. \square

Osservazione 1.31. Per determinare il comportamento asintotico di $n!$ per elevati valori di n , risulta di grande utilità la *formula di Stirling*:

$$n! \sim n^n e^{-n} \sqrt{2\pi n} \quad \text{per } n \rightarrow \infty, \quad (1.20)$$

intendendo che il rapporto tra i due membri ha limite 1. Per una dimostrazione elementare di questa formula (a meno del prefattore $\sqrt{2\pi}$, ma con una stima precisa sull'approssimazione), si veda la Proposizione 1.34 più sotto. Una dimostrazione indipendente, compreso il prefattore, verrà fornita nel Paragrafo 7.2.4. \square

Se A è un insieme finito, le funzioni $f : A \rightarrow A$ iniettive sono necessariamente suriettive, dunque biunivoche. Esse sono dette *permutazioni di A* e formano un gruppo rispetto alla composizione di applicazioni. A meno di corrispondenze biunivoche, non costa niente considerare il caso “speciale” $A = \{1, \dots, n\}$: in questo caso, il gruppo delle permutazioni è indicato con S_n ed è detto *gruppo simmetrico*:

$$S_n := \{ \sigma : \{1, \dots, n\} \rightarrow \{1, \dots, n\} \text{ biunivoche} \}.$$

Per quanto visto sopra, si ha $|S_n| = n!$. Munito della probabilità uniforme P , lo spazio di probabilità (S_n, P) ha proprietà interessanti e per certi versi sorprendenti, come vedremo nel Paragrafo 2.1.

Esempio 1.32. Mischiando un mazzo di carte da poker, la sequenza ordinata delle carte che ne risulta è una permutazione delle carte del mazzo. Il numero delle possibili sequenze ottenute in questo modo è dunque pari a $52! \approx 8 \cdot 10^{67}$. \square

Esempio 1.33 (Paradosso dei compleanni). Determiniamo la probabilità p_n che, in un gruppo di n persone selezionate in modo casuale (nate tutte in un anno non bisestile, per semplicità), almeno due di esse compiano gli anni lo stesso giorno. Quanto deve essere grande n affinché $p_n > 1/2$?

Numerati i giorni dell'anno e le n persone, lo spazio campionario naturale è dato dalle disposizioni con ripetizione di n elementi dall'insieme dei giorni dell'anno, ossia $\Omega = \{f : \{1, \dots, n\} \rightarrow \{1, \dots, 365\}\}$. Abbiamo visto che $|\Omega| = 365^n$. L'evento di interesse è dato da $A = \{f \in \Omega : f \text{ non è iniettiva}\}$. Munendo lo spazio Ω della probabilità uniforme P , si ha dunque

$$p_n = P(A) = 1 - P(A^c) = 1 - \frac{|A^c|}{|\Omega|},$$

dove, per quanto detto sopra,

$$|A^c| = |\{f \in \Omega : f \text{ è iniettiva}\}| = \frac{365!}{(365-n)!} = \prod_{i=0}^{n-1} (365-i).$$

Si ottiene dunque

$$p_n = 1 - \frac{\prod_{i=0}^{n-1} (365-i)}{365^n} = 1 - \prod_{i=0}^{n-1} \frac{365-i}{365} = 1 - \prod_{i=0}^{n-1} \left(1 - \frac{i}{365}\right). \quad (1.21)$$

Abbiamo ottenuto un'espressione esatta, anche se non del tutto "esplicita". Con l'ausilio di un calcolatore, si verifica facilmente che $p_n > \frac{1}{2}$ se e solo se $n \geq 23$. In altri termini, *in un gruppo di almeno 23 persone c'è una probabilità maggiore del 50% che almeno due persone abbiano lo stesso compleanno*, un risultato a prima vista sorprendente. Per un gruppo di 50 persone, la probabilità è maggiore del 97%.

È interessante notare che l'espressione (1.21) può essere approssimata in modo efficace ed esplicito: sfruttando la disuguaglianza $1 - x \leq e^{-x}$, valida per ogni $x \in \mathbb{R}$, e ricordando la relazione (0.7), si ottiene

$$p_n \geq 1 - \exp\left(-\sum_{i=0}^{n-1} \frac{i}{365}\right) = 1 - \exp\left(-\frac{(n-1)n}{2 \cdot 365}\right). \quad (1.22)$$

Questa espressione è più grande di $\frac{1}{2}$ per $n > n_0 = \frac{1}{2}(1 + \sqrt{1 + 4 \cdot 730 \cdot \log 2}) \approx 22.9$, dunque non appena $n \geq 23$, esattamente come l'espressione esatta.

Per avere un po' di intuizione, si può osservare che non conta tanto il numero n di persone, quanto piuttosto il numero $c_n := \frac{1}{2}n(n-1)$ di *coppie di persone* nel gruppo (una quantità che appare in (1.22)). Il punto è che c_n diventa comparabile con il numero 365 di giorni dell'anno per valori relativamente piccoli di n : per esempio $c_n = 253$ per $n = 23$. Ritorneremo su questo problema nell'Esempio 3.56. \square

Dimostriamo ora una versione rafforzata della formula di Stirling (1.20).

Proposizione 1.34. *Per ogni $n \geq 1$ esiste $\theta(n) \in [0, 1]$ tale che*

$$n! = \sqrt{2\pi} n^n e^{-n} \sqrt{n} e^{\frac{\theta(n)}{12n}}, \quad \forall n \geq 1. \quad (1.23)$$

Dimostrazione. Dimostriamo che esiste una costante $C \in (0, \infty)$ tale che

$$n! = C n^n e^{-n} \sqrt{n} e^{\frac{\theta(n)}{12n}}, \quad \forall n \geq 1, \quad (1.24)$$

con $\theta(n) \in [0, 1]$. Il fatto che $C = \sqrt{2\pi}$ sarà mostrato nel Paragrafo 7.2.4. Se definiamo

$$d_n := \log n! - \left(n + \frac{1}{2}\right) \log n + n,$$

la relazione (1.24) è equivalente al fatto che esista una costante reale $c (= \log C)$ tale che

$$c \leq d_n \leq c + \frac{1}{12n}, \quad \forall n \geq 1. \quad (1.25)$$

Affermiamo che:

- (i) la successione d_n è decrescente;
- (ii) la successione $d_n - \frac{1}{12n}$ è crescente.

Da queste proprietà segue facilmente (1.25), con $c = \lim_{n \rightarrow \infty} d_n$ (esercizio).

Restano da dimostrare le affermazioni (i) e (ii). Con semplici calcoli si ottiene:

$$d_n - d_{n+1} = \left(n + \frac{1}{2}\right) \log \frac{n+1}{n} - 1 = \frac{2n+1}{2} \log \frac{1 + \frac{1}{2n+1}}{1 - \frac{1}{2n+1}} - 1.$$

Ricordando la serie di Taylor (0.6) del logaritmo, si ha

$$\log \frac{1+t}{1-t} = \log(1+t) - \log(1-t) = \sum_{k=1}^{\infty} (-1)^{k+1} \frac{t^k}{k} + \sum_{k=1}^{\infty} \frac{t^k}{k} = 2 \sum_{k=0}^{\infty} \frac{t^{2k+1}}{2k+1},$$

che converge per $|t| < 1$. Usando tale serie per $t = \frac{1}{2n+1}$ si trova

$$d_n - d_{n+1} = (2n+1) \sum_{k=0}^{+\infty} \frac{1}{2k+1} \frac{1}{(2n+1)^{2k+1}} - 1 = \sum_{k=1}^{+\infty} \frac{1}{2k+1} \frac{1}{(2n+1)^{2k}} \geq 0, \quad (1.26)$$

essendo quest'ultima una serie a termini positivi. Ciò dimostra (i).

Usando di nuovo (1.26) e il fatto che $2k+1 \geq 3$ per $k \geq 1$, si ottiene

$$\begin{aligned} d_n - d_{n+1} &= \sum_{k=1}^{+\infty} \frac{1}{2k+1} \frac{1}{(2n+1)^{2k}} \leq \frac{1}{3} \sum_{k=1}^{+\infty} \frac{1}{(2n+1)^{2k}} = \frac{1}{3} \frac{(2n+1)^{-2}}{1 - (2n+1)^{-2}} \\ &= \frac{1}{3[(2n+1)^2 - 1]} = \frac{1}{12n(n+1)} = \frac{1}{12n} - \frac{1}{12(n+1)}, \end{aligned}$$

dove abbiamo applicato la serie geometrica (0.4). Abbiamo dunque dimostrato che

$$d_n - \frac{1}{12n} \leq d_{n+1} - \frac{1}{12(n+1)},$$

cioè la relazione (ii). □

1.2.5 Combinazioni

Definizione 1.35. Dati $k \in \mathbb{N}$ e un insieme finito A , si dicono *combinazioni di k elementi estratti da A* i sottoinsiemi di A di cardinalità k .

Mentre una disposizione (semplice o con ripetizione) corrisponde a una sequenza ordinata, una combinazione può essere vista come una collezione *non ordinata* di elementi dell'insieme A .

Vogliamo ora determinare il numero $C_{n,k}$ di combinazioni di k elementi estratti da un insieme A che ne contiene n . Chiaramente $C_{n,0} = 1$, perché c'è un solo insieme con zero elementi, l'insieme vuoto. Per determinare $C_{n,k}$ per $k \in \{1, \dots, n\}$, ricordiamo che ci sono $D_{n,k} = n!/(n-k)!$ disposizioni (cioè sequenze ordinate) di k elementi distinti estratti da A . Dato che nelle combinazioni l'ordine degli elementi non conta, dobbiamo identificare le disposizioni che danno origine alla stessa combinazione, cioè che selezionano lo stesso sottoinsieme di A . Dato che ci sono $k!$ riordinamenti possibili (cioè permutazioni) di k elementi fissati, si ottiene

$$C_{n,k} = \frac{D_{n,k}}{k!} = \binom{n}{k}, \quad (1.27)$$

dove abbiamo introdotto il *coefficiente binomiale* $\binom{n}{k}$, detto “ n su k ”, definito da

$$\binom{n}{k} := \frac{n!}{k!(n-k)!}, \quad \text{per } n \in \mathbb{N}_0, k \in \{0, \dots, n\}.$$

Si noti che la formula $C_{n,k} = \binom{n}{k}$ vale anche per $k = 0$. Abbiamo dunque ottenuto:

Proposizione 1.36. *Le combinazioni di k elementi estratti da un insieme che ne contiene n , dove $0 \leq k \leq n$, sono in numero $C_{n,k} = \binom{n}{k}$.*

L'argomento con cui abbiamo derivato la formula (1.27) può essere formalizzato in modo più preciso, ma decisamente più tecnico (si veda la fine del Paragrafo A.3 nell'appendice). Basta esempio concreto per chiarire l'idea: se $A = \{1, 2, 3, 4, 5\}$, il sottoinsieme di tre elementi $\{1, 3, 4\}$ può essere ottenuto mediante $3 \cdot 2 \cdot 1 = 6$ disposizioni distinte, ossia $(1, 3, 4)$, $(1, 4, 3)$, $(3, 4, 1)$, $(3, 1, 4)$, $(4, 1, 3)$, $(4, 3, 1)$. Lo stesso conto vale per qualsiasi altro sottoinsieme di tre elementi. Dato che ci sono $D_{5,3} = 5 \cdot 4 \cdot 3 = 60$ disposizioni di 3 elementi distinti estratti da A , segue che ci sono $D_{5,3}/3! = 60/6 = 10$ combinazioni (sottoinsiemi) di 3 elementi estratti da A .

Esempio 1.37. Ritornando brevemente all'Esempio 1.27, il numero di modi di scegliere 3 “semi” tra i quattro possibili $\{\heartsuit, \diamondsuit, \clubsuit, \spadesuit\}$ è pari al numero di combinazioni di 3 elementi estratti da un insieme che ne contiene 4 ed è dunque dato da $\binom{4}{3} = 4$ (come avevamo concluso per enumerazione diretta). Analogamente, il numero di modi

di scegliere 2 semi è pari a $\binom{4}{2} = 6$ e il numero di modi di scegliere due “tipi” tra i 13 possibili è pari a $\binom{13}{2} = 78$.

Una “mano” a Poker è un sottoinsieme di 5 carte distinte estratte da un mazzo che ne contiene 52. Il numero di possibili mani è dato dunque da $\binom{52}{5} = 2\,598\,960$. Ricordando l'Esempio 1.27, le probabilità di fare *full* oppure *doppia coppia* valgono rispettivamente $3\,744/2\,598\,960 \approx 0.14\%$ e $123\,552/2\,598\,960 \approx 4.8\%$. \square

Elenchiamo alcune semplici proprietà dei coefficienti binomiali:

$$\binom{n}{0} = \binom{n}{n} = 1, \quad \binom{n}{k} = \binom{n}{n-k}, \quad \forall n \in \mathbb{N}_0, \forall k \in \{0, \dots, n\}.$$

Vale inoltre la relazione

$$\binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k}, \quad \forall n \in \mathbb{N}, \forall k \in \{0, \dots, n\}, \quad (1.28)$$

come si verifica facilmente.[†] Ricordiamo infine la formula detta *binomio di Newton*:

$$(a+b)^n = \sum_{k=0}^n \binom{n}{k} a^k b^{n-k}, \quad \forall n \in \mathbb{N}_0, \forall a, b \in \mathbb{R}, \quad (1.29)$$

che si dimostra per induzione usando (1.28).

1.2.6 Estrazioni di palline da un'urna

Un esempio classico di applicazione del calcolo combinatorio è costituito dai problemi di estrazione di palline da un'urna. Al di là dell'interesse matematico, questo genere di problemi ha grande rilevanza applicativa: modelli basati su urne sono usati ad esempio per descrivere meccanismi di rinforzo in teoria dell'apprendimento. Più in generale, molti problemi di *campionamento* da una popolazione possono essere riformulati astrattamente come problemi di estrazione di palline da un'urna.

Ci limiteremo a considerare l'esempio di un'urna contenente N palline di due colori, diciamo M palline rosse e $N - M$ verdi (con $M \leq N$). Supponiamo di eseguire n estrazioni successive, secondo uno dei due schemi seguenti:

- *Estrazioni con reimmissione* (o reimbussolamento). Dopo ogni estrazione, la pallina estratta viene reinserita nell'urna.
- *Estrazioni senza reimmissione*. Le palline estratte non vengono reinserite. In questo caso dev'essere $n \leq N$.

[†] Esiste una dimostrazione “combinatoria” di (1.28). Le combinazioni di k elementi estratti da $\{1, \dots, n\}$ sono $\binom{n}{k}$ e possono essere divise in due sottoinsiemi disgiunti: quelle che contengono n e quelle che non lo contengono; le prime sono in corrispondenza biunivoca con le combinazioni di $(k-1)$ elementi estratti da $\{1, \dots, n-1\}$, che sono $\binom{n-1}{k-1}$, mentre le seconde sono in corrispondenza biunivoca con le combinazioni di k elementi estratti da $\{1, \dots, n-1\}$, che sono $\binom{n-1}{k}$.

Calcoliamo, per ciascuno dei due schemi, la probabilità che esattamente k delle n palline estratte siano rosse. Supponiamo di numerare da 1 a M le palline rosse e da $(M+1)$ a N le palline verdi.

Esempio 1.38 (Estrazioni con reimmissione). L'esito di n estrazioni successive è descritto da una disposizione con ripetizione di n elementi estratti dall'insieme $\{1, 2, \dots, N\}$. Sia dunque $\Omega := \{f : \{1, \dots, n\} \rightarrow \{1, \dots, N\}\}$ l'insieme di tali disposizioni e sia P la probabilità uniforme su Ω . Denotiamo infine con A l'insieme delle disposizioni contenenti esattamente k palline rosse. Si tratta di calcolare

$$P(A) = \frac{|A|}{|\Omega|}.$$

Per determinare $|A|$, utilizziamo il principio fondamentale del calcolo combinatorio. Un elemento di A è determinato dalle seguenti scelte successive.

- Si scelgono le k posizioni, su n possibili, in cui mettere le palline rosse: per questa scelta ci sono $\binom{n}{k}$ esiti possibili.
- Si dispongono k palline rosse (prese dalle M presenti nell'urna) nelle posizioni prescelte: ci sono M^k tali disposizioni.
- Si dispongono $(n-k)$ palline verdi (prese dalle $(N-M)$ presenti nell'urna) nelle rimanenti posizioni: ci sono $(N-M)^{n-k}$ tali disposizioni.

Si ottiene pertanto

$$|A| = \binom{n}{k} M^k (N-M)^{n-k},$$

e dato che $|\Omega| = N^n$, segue facilmente che

$$P(A) = \binom{n}{k} p^k (1-p)^{n-k}, \quad \text{dove } p := \frac{M}{N}. \quad (1.30)$$

Nel Paragrafo 1.3.4 daremo un'interpretazione alternativa a questa espressione. \square

Esempio 1.39 (Estrazioni senza reimmissione). Enumeriamo le palline come nel caso precedente. Un naturale spazio campionario, in cui la probabilità uniforme esprime la casualità dell'estrazione, è quello delle *disposizioni senza ripetizione*. Tuttavia, poiché l'evento "il numero di palline rosse estratte è k " non dipende dall'ordine di estrazione, è forse ancora più naturale scegliere come spazio campionario l'insieme delle *combinazioni*, e così faremo.

Sia dunque Ω l'insieme dei sottoinsiemi di n elementi dell'insieme $\{1, 2, \dots, N\}$, e P la probabilità uniforme su di esso. L'evento di cui vogliamo calcolare la probabilità è

$$A = \{\omega \in \Omega : |\omega \cap \{1, 2, \dots, M\}| = k\} = \{\omega \in \Omega : |\omega \cap \{M+1, \dots, N\}| = n-k\}.$$

Chiaramente $A = \emptyset$ se $k > M$ oppure se $(n-k) > (N-M)$. Supponiamo dunque che $k \leq M$ e $(n-k) \leq (N-M)$. Ogni elemento di A è determinato da due scelte successive: occorre scegliere k elementi da $\{1, 2, \dots, M\}$ e $(n-k)$ da $\{M+1, \dots, N\}$. Di

conseguenza possiamo scrivere

$$|A| = \binom{M}{k} \binom{N-M}{n-k},$$

dove usiamo la convenzione secondo cui $\binom{i}{j} = 0$ se $j < 0$ o $j > i$. Ricordando che $|\Omega| = \binom{N}{n}$, possiamo dunque concludere che

$$P(A) = \frac{\binom{M}{k} \binom{N-M}{n-k}}{\binom{N}{n}}. \quad (1.31)$$

□

Osservazione 1.40 (Estrazioni da un'urna popolosa). Intuitivamente, se l'urna contiene molte palline di ciascun colore (ossia $M \gg 1$ e $N - M \gg 1$), non ci dovrebbe essere grande differenza tra i due schemi di estrazioni considerati, dal momento che la rimozione di una pallina modifica in modo trascurabile la composizione dell'urna. Questo argomento può essere formalizzato matematicamente nel modo seguente.

Sia $(M_N)_{N \in \mathbb{N}}$ una successione di interi positivi tali che

$$\lim_{N \rightarrow \infty} \frac{M_N}{N} = p \in (0, 1). \quad (1.32)$$

Se effettuiamo n estrazioni *senza* reimmissione da un'urna che contiene N palline di cui M_N rosse, la probabilità di estrarne esattamente k rosse è data dalla formula (1.31). Mostriamo che il limite per $N \rightarrow \infty$ di tale probabilità, quando i numeri k, n sono fissati, è dato dalla formula (1.30).

Scrivendo $\binom{M_N}{k} = \frac{1}{k!} \prod_{i=0}^{k-1} (M_N - i)$ e analogamente per $\binom{N-M_N}{n-k}$ e $\binom{N}{n}$, con qualche manipolazione algebrica a partire da (1.31) si arriva all'espressione

$$\begin{aligned} P(A) &= \binom{n}{k} \left(\prod_{i=0}^{k-1} \frac{M_N - i}{N - i} \right) \left(\prod_{j=0}^{(n-k)-1} \frac{N - M_N - j}{N - k - j} \right) \\ &= \binom{n}{k} \left(\frac{M_N}{N} \right)^k \left(\prod_{i=0}^{k-1} \frac{1 - \frac{i}{M_N}}{1 - \frac{i}{N}} \right) \left(1 - \frac{M_N}{N} \right)^{n-k} \left(\prod_{j=0}^{(n-k)-1} \frac{1 - \frac{j}{N-M_N}}{1 - \frac{k+j}{N}} \right). \end{aligned}$$

Osserviamo che $N \rightarrow \infty$, $M_N \rightarrow \infty$ e anche $(N - M_N) \rightarrow \infty$, grazie a (1.32). Dato che n e k sono fissati, i termini tra parentesi graffe nella relazione precedente tendono a 1 per $N \rightarrow \infty$. Poiché $M_N/N \rightarrow p$, segue che

$$\lim_{N \rightarrow \infty} P(A) = \binom{n}{k} p^k (1-p)^{n-k}, \quad (1.33)$$

che coincide esattamente con l'espressione (1.30), come volevasi dimostrare.

Per questa ragione, quando l'urna contiene molte palline di ciascun colore, il calcolo della probabilità di estrarre k palline rosse in n estrazioni *senza* reimmissione viene fatto tipicamente usando la formula (1.30), valida per il caso di estrazioni *con*

reimmissione, che è più maneggevole di (1.31) e ne costituisce un'ottima approssimazione (per lo meno se n e k sono molto minori di M e $N - M$). \square

Concludiamo con un esempio di grande rilevanza teorica e applicativa, che ritroveremo nel seguito (si vedano gli Esempi 2.11 e 3.105). Sebbene sia formulato in termini di “oggetti” inseriti in “cassetti”, piuttosto che di palline estratte da un'urna, la struttura matematica sottostante è la stessa dell'Esempio 1.38.

Esempio 1.41 (Inserimento di oggetti in cassetti). Supponiamo di inserire casualmente n oggetti distinti in r cassetti disponibili (ciascun cassetto può contenere un numero qualunque di oggetti). Fissiamo $k \in \{0, \dots, n\}$ e chiediamoci qual è la probabilità che esattamente k oggetti finiscano nel primo cassetto.

Numerati gli oggetti da 1 a n e i cassetti da 1 a r , uno spazio campionario naturale è dato dall'insieme Ω delle disposizioni con ripetizione di n elementi estratti dall'insieme $\{1, \dots, r\}$ (e non viceversa!), ossia

$$\Omega := \{f : \{1, \dots, n\} \rightarrow \{1, \dots, r\}\}$$

Intuitivamente, $f \in \Omega$ codifica l'esito dell'esperimento in cui ciascun oggetto $i \in \{1, \dots, n\}$ viene inserito nel cassetto $f(i) \in \{1, \dots, r\}$. L'evento “il primo cassetto contiene k oggetti” corrisponde dunque al sottoinsieme delle funzioni che assumono il valore “1” esattamente k volte, ossia

$$A = \{f \in \Omega : |f^{-1}(1)| = k\} \subseteq \Omega.$$

La casualità dell'inserimento induce a munire Ω della probabilità uniforme P . Di conseguenza, lo spazio di probabilità (Ω, P) è *esattamente lo stesso* che descrive l'estrazione con reimmissione di n palline da un'urna che ne contiene $N = r$. Inoltre, quando nell'urna c'è solo una pallina rossa, la numero 1, l'evento A corrisponde all'evento “vengono estratte k palline rosse”. Pertanto, ricordando la formula (1.30) per $M = 1$, la probabilità che esattamente k oggetti finiscano nel primo cassetto vale

$$P(A) = \binom{n}{k} p^k (1-p)^{n-k}, \quad \text{con } p = \frac{1}{r}. \quad (1.34)$$

Questa formula verrà reinterpretata più avanti, nel Paragrafo 1.3.4. \square

Esercizi

Esercizio 1.6. Siano Ω_1 e Ω_2 due insiemi finiti, sia P_1 la probabilità uniforme su Ω_1 e P la probabilità uniforme su $\Omega_1 \times \Omega_2$. Si mostri che per ogni $A \subset \Omega_1$

$$P_1(A) = P(A \times \Omega_2).$$

Esercizio 1.7. Si consideri un mazzo di 52 carte da Poker, e si scelgano *a caso* 5 carte. Si calcoli la probabilità che:

- (i) nelle 5 carte ci sia *almeno* una coppia (cioè due carte di semi diversi ma con lo stesso numero o figura);
- (ii) nelle 5 carte ci sia *esattamente* una coppia, cioè ci sia una coppia ma nessuna combinazione migliore (doppia coppia, tris....)

Esercizio 1.8. Una classe è costituita da 30 persone, tra cui Giacomo, Claudio e Nicola. Un insegnante divide in modo casuale la classe in tre gruppi di 10 persone.

- (i) Qual è la probabilità che Giacomo, Claudio e Nicola finiscano in tre gruppi distinti? (Non semplificare i coefficienti binomiali)
 - (ii) Qual è la probabilità che finiscano nello stesso gruppo?
-

1.3 Probabilità condizionale e indipendenza

Nello studio di un modello probabilistico, risulta interessante analizzare l'influenza che il verificarsi di un dato evento B ha sulla probabilità di occorrenza di un altro evento A . Questo conduce alle nozioni di probabilità condizionale e di indipendenza di eventi, di importanza fondamentale.

1.3.1 Probabilità condizionale

Consideriamo un evento A di un esperimento aleatorio, che abbia probabilità $P(A)$. Se veniamo a conoscenza del fatto che un altro evento B si è verificato, come è sensato aggiornare il valore di $P(A)$ per tenere conto di questa informazione? La risposta è la probabilità condizionale (o condizionata) $P(A|B)$ di A sapendo B , che definiremo tra un momento.

Per motivare e “indovinare” la definizione, ritorniamo per un istante all'*interpretazione frequentista*. Ripetendo $N \gg 1$ volte l'esperimento aleatorio in condizioni “analoghe e indipendenti” e contando il numero $\mathcal{S}_N(A)$ di volte in cui l'evento A si verifica, si ha $P(A) \simeq \mathcal{S}_N(A)/N$. Per tenere conto dell'informazione aggiuntiva che l'evento B si è verificato, è naturale limitarsi a considerare le volte in cui l'esperimento ha dato esito in B , che sono in numero $\mathcal{S}_N(B)$, e contare in quante di queste volte anche A si è verificato, ossia $\mathcal{S}_N(A \cap B)$. Definendo la probabilità condizionale $P(A|B)$ come il rapporto di tali numeri, per N grande si ottiene

$$P(A|B) \simeq \frac{\mathcal{S}_N(A \cap B)}{\mathcal{S}_N(B)} = \frac{\mathcal{S}_N(A \cap B)/N}{\mathcal{S}_N(B)/N} \simeq \frac{P(A \cap B)}{P(B)}.$$

Queste considerazioni euristiche motivano la seguente definizione.

Definizione 1.42 (Probabilità condizionale). Siano A e B due eventi di uno spazio di probabilità discreto (Ω, P) , con $P(B) > 0$. Si dice *probabilità condizionale di A dato B* (o sapendo B , o rispetto a B) la quantità

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

Esempio 1.43. Nelle estrazioni per una ruota del Lotto, vengono scelte “a caso” cinque palline da un’urna che ne contiene novanta, numerate da 1 a 90. Supponiamo di giocare due numeri su quella ruota, e precisamente l’1 e il 7. Una persona, presente all’estrazione, mi avvisa che dei cinque numeri estratti tre sono dispari. Qual è la probabilità di fare “ambo” sulla base di questa informazione? E qual è la probabilità in assenza di tale informazione?

Lo spazio campionario naturale per l’esperimento aleatorio in questione è $\Omega =$ “insieme di tutte le cinque di numeri tra 1 e 90” $= \{C \subseteq \{1, \dots, 90\} : |C| = 5\}$. Con la terminologia del paragrafo precedente, Ω è l’insieme delle combinazioni di cinque elementi estratti dall’insieme $\{1, \dots, 90\}$, dunque $|\Omega| = \binom{90}{5}$. Assumendo l’equità dell’estrazione, scegliamo come probabilità P su Ω quella uniforme. I due eventi che compaiono nell’enunciato del problema sono

$$A = \text{“i numeri estratti contengono l’1 e il 7”} = \{C \in \Omega : \{1, 7\} \subseteq C\},$$

$$B = \text{“tre dei numeri estratti sono dispari”} = \{C \in \Omega : |C \cap (2\mathbb{N}_0 + 1)| = 3\},$$

dove indichiamo con $2\mathbb{N}_0 + 1$ l’insieme dei numeri dispari. La scelta di un elemento di A corrisponde alla scelta di tre numeri diversi da 1 e 7, pertanto $|A| = \binom{88}{3}$. Di conseguenza, in assenza dell’informazione sull’occorrenza di B si ha

$$P(A) = \frac{|A|}{|\Omega|} = \frac{\binom{88}{3}}{\binom{90}{5}} = \frac{88! \cdot 5!}{90! \cdot 3!} = \frac{5 \cdot 4}{90 \cdot 89} = \frac{2}{801} \simeq 0.0025.$$

La probabilità di A *condizionale* all’occorrenza di B è data da

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{|A \cap B|/|\Omega|}{|B|/|\Omega|} = \frac{|A \cap B|}{|B|}.$$

Gli elementi di $A \cap B$ sono costituiti dalle cinque contenenti i numeri 1 e 7, un altro numero dispari diverso da 1 e 7 (per cui ci sono 43 scelte) e due numeri pari (per cui ci sono $\binom{45}{2}$ scelte), dunque $|A \cap B| = 43 \binom{45}{2}$. Dato che gli elementi di B contengono tre numeri dispari e due pari, si ha $|B| = \binom{45}{3} \binom{45}{2}$ e dunque

$$P(A|B) = \frac{|A \cap B|}{|B|} = \frac{43}{\binom{45}{3}} = \frac{43 \cdot 42! \cdot 3!}{45!} = \frac{6}{45 \cdot 44} = \frac{1}{330} \simeq 0.003,$$

che è maggiore della probabilità in assenza di informazioni. \square

Alcune proprietà formali della probabilità condizionale sono sintetizzate nella seguente proposizione, la cui dimostrazione è lasciata per esercizio.

Proposizione 1.44. *Sia B un evento fissato di uno spazio di probabilità discreto (Ω, \mathcal{P}) . La funzione $P(\cdot | B)$, ossia*

$$\begin{aligned} \mathcal{P}(\Omega) &\longrightarrow [0, 1] \\ A &\longmapsto P(A|B), \end{aligned}$$

è una probabilità su Ω .

Sottolineiamo che, fissato un evento A , la funzione $B \mapsto P(A|B)$ non è una probabilità (ad esempio, non è nemmeno definita per $B = \emptyset$).

Notiamo che la probabilità condizionale permette di esprimere in modo ricorsivo la probabilità dell'intersezione di n eventi.

Proposizione 1.45 (Regola della catena). *Per $n \in \mathbb{N}$, $n \geq 2$ siano A_1, \dots, A_n eventi tali che $P(A_1 \cap A_2 \cap \dots \cap A_{n-1}) > 0$. Allora*

$$\begin{aligned} P\left(\bigcap_{i=1}^n A_i\right) &= P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2) \cdots P(A_n|A_1 \cap \dots \cap A_{n-1}) \\ &= P(A_1) \cdot \prod_{i=2}^n P(A_i|A_1 \cap \dots \cap A_{i-1}). \end{aligned} \quad (1.35)$$

Dimostrazione. Notiamo innanzitutto che $(A_1 \cap A_2 \cap \dots \cap A_j) \subseteq (A_1 \cap A_2 \cap \dots \cap A_i)$ per $i \leq j$. In particolare, se $P(A_1 \cap A_2 \cap \dots \cap A_{n-1}) > 0$ anche $P(A_1 \cap A_2 \cap \dots \cap A_i) > 0$ per ogni $1 \leq i \leq n-1$, dunque le probabilità condizionali nel membro destro in (1.35) sono ben definite.

La dimostrazione procede per induzione. Dati due eventi B, C con $P(C) > 0$ si ha, per definizione di probabilità condizionale,

$$P(B \cap C) = P(B)P(C|B), \quad (1.36)$$

dunque la relazione (1.35) è verificata per $n = 2$. Per il passo induttivo, notiamo che grazie alla relazione (1.36) si ha

$$P\left(\bigcap_{i=1}^n A_i\right) = P\left(\left(\bigcap_{i=1}^{n-1} A_i\right) \cap A_n\right) = P\left(\bigcap_{i=1}^{n-1} A_i\right) P\left(A_n \middle| \bigcap_{i=1}^{n-1} A_i\right),$$

da cui, applicando l'ipotesi induttiva, si ottiene la relazione (1.35). \square

1.3.2 Bayes e dintorni

In molte situazioni, la nozione di probabilità condizionale è utile nella costruzione stessa di un modello probabilistico: talvolta è “naturale” assegnare il valore di alcune probabilità condizionali, e da esse dedurre il valore di probabilità non condizionali.

Esempio 1.46. Due urne, che indichiamo con α e β , contengono rispettivamente 3 palline rosse e 1 verde (l’urna α) e 1 pallina rossa e 1 verde (l’urna β). Si sceglie, con la stessa probabilità, una delle due urne e poi, dall’urna scelta, si estrae una pallina a caso. Qual è la probabilità di estrarre una pallina rossa?

Come spazio campionario si può scegliere l’insieme di quattro elementi $\Omega = \{\alpha r, \alpha v, \beta r, \beta v\}$, dove la prima lettera di ogni elemento $\omega \in \Omega$ indica l’urna scelta e la seconda il colore della pallina estratta. L’evento $A = \{\alpha r, \alpha v\}$ corrisponde a “l’urna scelta è α ”, l’evento $R = \{\alpha r, \beta r\}$ corrisponde a “la pallina estratta è rossa”. Dev’essere senz’altro $P(A) = 1/2$, visto che le urne vengono scelte con uguale probabilità. Inoltre, *supponendo* di aver scelto l’urna α , la probabilità di estrarre una pallina rossa è $3/4$, perciò porremo $P(R|A) = 3/4$. Analogamente $P(R|A^c) = 1/2$. Osservando che $P(A \cap R) = P(A)P(R|A)$ per definizione di speranza condizionale, si ricava

$$P(\{\alpha r\}) = P(A \cap R) = P(A)P(R|A) = \frac{1}{2} \frac{3}{4} = \frac{3}{8}.$$

Analogamente

$$P(\{\beta r\}) = P(A^c \cap R) = P(A^c)P(R|A^c) = \frac{1}{2} \frac{1}{2} = \frac{1}{4},$$

e sfruttando la Proposizione 1.44

$$P(\{\alpha v\}) = P(A \cap R^c) = P(A)P(R^c|A) = P(A)(1 - P(R|A)) = \frac{1}{2} \frac{1}{4} = \frac{1}{8},$$

$$P(\{\beta v\}) = P(A^c \cap R^c) = P(A^c)P(R^c|A^c) = P(A^c)(1 - P(R|A^c)) = \frac{1}{2} \frac{1}{2} = \frac{1}{4}.$$

In questo modo abbiamo calcolato la densità $p(\omega) := P(\{\omega\})$ per ogni $\omega \in \Omega$. Per quanto visto nel Paragrafo 1.1.3 (o per verifica diretta) sappiamo che ciò determina completamente la probabilità P su $(\Omega, \mathcal{P}(\Omega))$, per cui possiamo calcolare la probabilità di qualunque evento. In particolare otteniamo

$$P(R) = P(\{\alpha r, \beta r\}) = P(\{\alpha r\}) + P(\{\beta r\}) = \frac{3}{8} + \frac{1}{4} = \frac{5}{8}, \quad (1.37)$$

la risposta cercata. □

Nell’esempio precedente, scrivere esplicitamente lo spazio campionario Ω e determinare completamente la probabilità P è stato un esercizio istruttivo. Tuttavia è interessante notare che ciò *non è strettamente necessario* per rispondere alla domanda posta, come mostra la proposizione seguente.

Proposizione 1.47 (Formule di disintegrazione e delle probabilità totali). Sia (Ω, \mathcal{P}) uno spazio di probabilità discreto e sia $(B_i)_{i \in I}$ una partizione finita o numerabile di Ω , ossia $|I| \leq |\mathbb{N}|$, $B_i \cap B_j = \emptyset$ per ogni $i \neq j$ e $\bigcup_{i \in I} B_i = \Omega$.

Per ogni evento A vale la formula di disintegrazione

$$P(A) = \sum_{i \in I} P(A \cap B_i).$$

Se inoltre $P(B_i) > 0$ per ogni $i \in I$, vale la formula delle probabilità totali:

$$P(A) = \sum_{i \in I} P(A|B_i) P(B_i).$$

In particolare, per ogni coppia di eventi A, B con $0 < P(B) < 1$, si ha

$$P(A) = P(A|B) P(B) + P(A|B^c) P(B^c) = P(A|B) P(B) + P(A|B^c)(1 - P(B)).$$

Dimostrazione. Per ipotesi $\Omega = \bigcup_{i \in I} B_i$, pertanto

$$A = A \cap \Omega = A \cap \left(\bigcup_{i \in I} B_i \right) = \bigcup_{i \in I} (A \cap B_i).$$

Dato che gli eventi $(B_i)_{i \in I}$ sono disgiunti, anche gli eventi $(A \cap B_i)_{i \in I}$ lo sono. Dato che $|I| \leq |\mathbb{N}|$, per la σ -addittività di P si ha

$$P(A) = \sum_{i \in I} P(A \cap B_i).$$

Infine, se $P(B_i) > 0$, possiamo scrivere $P(A \cap B_i) = P(A|B_i) P(B_i)$ per definizione di probabilità condizionale. \square

Esempio 1.48. Ritornando all'Esempio 1.46, i dati dell'esercizio fornivano i valori $P(A) = \frac{1}{2}$, $P(R|A) = \frac{3}{4}$ e $P(R|A^c) = \frac{1}{2}$. Applicando la formula delle probabilità totali si ottiene dunque

$$P(R) = P(R|A) P(A) + P(R|A^c)(1 - P(A)) = \frac{3}{4} \frac{1}{2} + \frac{1}{2} \frac{1}{2} = \frac{5}{8}, \quad (1.38)$$

che coincide (naturalmente) con il risultato (1.37) trovato in precedenza. \square

Supponiamo ora che A e B siano due eventi tali che $P(A) > 0$, $P(B) > 0$, sicché entrambe le probabilità condizionali $P(A|B)$ e $P(B|A)$ sono definite. È pressoché immediato verificare la seguente relazione.

Teorema 1.49 (Formula di Bayes). *Se A, B sono eventi di uno spazio di probabilità discreto (Ω, \mathcal{P}) , con $P(A) > 0$ e $P(B) > 0$, allora vale la formula*

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)}. \quad (1.39)$$

Dimostrazione. La formula di Bayes (1.39) è equivalente a

$$P(B|A)P(A) = P(A|B)P(B),$$

che è vera in quanto, per definizione di probabilità condizionale, entrambi i membri sono uguali a $P(A \cap B)$. \square

Osservazione 1.50. Nell'ipotesi che $0 < P(B) < 1$, usando la formula delle probabilità totali, la formula di Bayes può essere riscritta nella forma

$$P(B|A) = \frac{P(A|B)P(B)}{P(A|B)P(B) + P(A|B^c)P(B^c)}. \quad (1.40)$$

Analogamente, se $(B_i)_{i \in I}$ è una famiglia di eventi finita o numerabile che soddisfa alle ipotesi della Proposizione 1.47, per ogni $i \in I$ si ha

$$P(B_i|A) = \frac{P(A|B_i)P(B_i)}{P(A)} = \frac{P(A|B_i)P(B_i)}{\sum_{j \in I} P(A|B_j)P(B_j)}. \quad (1.41)$$

Le versioni (1.40) e (1.41) della formula di Bayes sono quelle che più spesso capita di usare negli esercizi. \square

La formula di Bayes, a dispetto della sua semplicità, ha un'importanza fondamentale ed è all'origine di un'intera area della statistica, detta appunto *bayesiana*. La rilevanza della formula di Bayes si può già apprezzare in applicazioni elementari, ma dai risultati a prima vista sorprendenti, come quelle descritte negli esempi che seguono (si veda l'Osservazione 1.53 seguente per una spiegazione intuitiva).

Esempio 1.51. Ritorniamo un'ultima volta all'Esempio 1.46, con le stesse notazioni. Se la pallina estratta è rossa, qual è la probabilità che l'urna scelta sia stata α ?

Introducendo gli eventi $A = \text{"l'urna scelta è } \alpha\text{"}$ e $R = \text{"la pallina estratta è rossa"}$, dobbiamo calcolare la probabilità condizionale $P(A|R)$. Sappiamo che $P(A) = \frac{1}{2}$, $P(R|A) = \frac{3}{4}$ e $P(R|A^c) = \frac{1}{2}$ e abbiamo già calcolato $P(R) = \frac{5}{8}$, cf. (1.38). Applicando la formula di Bayes otteniamo dunque

$$P(A|R) = \frac{P(R|A)P(A)}{P(R)} = \frac{\frac{3}{4} \cdot \frac{1}{2}}{\frac{5}{8}} = \frac{3}{5}.$$

Si noti che tale probabilità è diversa dalla probabilità originale (non condizionale) di scegliere l'urna α , che vale $P(A) = \frac{1}{2}$, a dispetto di quanto un'intuizione errata avrebbe potuto suggerire. \square

Esempio 1.52 (Test clinico). Per determinare la presenza di un virus, viene elaborato un test clinico avente la seguente efficacia: se il virus è presente, il test risulta positivo nel 99% dei casi; se il virus è assente, il test risulta positivo nel 2% dei casi. È noto che 4 persone su 10 000 hanno il virus. Supponiamo che un individuo scelto a caso risulti positivo al test. Con quale sicurezza possiamo affermare che sia malato?

Come accade sovente, non è rilevante descrivere nel dettaglio lo spazio campionario. Si considerino gli eventi descritti informalmente da $A = \text{“il test è risultato positivo”}$ e $B = \text{“l'individuo è malato”}$. I dati del problema sono:

$$P(B) = 0.0004, \quad P(A|B) = 0.99, \quad P(A|B^c) = 0.02. \quad (1.42)$$

Calcoliamo $P(B|A)$. Utilizzando la formula di Bayes e la formula delle probabilità totali, si ha

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)} = \frac{P(A|B)P(B)}{P(A|B)P(B) + P(A|B^c)P(B^c)} = \frac{0.000396}{0.020388} \simeq 0.02$$

che è estremamente bassa. Quindi, anche se un individuo risulta positivo al test, è estremamente improbabile che sia malato! Questo test dunque darà una grandissima percentuale di falsi positivi. \square

Osservazione 1.53. Per interpretare più intuitivamente i risultati degli esempi precedenti, può essere utile tenere a mente l'interpretazione frequentista della probabilità condizionale, accennata all'inizio del paragrafo.

Per quanto riguarda l'Esempio 1.51, immaginiamo di ripetere l'intero esperimento (scelta dell'urna e estrazione della pallina) un numero elevato di volte $N \gg 1$. La pallina estratta sarà rossa (cioè si verificherà l'evento R) un numero di volte pari a $\mathcal{S}_N(R) \simeq P(R) \cdot N$. Ignorando le volte in cui la pallina estratta *non* è rossa, ci si può chiedere quante volte l'urna scelta è α (cioè si verifica l'evento A): la risposta è $\mathcal{S}_N(A \cap R) \simeq P(A \cap R) \cdot N$. Di conseguenza, *restringendosi alle ripetizioni dell'esperimento in cui la pallina estratta è rossa*, la frazione di volte in cui l'urna scelta è α è proprio $\mathcal{S}_N(A \cap R) / \mathcal{S}_N(R) \simeq P(A \cap R) / P(R) = P(A|R)$. In altre parole, $P(A|R) = \frac{3}{5}$ significa che l'urna scelta è α circa $3/5$ delle volte in cui la pallina estratta è rossa.

Un discorso del tutto analogo può essere fatto per l'Esempio 1.52: tra gli individui che risultano positivi al test, soltanto 2 su 100 sono effettivamente malati. Per capire più concretamente che cosa succede in questo caso, prendiamo in considerazione una popolazione numerosa, diciamo di 100 000 persone, di cui 40 sono malate e le restanti 99 960 sono sane (per ipotesi 4 persone su 10 000 sono malate). Immaginiamo ora di sottoporre *tutta la popolazione* al test:

- per le persone malate, il test darà (correttamente) esito positivo sul 99% di esse, ossia su $40 \cdot 0.99 \simeq 40$ persone;
- per le persone sane, il test darà (erroneamente!) esito positivo sul 2% di esse, ossia su $99\,960 \cdot 0.02 \simeq 2000$ persone.

In definitiva, ci saranno $2000 + 40 = 2040$ persone su cui il test ha dato esito positivo, ma soltanto 40 tra di loro sono effettivamente malate! La frazione di indi-

vidui effettivamente malati, rispetto a quelli risultati positivi al test, vale pertanto $40/2040 \simeq 0.02$, come calcolato in precedenza. \square

Osservazione 1.54. La conclusione dell'Esempio 1.52 appare paradossale perché la frazione $\alpha = P(A|B^c)$ di falsi positivi è relativamente bassa ($\alpha = 2\%$). Ma il punto è che non è rilevante la grandezza *assoluta* di α , quanto piuttosto la sua grandezza *relativa* rispetto a $\beta = P(B)$, la frazione di persone malate ($\beta = 0.04\%$). In effetti, ricordando la relazione (1.40), se $P(A|B)$ è molto vicino a 1 e se β è piccolo, come è tipico, si può approssimare $P(A|B) \simeq 1$ e $P(B^c) \simeq 1$, ottenendo

$$P(B|A) \simeq \frac{P(B)}{P(B) + P(A|B^c)} = \frac{\beta}{\beta + \alpha} = \frac{1}{1 + \alpha/\beta}.$$

Se $\alpha \gg \beta$, come nell'Esempio 1.52 ($\alpha/\beta = 50$), allora $P(B|A)$ sarà piccola. Se invece si vuole avere $P(A|B)$ vicino a 1 (questo significa che gli individui risultati positivi al test sono molto probabilmente malati) occorre che $\alpha \ll \beta$. \square

Osservazione 1.55. E se avessimo voluto specificare per bene lo spazio campionario dell'Esempio 1.52? Si sarebbe potuto procedere così. Definiamo

$$\Omega = \{(m, p), (m, n), (s, p), (s, n)\} = \{m, s\} \times \{p, n\}$$

dove m e s indicano la presenza (m) o l'assenza (s) del virus, p e n il risultato del test (p = positivo, n = negativo). Qual è la probabilità P su Ω ? Per individuare P dobbiamo usare i dati del problema. Si noti che gli eventi A e B definiti sopra corrispondono ai seguenti sottoinsiemi di Ω :

$$A = \{(m, p), (s, p)\}, \quad B = \{(m, p), (m, n)\}.$$

Usando i dati in (1.42), è effettivamente possibile calcolare la probabilità di tutti i sottoinsiemi di Ω (esercizio), da cui si può dedurre il valore di $P(B|A)$. Tuttavia, per rispondere al quesito posto, questi dettagli sono poco rilevanti. \square

Come abbiamo visto, in molti casi è prassi non scrivere esplicitamente lo spazio campionario Ω ma limitarsi a introdurre gli eventi rilevanti, specificandone alcune probabilità assolute e/o condizionali. Questo accadrà frequentemente nel seguito. Sottolineiamo tuttavia che le probabilità (assolute o condizionali) di eventi *non possono essere assegnate arbitrariamente*: ad esempio, non è possibile che due eventi A e B siano tali che $P(A \cap B) = 0$, $P(A) = \frac{1}{2}$ e $P(B) = \frac{2}{3}$, perché in tal caso si dovrebbe avere $P(A \cup B) = P(A) + P(B) = \frac{7}{6} > 1$, il che è impossibile.

Negli esempi ed esercizi che incontreremo, con un po' di pazienza sarà sempre possibile scrivere esplicitamente uno spazio campionario Ω che realizza le specifiche assegnate (si veda ad esempio l'Esercizio 1.12).

1.3.3 Indipendenza di eventi

Si è visto come la probabilità condizionale $P(A|B)$ rappresenti la probabilità dell'evento A sotto la condizione del verificarsi dell'evento B . È possibile che tale condi-

zione non modifichi la probabilità di A , ossia

$$P(A|B) = P(A). \quad (1.43)$$

Usando la definizione di probabilità condizionale, si vede che l'identità (1.43) equivale a:

$$P(A \cap B) = P(A)P(B). \quad (1.44)$$

L'identità in (1.44), rispetto a quella in (1.43) ha il vantaggio di essere esplicitamente simmetrica in A e B , e di essere definita (e banalmente vera) anche quando $P(B) = 0$. Essa viene dunque scelta per caratterizzare la nozione di *indipendenza*.

Definizione 1.56 (Indipendenza di due eventi). Due eventi A e B di uno spazio di probabilità discreto (Ω, P) si dicono *indipendenti* se

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

Esempio 1.57. Da due mazzi di carte da Poker si estraggono due carte, una per mazzo. Lo spazio campionario naturale è l'insieme delle coppie (i, j) nel prodotto cartesiano $\Omega = X \times X$, dove X è l'insieme delle carte di un mazzo. Possiamo assumere che la scelta sia “casuale”, cioè descritta dalla probabilità P uniforme su Ω . Consideriamo due eventi A e B , di cui l'evento A dipende solo dall'estrazione dal primo mazzo, l'evento B solo dall'estrazione dal secondo mazzo. In altre parole, devono esistere $F, G \subseteq X$ tali che A e B sono della forma:

$$\begin{aligned} A &= \{(i, j) \in \Omega : i \in F\} = F \times X, \\ B &= \{(i, j) \in \Omega : j \in G\} = X \times G, \end{aligned}$$

da cui segue che $A \cap B = F \times G$. Dato che $|X| = 52$, la formula per la cardinalità degli insiemi prodotto dà $|\Omega| = |X|^2 = 52^2$, $|A| = |F||X| = 52|F|$, $|B| = |X||G| = 52|G|$, $|A \cap B| = |F||G|$, quindi

$$P(A \cap B) = \frac{|A \cap B|}{|\Omega|} = \frac{|F||G|}{52^2} = \frac{|A|}{|\Omega|} \frac{|B|}{|\Omega|} = P(A)P(B).$$

Dunque A e B sono indipendenti. Si noti che gli eventi A e B si riferiscono a due *ripetizioni* dello stesso esperimento aleatorio. L'indipendenza esprime il fatto che l'esito di un esperimento non “influenza” l'esito dell'altro esperimento. Questo contesto di *prove indipendenti ripetute*, rilevante in molti aspetti della probabilità e della statistica, è quello in cui la nozione di indipendenza appare in modo naturale. \square

L'indipendenza di due eventi non vuol necessariamente dire che gli eventi “non hanno niente a che fare”, o che “non si influenzano in alcun modo”, come mostra l'esempio seguente.

Esempio 1.58. Gli eventi A e B descritti nei punti seguenti sono indipendenti:

- estraendo una carta da un mazzo di carte da Poker da 52, gli eventi $A :=$ “la carta è di cuori” e $B :=$ “la carta è un 3”;
- lanciando due dadi regolari a 6 facce, gli eventi $A :=$ “la somma vale 7” e $B :=$ “il primo dado dà come risultato 3”;
- se Andrea, Livio e Simone lanciano ciascuno un dado regolare a sei facce, gli eventi $A :=$ “Andrea e Livio ottengono lo stesso risultato” e $B :=$ “Livio e Simone ottengono lo stesso risultato”.

In tutti questi esempi, sebbene ci sia intuitivamente qualche tipo di “influenza” tra gli eventi A e B , essi risultano indipendenti secondo la definizione, perché *il fatto che uno si verifichi non modifica la probabilità dell'altro*.

Verifichiamo l'indipendenza nel terzo caso, lasciando i primi due come esercizio. Posto $X = \{1, 2, 3, 4, 5, 6\}$, prendiamo come spazio campionario

$$\Omega := X \times X \times X = \{(x, y, z) : x, y, z \in \mathbb{N}, 1 \leq x, y, z \leq 6\},$$

munito della probabilità P uniforme. Si noti che

$$A = \{(x, y, z) \in \Omega : x = y\}, \quad B = \{(x, y, z) \in \Omega : y = z\}.$$

Per il principio fondamentale del calcolo combinatorio, gli elementi di A sono determinati dalla scelta delle componenti x e z , per cui $|A| = 6 \cdot 6 = 36$ e, analogamente, $|B| = 36$. Allo stesso modo, $A \cap B = \{(x, y, z) \in \Omega : x = y = z\}$ e dunque $|A \cap B| = 6$. Dato che $|\Omega| = |X|^3 = 6^3 = 216$, segue che

$$P(A) = P(B) = \frac{36}{216} = \frac{1}{6}, \quad P(A \cap B) = \frac{|A \cap B|}{|\Omega|} = \frac{6}{216} = \frac{1}{36} = P(A)P(B),$$

come richiesto. \square

Benché sia ovvio dalla definizione, vale la pena di sottolineare che due eventi indipendenti *non sono disgiunti*, escludendo il caso banale in cui uno dei due eventi abbia probabilità zero. Infatti se A, B sono indipendenti e $P(A) > 0$, $P(B) > 0$, segue che $P(A \cap B) = P(A)P(B) > 0$ e dunque $A \cap B \neq \emptyset$. Detto in modo grossolano, due eventi indipendenti si devono intersecare “nelle giuste proporzioni”.

Osservazione 1.59. Se A e B sono eventi indipendenti, allora anche A^c e B lo sono:

$$\begin{aligned} P(A^c \cap B) &= P(B \setminus (A \cap B)) = P(B) - P(A \cap B) = P(B) - P(A)P(B) \\ &= (1 - P(A))P(B) = P(A^c)P(B). \end{aligned}$$

Analogamente, anche A e B^c sono eventi indipendenti, così come A^c e B^c . Vedremo nella Proposizione 1.64 una formulazione più generale di questa proprietà. \square

Vogliamo ora capire come definire l'indipendenza di più di due eventi. Consideriamo la generalizzazione dell'Esempio 1.57 al caso di 3 mazzi di carte: lo spazio

campionario diventa $\Omega = X^3 = X \times X \times X$ munito della probabilità uniforme P e, dati $F_1, F_2, F_3 \subseteq X$, poniamo $A_i = \{\omega = (\omega_1, \omega_2, \omega_3) \in X^3 : \omega_i \in F_i\}$. In questo modo l'evento A_i dipende solo dall'estrazione dell' i -esimo mazzo e sembra naturale considerare A_1, A_2, A_3 eventi indipendenti. Dato che

$$|A_1 \cap A_2 \cap A_3| = |F_1 \times F_2 \times F_3| = |F_1| |F_2| |F_3|,$$

segue facilmente che

$$P(A_1 \cap A_2 \cap A_3) = P(A_1) P(A_2) P(A_3). \quad (1.45)$$

Inoltre, considerando le tre coppie di eventi $\{A_1, A_2\}$, $\{A_2, A_3\}$, $\{A_1, A_3\}$, analoghi calcoli sulle cardinalità mostrano che

$$\begin{aligned} P(A_1 \cap A_2) &= P(A_1) P(A_2), & P(A_2 \cap A_3) &= P(A_2) P(A_3), \\ P(A_1 \cap A_3) &= P(A_1) P(A_3). \end{aligned} \quad (1.46)$$

Tutto ciò suggerisce che, per definire l'indipendenza di tre eventi A_1, A_2, A_3 , sia opportuno richiedere la proprietà "moltiplicativa" sia per la terna di eventi, cf. (1.45), sia per le coppie, cf. (1.46). È interessante notare che queste due richieste *non* sono implicate l'una dall'altra, come mostrano i seguenti esempi.

Esempio 1.60. Ritornando al terzo caso descritto nell'Esempio 1.58, supponiamo che Andrea, Livio e Simone lancino ciascuno un dado regolare a sei facce e consideriamo gli eventi $A :=$ "Andrea e Livio ottengono lo stesso risultato", $B :=$ "Livio e Simone ottengono lo stesso risultato" e $C :=$ "Andrea e Simone ottengono lo stesso risultato". Abbiamo già mostrato che gli eventi $\{A, B\}$ sono indipendenti. Con identici calcoli (o per simmetria) si mostra che anche $\{A, C\}$ sono indipendenti, così come $\{B, C\}$: valgono dunque le relazioni

$$\begin{aligned} P(A \cap B) &= P(A) P(B), & P(B \cap C) &= P(B) P(C), \\ P(A \cap C) &= P(A) P(C). \end{aligned}$$

Tuttavia, notando che $A \cap B \cap C = A \cap B$, si ha che

$$P(A \cap B \cap C) = P(A \cap B) = \frac{1}{36} \neq \frac{1}{216} = \frac{1}{6^3} = P(A) P(B) P(C). \quad \square$$

Esempio 1.61. Consideriamo l'esperimento aleatorio dato dal lancio di due dadi regolari a sei facce, descritto dallo spazio campionario $\Omega = \{1, 2, \dots, 6\}^2$ munito della probabilità uniforme P . Consideriamo gli eventi

$$\begin{aligned} A &= \text{"il risultato del secondo dado è 1, 2 o 5"} = \{(i, j) \in \Omega : j = 1, 2, \text{ o } 5\}, \\ B &= \text{"il risultato del secondo dado è 4, 5 o 6"} = \{(i, j) \in \Omega : j = 4, 5, \text{ o } 6\}, \\ C &= \text{"la somma dei risultati dei due dadi vale 9"} = \{(i, j) \in \Omega : i + j = 9\}. \end{aligned}$$

Si ha allora

$$P(A \cap B \cap C) = \frac{1}{36} = P(A)P(B)P(C),$$

ma

$$P(A \cap B) = \frac{1}{6} \neq \frac{1}{4} = P(A)P(B),$$

$$P(A \cap C) = \frac{1}{36} \neq \frac{1}{18} = P(A)P(C),$$

$$P(B \cap C) = \frac{1}{12} \neq \frac{1}{18} = P(B)P(C).$$

□

Le considerazioni precedenti motivano la seguente definizione.

Definizione 1.62 (Indipendenza di eventi). Sia $(A_i)_{i \in I}$ una famiglia di eventi in uno spazio di probabilità discreto (Ω, P) , dove l'insieme di indici I è arbitrario. Diremo che tali eventi sono *indipendenti* se per ogni sottoinsieme *finito* $J \subseteq I$ (con $|J| \geq 2$) si ha

$$P\left(\bigcap_{j \in J} A_j\right) = \prod_{j \in J} P(A_j). \quad (1.47)$$

Per l'indipendenza di n eventi A_1, \dots, A_n non basta dunque verificare che

$$P(A_1 \cap A_2 \cap \dots \cap A_n) = P(A_1) \cdot P(A_2) \cdots P(A_n), \quad (1.48)$$

ma occorre mostrare che questa proprietà di fattorizzazione vale anche *per ogni sottofamiglia*, come richiesto dalla relazione (1.47).

In alternativa, mostriamo ora che l'indipendenza di A_1, \dots, A_n è *equivalente* al fatto che valga la relazione (1.48) e, in aggiunta, tutte le relazioni ottenute da questa rimpiazzando alcuni degli eventi A_i con i rispettivi complementari A_i^c .

Proposizione 1.63. Siano A_1, \dots, A_n eventi in uno spazio di probabilità discreto (Ω, P) . Le seguenti affermazioni sono equivalenti:

- (i) gli eventi A_1, \dots, A_n sono indipendenti;
- (ii) per ogni scelta di B_1, \dots, B_n , in cui $B_i = A_i$ oppure $B_i = A_i^c$, si ha

$$P(B_1 \cap B_2 \cap \dots \cap B_n) = P(B_1) \cdot P(B_2) \cdots P(B_n). \quad (1.49)$$

Dimostrazione. Cominciamo mostrando l'implicazione (ii) \Rightarrow (i), procedendo per induzione su $n \geq 2$. Si noti che il passo base $n = 2$ è immediato: basta scegliere $B_1 = A_1$ e $B_2 = A_2$ in (1.49). Concentriamoci dunque sul passo induttivo.

Assumiamo che valga la relazione (1.49) e mostriamo che gli eventi A_1, \dots, A_n sono indipendenti, ossia che vale la relazione (1.47) per ogni $J \subseteq \{1, \dots, n\}$. Il caso

estremo in cui $J = \{1, \dots, n\}$ segue immediatamente da (1.49) scegliendo $B_i = A_i$, quindi possiamo supporre che $|J| \leq n-1$. Per semplificare le notazioni, limitiamoci a considerare $J \subseteq \{1, \dots, n-1\}$ (gli altri casi sono analoghi). Dimostrare la relazione (1.47) per tali J equivale a mostrare l'indipendenza di A_1, \dots, A_{n-1} . Applicando l'ipotesi induttiva, ci basta mostrare che vale la relazione (1.49) con $n-1$ invece di n . Ma questo è semplice: scrivendo $\Omega = A_n \cup A_n^c$, segue immediatamente che

$$P(C) = P(C \cap A_n) + P(C \cap A_n^c),$$

per ogni evento C . Scegliendo $C := B_1 \cap \dots \cap B_{n-1}$, per ogni scelta di $B_i = A_i$ oppure $B_i = A_i^c$, e applicando la relazione (1.49), che vale per ipotesi, otteniamo

$$\begin{aligned} P(B_1 \cap \dots \cap B_{n-1}) &= P(B_1 \cap \dots \cap B_{n-1} \cap A_n) + P(B_1 \cap \dots \cap B_{n-1} \cap A_n^c) \\ &= P(B_1) \cdots P(B_{n-1}) \cdot P(A_n) + P(B_1) \cdots P(B_{n-1}) \cdot P(A_n^c) \\ &= P(B_1) \cdots P(B_{n-1}) \cdot (P(A_n) + P(A_n^c)) = P(B_1) \cdots P(B_{n-1}), \end{aligned}$$

dunque (1.49) vale con $n-1$ invece di n . In questo modo abbiamo completato la dimostrazione dell'implicazione (ii) \Rightarrow (i).

Mostriamo ora che (i) \Rightarrow (ii). Assumiamo che gli eventi A_1, \dots, A_n siano indipendenti e mostriamo che vale la relazione (1.49), per ogni scelta di B_1, \dots, B_n , in cui $B_i = A_i$ oppure $B_i = A_i^c$. Dato che entrambi i membri di (1.49) sono invarianti per permutazioni degli insiemi che vi compaiono, è sufficiente mostrare che

$$P(A_1^c \cap \dots \cap A_k^c \cap A_{k+1} \cap \dots \cap A_n) = P(A_1^c) \cdots P(A_k^c) \cdot P(A_{k+1}) \cdots P(A_n),$$

per ogni $k \in \{0, 1, \dots, n\}$. Il caso $k = 0$ non è altro che la relazione (1.48), che segue immediatamente dall'indipendenza degli eventi A_1, \dots, A_n . Procediamo dunque per induzione su k . Sia $k \geq 1$ e notiamo che $C \cap D \cap A_k^c = (C \cap D) \setminus (C \cap A_k \cap D)$, per ogni scelta degli eventi C, D , dunque

$$P(C \cap A_k^c \cap D) = P(C \cap D) - P(C \cap A_k \cap D).$$

Scegliendo $C := A_1^c \cap \dots \cap A_{k-1}^c$ e $D := A_{k+1} \cap \dots \cap A_n$, si ottiene

$$\begin{aligned} P(A_1^c \cap \dots \cap A_k^c \cap A_{k+1} \cap \dots \cap A_n) &= P(A_1^c \cap \dots \cap A_{k-1}^c \cap A_{k+1} \cap \dots \cap A_n) \\ &\quad - P(A_1^c \cap \dots \cap A_{k-1}^c \cap A_k \cap A_{k+1} \cap \dots \cap A_n), \end{aligned}$$

e usando due volte l'ipotesi induttiva otteniamo

$$\begin{aligned} P(A_1^c \cap \dots \cap A_k^c \cap A_{k+1} \cap \dots \cap A_n) &= P(A_1^c) \cdots P(A_{k-1}^c) P(A_{k+1}) \cdots P(A_n) \\ &\quad - P(A_1^c) \cdots P(A_{k-1}^c) P(A_k) P(A_{k+1}) \cdots P(A_n) \\ &= P(A_1^c) \cdots P(A_{k-1}^c) (1 - P(A_k)) P(A_{k+1}) \cdots P(A_n) \\ &= P(A_1^c) \cdots P(A_{k-1}^c) P(A_k^c) P(A_{k+1}) \cdots P(A_n), \end{aligned}$$

che è proprio quanto dovevamo mostrare. \square

La seguente proposizione mostra che se in un'arbitraria famiglia di eventi indipendenti si rimpiazzano alcuni eventi con i loro complementari, si ottiene ancora una famiglia di eventi indipendenti. La dimostrazione è un semplice corollario della Proposizione 1.63 ed è lasciata come esercizio al lettore.

Proposizione 1.64. *Sia $(A_i)_{i \in I}$ una famiglia di eventi indipendenti. Sia I' un sottoinsieme fissato di I e definiamo*

$$B_i = \begin{cases} A_i^c & \text{se } i \in I' \\ A_i & \text{se } i \in I \setminus I'. \end{cases}$$

Allora $(B_i)_{i \in I}$ è una famiglia di eventi indipendenti.

Esempio 1.65 (Prodotto di Eulero). Diamo una dimostrazione probabilistica di una formula dovuta a Eulero. Per ogni $s \in (1, \infty)$ definiamo la funzione *zeta di Riemann* $\zeta(s)$ mediante la serie (si ricordi la relazione (0.2))

$$\zeta(s) := \sum_{m \in \mathbb{N}} \frac{1}{m^s} \in (0, \infty).$$

La formula prodotto di Eulero afferma che, indicando con $\mathfrak{P} := \{2, 3, 5, \dots\}$ l'insieme dei numeri primi, per ogni $s \in (1, \infty)$ vale l'identità

$$\zeta(s) = \prod_{p \in \mathfrak{P}} \frac{1}{1 - p^{-s}}. \quad (1.50)$$

Sia $\Omega = \mathbb{N}$ e introduciamo su Ω la probabilità P definita sui singoletti (si ricordi il Paragrafo 1.1.3) mediante

$$P(\{n\}) := \frac{1}{\zeta(s)} \frac{1}{n^s}.$$

Per ogni $k \in \mathbb{N}$ sia $M_k := \{k, 2k, 3k, \dots\} = k\mathbb{N} \subseteq \Omega$ l'insieme dei multipli di k . La probabilità dell'evento M_k è data semplicemente da

$$P(M_k) = \sum_{n \in \mathbb{N}} P(\{nk\}) = \sum_{n \in \mathbb{N}} \frac{1}{\zeta(s)} \frac{1}{n^s} \frac{1}{k^s} = \frac{1}{k^s}. \quad (1.51)$$

Osserviamo che $M_p \cap M_q = M_{pq}$ se p e q non hanno divisori comuni. Con un facile argomento induttivo segue che se $p_1, \dots, p_\ell \in \mathfrak{P}$ sono numeri primi distinti si ha

$$M_{p_1} \cap M_{p_1} \cap \dots \cap M_{p_\ell} = M_{p_1 \cdot p_2 \cdots p_\ell},$$

quindi grazie a (1.51)

$$P(M_{p_1} \cap M_{p_1} \cap \dots \cap M_{p_\ell}) = \frac{1}{(p_1 \cdot p_2 \cdots p_\ell)^s} = P(M_{p_1}) \cdot P(M_{p_1}) \cdots P(M_{p_\ell}).$$

Dato che questa formula vale per ogni scelta di $\ell \in \mathbb{N}$ e di $p_1, \dots, p_\ell \in \mathfrak{P}$, abbiamo mostrato che gli eventi $(M_p)_{p \in \mathfrak{P}}$ sono indipendenti. Per la Proposizione 1.64, anche gli eventi $(M_p^c)_{p \in \mathfrak{P}}$ sono indipendenti: di conseguenza, se numeriamo i primi $\mathfrak{P} = \{p_1, p_2, \dots\}$, per ogni $n \in \mathbb{N}$ vale la relazione

$$P(M_{p_1}^c \cap M_{p_2}^c \cap \dots \cap M_{p_n}^c) = P(M_{p_1}^c) \cdot P(M_{p_2}^c) \cdots P(M_{p_n}^c) = \prod_{i=1}^n \left(1 - \frac{1}{p_i^s}\right).$$

Si noti che $(M_{p_1}^c \cap M_{p_2}^c \cap \dots \cap M_{p_n}^c)_{n \in \mathbb{N}}$ è una successione decrescente di eventi, la cui intersezione è $\bigcap_{p \in \mathfrak{P}} M_p^c = \{1\}$, dal momento che 1 è l'unico intero che non è multiplo di nessun numero primo. Grazie alla Proposizione 1.19 (continuità dall'alto della probabilità) possiamo allora scrivere

$$P(\{1\}) = \lim_{n \rightarrow \infty} P(M_{p_1}^c \cap M_{p_2}^c \cap \dots \cap M_{p_n}^c) =: \prod_{p \in \mathfrak{P}} \left(1 - \frac{1}{p^s}\right).$$

Dato che $P(\{1\}) = 1/\zeta(s)$, otteniamo la formula cercata (1.50). □

1.3.4 Prove ripetute e indipendenti

Analizziamo ora un importante paradigma probabilistico, che appare con grande frequenza nelle applicazioni (e negli esercizi...). Molti esperimenti aleatori sono costituiti da n “prove ripetute” che possono avere due soli esiti possibili, detti convenzionalmente “successo” e “insuccesso”. Siamo interessati al caso in cui ciascuna prova ha successo con probabilità $p \in [0, 1]$ e i risultati di prove distinte sono indipendenti. Esempi tipici sono dati dalla ripetizione di uno stesso gioco (il lancio di n dadi o di n monete, l'estrazione con reimmissione di n palline da un'urna, ...) intendendo come “successo” in ciascuna prova un particolare risultato o sottoinsieme di risultati (esce il numero 6, esce “testa”, la pallina è rossa, ...). Tale situazione è detta *schema di prove ripetute e indipendenti*, e può essere formalizzata come segue.

Definizione 1.66 (Prove ripetute e indipendenti). n eventi C_1, \dots, C_n indipendenti e con la stessa probabilità p , in uno spazio di probabilità discreto (Ω, P) , saranno detti *n prove ripetute e indipendenti con probabilità di successo p* .

Intuitivamente, C_i rappresenta l'evento “l' i -esima prova ha successo”.

Questa formalizzazione generale, dall'apparenza piuttosto astratta, permette di calcolare molte probabilità di interesse, che si applicano a svariati esempi concreti. Ad esempio, la probabilità che *nessuna delle n prove abbia successo* vale

$$P(C_1^c \cap \dots \cap C_n^c) = P(C_1^c) \cdots P(C_n^c) = (1 - p)^n,$$

per la Proposizione 1.63. Passando all'evento complementare e applicando le leggi

di De Morgan, la probabilità che *almeno una delle n prove abbia successo* è dunque

$$P(C_1 \cup \dots \cup C_n) = 1 - P(C_1^c \cap \dots \cap C_n^c) = 1 - (1 - p)^n.$$

A partire da questa semplice espressione, si possono dedurre conclusioni a prima vista sorprendenti. Ad esempio, per ogni $p > 0$ fissato, si ha $\lim_{n \rightarrow \infty} (1 - p)^n = 0$, quindi esiste $n_0 = n_0(p) < \infty$ tale che $(1 - p)^{n_0} < 0.0001$, ossia

$$P(C_1 \cup \dots \cup C_{n_0}) \geq 99.99\%.$$

In altri termini, *per quanto piccola (purché strettamente positiva) sia la probabilità p che una singola prova abbia successo, con un numero sufficientemente elevato di prove si è quasi certi di avere almeno un successo.*

Esempio 1.67 (Paradosso della scimmia di Borel). Se una scimmia schiaccia tasti a caso su una tastiera, riuscirà prima o poi a comporre la Divina Commedia?

Diciamo che la tastiera contiene $T := 100$ tasti e supponiamo che ogni tasto abbia la stessa probabilità $q := \frac{1}{T}$ di essere premuto. Assumiamo inoltre che i tasti premuti dalla scimmia vengano “scelti” indipendentemente. La Divina Commedia consta circa di $N := 510\,000$ caratteri (spazi inclusi). Chiamiamo “prima prova” la sequenza dei primi N tasti premuti dalla scimmia, “seconda prova” la sequenza dei successivi N tasti premuti, ecc. e indichiamo con “successo” in ciascuna prova l’evento che la sequenza di tasti premuti coincida *esattamente* con il testo della Divina Commedia. La probabilità p di singolo successo (ossia, che N tasti premuti consecutivamente siano proprio quelli “giusti”) è ridicolmente bassa, ma pur sempre strettamente positiva:

$$p = q^N = \frac{1}{T^N} = 10^{-1020000} > 0.$$

Di conseguenza, se consideriamo un numero di prove pari a n_0 , determinato sopra, *la probabilità che la scimmia componga la Divina Commedia almeno una volta è maggiore di 99.99%!* La “soluzione” del paradosso è che il numero di prove richiesto è *enorme*, ben al di là di qualunque esperimento praticabile: infatti (esercizio) si ha $n_0 \geq 1/p = 10^{1020000}$, mentre l’età dell’universo non supera i 10^{18} secondi.

In una versione lievemente più sofisticata di questo paradosso, proposta dal matematico francese E. Borel, la scimmia prosegue indefinitamente nella sua opera, ossia effettua una successione *infinita* di prove. Questo va al di là degli strumenti che abbiamo attualmente a disposizione (si veda l’Osservazione 1.69 più sotto), ma in questo caso è possibile mostrare che, con probabilità uno, la scimmia compone infinite volte l’intero testo della Divina Commedia! \square

Proseguiamo con il calcolo di probabilità di interesse, relative a uno schema di n prove ripetute e indipendenti con probabilità di successo p . Poniamoci le seguenti domande, per $\ell \in \{1, \dots, n\}$ e $k \in \{0, \dots, n\}$:

- (i) Qual è la probabilità che il primo successo avvenga alla ℓ -esima prova?
- (ii) Qual è la probabilità che esattamente k prove abbiano successo?

Cominciamo con la domanda (i). Per definizione, i successi nelle n prove sono rappresentati da n eventi C_1, \dots, C_n indipendenti e con la stessa probabilità p , in uno spazio di probabilità discreto (Ω, \mathcal{P}) . Dobbiamo calcolare la probabilità dell'evento

$$B_\ell := \text{“il primo successo si verifica nella } \ell\text{-esima prova”} = C_1^c \cap \dots \cap C_{\ell-1}^c \cap C_\ell.$$

Usando l'indipendenza degli eventi coinvolti, per la Proposizione 1.63, si ottiene

$$P(B_\ell) = P(C_1^c) \cdots P(C_{\ell-1}^c) P(C_\ell) = p(1-p)^{\ell-1}. \quad (1.52)$$

Si noti che quest'ultimo valore non dipende da n .

Veniamo ora alla domanda (ii). Vogliamo calcolare la probabilità dell'evento

$$A_k := \text{“esattamente } k \text{ prove hanno successo”}, \quad (1.53)$$

che è rappresentato dal sottoinsieme degli $\omega \in \Omega$ che appartengono esattamente a k degli eventi C_1, \dots, C_n . Il passo fondamentale consiste nell'esprimere A_k come funzione più esplicita di C_1, \dots, C_n , specificando *quali* prove hanno avuto successo. A tal fine, associamo a ogni sottoinsieme $I \subseteq \{1, \dots, n\}$ l'evento $D_I \subseteq \Omega$ definito da “le prove che hanno avuto successo sono quelle con indici in I ”, ossia

$$D_I := \left(\bigcap_{i \in I} C_i \right) \cap \left(\bigcap_{j \in \{1, \dots, n\} \setminus I} C_j^c \right). \quad (1.54)$$

Ad esempio, $D_{\{1, \dots, k\}} = C_1 \cap C_2 \cap \dots \cap C_k \cap C_{k+1}^c \cap C_{k+2}^c \cap \dots \cap C_n^c$ è l'evento “le prove che hanno avuto successo sono esattamente le prime k ”. Possiamo allora scrivere

$$A_k = \bigcup_{I \subseteq \{1, \dots, n\}: |I|=k} D_I \quad \implies \quad P(A_k) = \sum_{I \subseteq \{1, \dots, n\}: |I|=k} P(D_I), \quad (1.55)$$

perché l'unione è disgiunta: $D_I \cap D_{I'} = \emptyset$ per $I \neq I'$.

Determiniamo ora $P(D_I)$. L'evento D_I è dato dall'intersezione di alcuni tra gli eventi C_1, \dots, C_n — quelli con indici in I — e dei complementari dei restanti eventi. Per la Proposizione 1.63, si ha dunque

$$P(D_I) = \left(\prod_{i \in I} P(C_i) \right) \cdot \left(\prod_{j \in \{1, \dots, n\} \setminus I} P(C_j^c) \right) = p^{|I|} \cdot (1-p)^{n-|I|},$$

perché $P(C_i) = p$ e $P(C_j^c) = 1 - P(C_j) = 1 - p$ per ogni i, j . In particolare, se $|I| = k$, il valore di $P(D_I) = p^k(1-p)^{n-k}$ non dipende da I . Dato che ci sono $\binom{n}{k}$ possibili sottoinsiemi $I \subseteq \{1, \dots, n\}$ con $|I| = k$, grazie a (1.55) otteniamo infine

$$P(A_k) = \binom{n}{k} p^k (1-p)^{n-k}. \quad (1.56)$$

Osservazione 1.68. L'espressione (1.56) coincide con le formule (1.30) e (1.34), ottenute rispettivamente negli Esempi 1.38 e 1.41. Più esplicitamente:

- per $p = M/N$, la formula (1.56) fornisce la probabilità che, effettuando n estrazioni con reimmissione da un'urna che contiene N palline, di cui M rosse, esattamente k delle palline estratte siano rosse;
- per $p = 1/r$, la formula (1.56) fornisce la probabilità che, inserendo casualmente n oggetti in r cassetti, ne finiscano k nel primo cassetto.

Naturalmente non si tratta di un caso: entrambi i problemi in esame sono esempi di n prove ripetute e indipendenti con probabilità di successo p (esercizio), considerando gli eventi $C_i = \text{"l}'i\text{-esima pallina estratta è rossa"}$ e, rispettivamente, $C_i = \text{"l}'i\text{-esimo oggetto viene inserito nel primo cassetto"}$. \square

Chiudiamo il paragrafo con una domanda all'apparenza sorprendente: siamo sicuri che *esistano* n prove ripetute e indipendenti con probabilità di successo p ? In altre parole, fissati $n \in \mathbb{N}$ e $p \in [0, 1]$, esiste uno spazio di probabilità discreto (Ω, \mathbf{P}) in cui sono definiti n eventi C_1, \dots, C_n indipendenti e con la stessa probabilità p ?

La risposta è affermativa. Abbiamo già incontrato costruzioni esplicite per alcuni valori di p , ma ora procediamo in modo sistematico, costruendo lo *spazio di probabilità canonico* $(\Omega_n, \mathbf{P}_{n,p})$ per n prove ripetute e indipendenti con probabilità di successo p . Scegliamo come spazio campionario l'insieme

$$\Omega_n := \{0, 1\}^n = \{\omega = (\omega_1, \omega_2, \dots, \omega_n) : \omega_i \in \{0, 1\}\}, \quad (1.57)$$

dove ogni componente ω_i descrive intuitivamente l'esito dell' i -esima prova ($1 = \text{successo}$, $0 = \text{insuccesso}$). Definiamo quindi gli eventi C_1, \dots, C_n mediante

$$C_i := \{\omega = (\omega_1, \dots, \omega_n) \in \Omega_n : \omega_i = 1\}.$$

Mostriamo che per ogni $n \in \mathbb{N}$ e $p \in [0, 1]$ esiste un'unica probabilità $\mathbf{P} = \mathbf{P}_{n,p}$ su Ω_n tale che C_1, \dots, C_n siano eventi indipendenti con la stessa probabilità p .

Per semplificare le notazioni, risulta conveniente definire gli eventi

$$F_i(t) := \{\omega \in \Omega_n : \omega_i = t\}, \quad \text{per } i \in \{1, \dots, n\}, \quad t \in \{0, 1\}, \quad (1.58)$$

così che $F_i(1) = C_i$ e $F_i(0) = C_i^c$. Per ogni $x = (x_1, \dots, x_n) \in \Omega_n$ possiamo scrivere

$$\{x\} = F_1(x_1) \cap F_2(x_2) \cap \dots \cap F_n(x_n).$$

Applichiamo ora la Proposizione 1.63: una probabilità \mathbf{P} su Ω_n è tale che gli eventi C_1, \dots, C_n sono indipendenti *se e solo se* per ogni scelta di x_1, \dots, x_n si ha

$$\mathbf{P}(\{x\}) = \mathbf{P}(F_1(x_1)) \cdot \mathbf{P}(F_2(x_2)) \cdots \mathbf{P}(F_n(x_n)). \quad (1.59)$$

La richiesta che ogni evento C_i abbia probabilità p significa che

$$\mathbf{P}(F_i(t)) = \begin{cases} p & \text{se } t = 1 \\ 1 - p & \text{se } t = 0 \end{cases},$$

ossia, in forma più compatta,

$$\mathbf{P}(F_i(t)) = q(t) := p^t (1 - p)^{1-t}, \quad \forall t \in \{0, 1\}, \quad (1.60)$$

dove conveniamo che $0^0 := 1$. La relazione (1.59) diventa dunque

$$P(\{x\}) = \prod_{i=1}^n q(x_i). \quad (1.61)$$

Risulta dunque determinato il valore di $P(\{x\})$ per ogni $x \in \Omega_n$. Ricordando la Proposizione 1.11, resta solo da mostrare che il membro destro dell'espressione (1.61) è una densità discreta su Ω_n . Procediamo per induzione su $n \in \mathbb{N}$.

- Per $n = 1$ si ha $q(0) + q(1) = (1 - p) + p = 1$, grazie a (1.60).
- Per il passo induttivo $(n - 1) \Rightarrow n$, dato che $\Omega_n = \Omega_{n-1} \times \Omega_1 = \Omega_{n-1} \times \{0, 1\}$, per la somma a blocchi (0.11) si ha

$$\begin{aligned} \sum_{x \in \Omega_n} \left(\prod_{i=1}^n q(x_i) \right) &= \sum_{(x_1, \dots, x_{n-1}) \in \Omega_{n-1}} \left(\prod_{i=1}^{n-1} q(x_i) \sum_{x_n \in \{0, 1\}} q(x_n) \right) \\ &= \sum_{x \in \Omega_{n-1}} \left(\prod_{i=1}^{n-1} q(x_i) (q(0) + q(1)) \right) = \sum_{x \in \Omega_{n-1}} \left(\prod_{i=1}^{n-1} q(x_i) \right) = 1. \end{aligned}$$

In definitiva, abbiamo costruito lo spazio di probabilità discreto $(\Omega_n, P_{n,p})$ cercato: lo spazio campionario è $\Omega_n = \{0, 1\}^n$ e la probabilità $P = P_{n,p}$ è definita sui singoletti $\{x\}$ da (1.61), per ogni $x = (x_1, \dots, x_n) \in \Omega_n$. Questa espressione può essere riscritta in modo più intuitivo nel modo seguente:

$$P_{n,p}(\{x\}) = p^{\sum_{i=1}^n x_i} (1 - p)^{n - \sum_{i=1}^n x_i} = p^k (1 - p)^{n-k}, \quad (1.62)$$

dove $k := \sum_{i=1}^n x_i$ è il numero di “1” che compaiono nella sequenza $x = (x_1, \dots, x_n)$, ossia il “numero di successi”; analogamente, $n - k$ è il “numero di insuccessi”.

Osservazione 1.69. Sarebbe naturale estendere l'analisi a una successione infinita di prove ripetute e indipendenti, i cui esiti sono descritti dallo spazio campionario

$$\Omega = \{0, 1\}^{\mathbb{N}} := \{\omega = \{\omega_i\}_{i \in \mathbb{N}} : \omega_i \in \{0, 1\}\}.$$

Il problema è che l'insieme Ω è *più che numerabile*. La costruzione della probabilità “giusta” su Ω necessita in questo caso degli strumenti avanzati di *teoria della misura*, a cui accenneremo nel Capitolo 5. \square

Osservazione 1.70 (Legge dei grandi numeri). In uno schema di n prove ripetute e indipendenti con probabilità di successo p , con n molto grande, è naturale aspettarsi una frazione di successi dell'ordine di p , ossia un numero k di successi dell'ordine di np , con grande probabilità (si ricordi l'interpretazione frequentista della probabilità). Ora disponiamo degli strumenti per formalizzare questa affermazione. Fissiamo $\varepsilon > 0$ e consideriamo l'evento

$$L_{n,p,\varepsilon} := \text{“la frazione di successi dista da } p \text{ più di } \varepsilon\text{”} = \bigcup_{k \in \{0, \dots, n\} : |\frac{k}{n} - p| > \varepsilon} A_k,$$

dove gli eventi A_k sono definiti informalmente in (1.53), e più precisamente in (1.55). Dato

che gli eventi $(A_k)_{0 \leq k \leq n}$ sono disgiunti, si ottiene dalla formula (1.56)

$$P(L_{n,p,\varepsilon}) = \sum_{k \in \{0, \dots, n\}: |\frac{k}{n} - p| > \varepsilon} \binom{n}{k} p^k (1-p)^{n-k}.$$

Abbiamo ottenuto un'espressione forse non proprio amichevole, ma relativamente esplicita, per la probabilità dell'evento di interesse, ed è possibile dimostrare che

$$\lim_{n \rightarrow \infty} P(L_{n,p,\varepsilon}) = 0, \quad \forall p \in [0, 1], \quad \forall \varepsilon > 0. \quad (1.63)$$

Questa è una formalizzazione matematicamente precisa dell'affermazione iniziale “per n grande, la frazione di successi è dell'ordine di p , con grande probabilità”.

Dimosteremo la relazione (1.63) nel Capitolo 7, come caso particolare di un teorema molto importante, la *legge dei grandi numeri*. Sebbene siano possibili dimostrazioni analitiche dirette, le tecniche probabilistiche che svilupperemo permetteranno un'analisi più elegante e trasparente. \square

1.3.5 Esempi e paradossi sul condizionamento

La probabilità condizionale è una delle principali sorgenti di “paradossi” in teoria della probabilità, intendendo con questo termine risultati veri ma poco intuitivi. Al di là del loro interesse intrinseco, lo studio di questi problemi è utile perché evidenzia i punti delicati che si nascondono dietro la nozione di condizionamento e, soprattutto, dietro la sua interpretazione.

In questo paragrafo discutiamo con un certo dettaglio due problemi classici, noti come paradosso di Monty Hall (Esempio 1.71) e paradosso dei figli (Esempio 1.72). Altri problemi si possono trovare tra gli esercizi. Chiudiamo quindi il paragrafo con due problemi liberamente ispirati a casi giudiziari reali (Esempi 1.73 e 1.74), che mostrano quanto la mancata comprensione della nozione di condizionamento possa condurre a conclusioni errate.

Esempio 1.71 (Paradosso di Monty Hall). Il presentatore di un gioco a premi vi propone di scegliere una di tre buste chiuse. Delle tre buste, una contiene un premio mentre le altre due sono vuote. Dopo che avete effettuato la scelta, il presentatore apre una delle due buste rimaste, mostrando che è vuota, e vi propone di cambiare la busta che avete scelto con quella rimanente. Che cosa vi conviene fare?

A dispetto della formulazione elementare, questo problema nasconde diverse insidie, che sono per certi versi paradigmatiche dei problemi che coinvolgono la probabilità condizionale. Per cominciare, la strategia con cui il presentatore sceglie la busta da aprire non è esplicitamente dichiarata nella formulazione del problema: come vedremo, strategie diverse conducono a modelli probabilistici distinti, che determinano *risposte diverse al problema*.

Numeriamo le buste da 1 a 3 e consideriamo innanzitutto la seguente strategia: il presentatore guarda di nascosto le due buste che gli sono rimaste in mano e ne apre *sempre* una vuota (nel caso in cui entrambe le buste siano vuote, diciamo che apre quella con numero più basso). Allora la risposta è che *vi conviene cambiare bu-*

sta, perché la probabilità di trovare il premio passa da $\frac{1}{3}$ a $\frac{2}{3}$. Per convincersene, notiamo che l'esperimento aleatorio può essere descritto dallo spazio campionario $\Omega = \{1, 2, 3\} \times \{1, 2, 3\} = \{\omega = (i, j) : 1 \leq i, j \leq 3\}$, dove $\omega = (i, j)$ significa che il premio è nella busta i e voi scegliete inizialmente la busta j . Definiamo gli eventi

$$A_k := \text{“il premio è nella busta } k\text{”} = \{(i, j) \in \Omega : i = k\},$$

$$B_\ell := \text{“voi scegliete inizialmente la busta } \ell\text{”} = \{(i, j) \in \Omega : j = \ell\},$$

dove $k, \ell \in \{1, 2, 3\}$. Quale probabilità P è sensato mettere su Ω ? Innanzitutto è ragionevole assumere che il premio sia in ciascuna busta con la stessa probabilità, ossia $P(A_1) = P(A_2) = P(A_3) = \frac{1}{3}$. In secondo luogo, è naturale supporre che la vostra scelta iniziale della busta sia indipendente da quale busta contenga il premio, ossia gli eventi A_k e B_ℓ siano indipendenti per ogni $k, \ell \in \{1, 2, 3\}$. Di conseguenza, una volta specificati i valori di $p_\ell := P(B_\ell)$ per $k = 1, 2, 3$ – per esempio $p_1 = p_2 = p_3 = \frac{1}{3}$ se scegliete inizialmente una busta “a caso” – risulta determinata la probabilità P su Ω , dato che $P(\{(k\ell)\}) = P(A_k \cap B_\ell) = P(A_k)P(B_\ell) = \frac{1}{3}p_\ell$. Possiamo finalmente determinare la probabilità degli eventi a cui siamo interessati, ossia

$$F := \text{“tenendo la busta inizialmente scelta, trovate il premio”}$$

$$G := \text{“cambiando la busta, trovate il premio”}. \quad (1.64)$$

È chiaro che possiamo riformulare

$$F = \text{“il premio è nella busta scelta inizialmente”} = \{(i, j) \in \Omega : i = j\},$$

pertanto

$$P(F) = P(\{(1, 1)\}) + P(\{(2, 2)\}) + P(\{(3, 3)\}) = \frac{1}{3}(p_1 + p_2 + p_3) = \frac{1}{3}.$$

L'osservazione cruciale è che, *cambiando busta, trovate il premio se e solo se esso non è nella busta da voi scelta inizialmente*: in tal caso infatti il premio è in una delle due buste rimaste inizialmente in mano al presentatore, e lui provvede ad aprire quella vuota. Questo significa che $G = F^c$ e dunque

$$P(G) = P(F^c) = 1 - P(F) = \frac{2}{3},$$

come annunciato. Si noti che il risultato non dipende dalle probabilità p_k con cui effettuate la vostra scelta iniziale, come è peraltro intuitivo.

Mentre è intuitivamente chiaro che la probabilità di trovare il premio tenendo la busta inizialmente scelta vale $\frac{1}{3}$, il fatto che la probabilità diventi $\frac{2}{3}$ se si cambia busta può apparire a prima vista sorprendente. Per convincersi intuitivamente di questo fatto, suggeriamo di riflettere sull'osservazione sopra esposta: cambiando busta, si trova il premio *ogniquale volta* esso non è nella busta inizialmente scelta, dunque in media due volte su tre. Chi non fosse convinto, può provare un esperimento concre-

to ripetendo il gioco molte volte, eventualmente con l'ausilio di un computer. Per esercitare l'intuizione su problemi analoghi, si vedano gli Esercizi 1.13 e 1.14.

Per concludere, mostriamo come la risposta cambia in funzione della strategia del presentatore. Supponiamo ad esempio che il presentatore scelga casualmente una delle due buste che gli sono rimaste in mano: se la busta scelta contiene il premio, il gioco finisce; se invece è vuota, vi viene proposta la possibilità di cambiare la busta che avete scelto inizialmente con quella rimanente. Mostriamo che in questo caso è indifferente cambiare o non cambiare busta: in entrambi i casi, se il presentatore apre una busta vuota, la probabilità (condizionale) di trovare il premio vale $\frac{1}{2}$.

Per formalizzare il problema, conviene “arricchire” lo spazio campionario Ω introdotto in precedenza, in modo che i suoi elementi descrivano, oltre alla busta che contiene il premio e a quella scelta inizialmente, anche la busta aperta dal presentatore. Una scelta naturale è dunque $\hat{\Omega} := \{\omega = (i, j, k) : i, j, k \in \{1, 2, 3\}, k \neq j\}$. Preferiamo tuttavia alleggerire le notazioni, immaginando che voi scegliate sempre la busta numero 1 (come abbiamo visto in precedenza, e come è intuitivo, la strategia con cui viene scelta la busta iniziale risulta alla fine irrilevante). Poniamo dunque

$$\hat{\Omega} := \{\omega = (i, k) : i \in \{1, 2, 3\}, k \in \{2, 3\}\},$$

dove $\omega = (i, k)$ significa che il premio è nella busta i e il presentatore apre la busta k (voi scegliete inizialmente la busta 1, da cui la restrizione $k \in \{2, 3\}$). Per determinare la probabilità \hat{P} da mettere su $\hat{\Omega}$, definiamo per $\ell \in \{1, 2, 3\}$ l'evento

$$A_\ell = \text{“il premio è nella busta } \ell\text{”} = \{(i, k) \in \hat{\Omega} : i = \ell\},$$

e richiediamo, come in precedenza, che $\hat{P}(A_1) = \hat{P}(A_2) = \hat{P}(A_3) = \frac{1}{3}$. Inoltre, definendo per $m \in \{2, 3\}$ l'evento

$$C_m := \text{“il presentatore sceglie la busta } m\text{”} = \{(i, k) \in \hat{\Omega} : k = m\},$$

richiediamo che $\hat{P}(C_m|A_\ell) = \frac{1}{2}$ per ogni $\ell \in \{1, 2, 3\}$ e $m \in \{2, 3\}$, perché il presentatore apre una busta “a caso”. Ciò significa che per ogni $\omega = (i, k) \in \hat{\Omega}$

$$\hat{P}(\{\omega\}) = \hat{P}(\{(i, k)\}) = P(A_i \cap C_k) = P(A_i)P(C_k|A_i) = \frac{1}{3} \frac{1}{2} = \frac{1}{6},$$

ossia \hat{P} è la probabilità uniforme su $\hat{\Omega}$.

Introduciamo ora l'evento $D := \text{“la busta aperta dal presentatore è vuota”}$, ossia

$$D := \{(i, k) \in \hat{\Omega} : k \neq i\} = \{(1, 2), (1, 3), (2, 3), (3, 2)\}.$$

Dato che inizialmente scegliete la busta numero 1, l'evento $F := \text{“tenendo la busta inizialmente scelta, trovate il premio”}$ è dato da

$$F = \{(i, k) \in \hat{\Omega} : i = 1\} = \{(1, 2), (1, 3)\},$$

pertanto

$$\hat{P}(F|D) = \frac{\hat{P}(F \cap D)}{\hat{P}(D)} = \frac{|F \cap D|}{|D|} = \frac{2}{4} = \frac{1}{2}.$$

Se cambiate busta, trovate il premio se e solo se il premio non è nella busta numero 1 né nella busta aperta dal presentatore. Quindi l'evento $G :=$ “cambiando busta, trovate il premio” è dato da

$$G = \{(i, k) \in \hat{\Omega} : i \neq 1, k \neq i\} = \{(2, 3), (3, 2)\},$$

e la sua probabilità condizionale vale

$$\hat{P}(G|D) = \frac{\hat{P}(G \cap D)}{\hat{P}(D)} = \frac{|G \cap D|}{|D|} = \frac{2}{4} = \frac{1}{2}.$$

in definitiva, se il presentatore apre una busta vuota, le probabilità (condizionale) di trovare il premio mantenendo la busta scelta inizialmente, oppure cambiando busta, valgono entrambe $\frac{1}{2}$, come annunciato in precedenza. \square

Esempio 1.72 (Paradosso dei figli). Una coppia ha due figli(e). Assumendo che ciascun figlio possa essere maschio o femmina con la stessa probabilità, indipendentemente dal sesso dell'altro figlio, rispondiamo alle domande seguenti:

- (i) Se il primogenito è maschio, qual è la probabilità che i figli siano entrambi maschi?
- (ii) Se il secondogenito è maschio, qual è la probabilità che i figli siano entrambi maschi?
- (iii) Se almeno un figlio è maschio, qual è la probabilità che i figli siano entrambi maschi?

Per descrivere il sesso dei due figli, introduciamo lo spazio campionario $\Omega = \{mm, mf, fm, ff\}$, dove ab indica che il primogenito è di sesso a e il secondogenito di sesso b . È facile convincersi che le ipotesi di indipendenza e di equiprobabilità del sesso dei figli corrispondono a munire Ω della probabilità uniforme, cioè $P(\{mm\}) = P(\{mf\}) = P(\{fm\}) = P(\{ff\}) = \frac{1}{4}$. Introducendo gli eventi

$$A := \text{“il primogenito è maschio”} = \{mm, mf\},$$

$$B := \text{“il secondogenito è maschio”} = \{mm, fm\},$$

e interpretando le domande poste come probabilità condizionali, le risposte si ottengono con semplici calcoli: essendo $A \cup B = \{mm, mf, fm\}$ e $A \cap B = \{mm\}$,

- (i) $P(A \cap B|A) = \frac{P(A \cap B)}{P(A)} = \frac{|A \cap B|}{|A|} = \frac{1}{2};$
- (ii) $P(A \cap B|B) = \frac{P(A \cap B)}{P(B)} = \frac{|A \cap B|}{|B|} = \frac{1}{2};$
- (iii) $P(A \cap B|A \cup B) = \frac{P((A \cap B) \cap (A \cup B))}{P(A \cup B)} = \frac{P(A \cap B)}{P(A \cup B)} = \frac{|A \cap B|}{|A \cup B|} = \frac{1}{3}.$

Sebbene i calcoli siano del tutto elementari, la risposta all'ultima domanda è a prima vista sorprendente, per il contrasto con le prime due, e merita una discussione. Il punto chiave sta nel significato da attribuire all'espressione "se almeno un figlio è maschio". Ricordiamoci dell'interpretazione frequentista della probabilità (condizionale e non): selezionando in modo indipendente un grande numero di famiglie con due figli, all'incirca un quarto di queste sarà del tipo mm (ossia avrà entrambi i figli maschi), un quarto sarà mf , un quarto sarà fm e un quarto sarà ff . Se ci restringiamo alle famiglie in cui almeno un figlio è maschio, otteniamo un sottoinsieme costituito dai tre tipi mm , mf e fm , tutti all'incirca con la stessa numerosità: di conseguenza, delle famiglie in cui almeno un figlio è maschio, all'incirca una su tre ha entrambi i figli maschi e due su tre hanno invece un maschio e una femmina, in accordo con la risposta trovata sopra.

L'argomento appena esposto contribuisce a chiarire il significato da attribuire alla probabilità condizionale e, allo stesso tempo, ne mette in evidenza le importanti limitazioni. Nella soluzione del problema abbiamo tradotto automaticamente l'informazione "sappiamo che almeno un figlio è maschio" con il concetto matematico di condizionamento. Tuttavia, affinché questo procedimento sia giustificato, occorre che il processo con cui tale informazione è stata ottenuta corrisponda realmente a un "campionamento uniforme" dell'insieme delle famiglie in cui almeno un figlio è maschio. Ad esempio, immaginiamo di accedere ai dati del censimento Istat e di costruire un database contenente tutte e sole le coppie italiane con due figli, di cui almeno uno maschio: se si sceglie una famiglia a caso nel database, la probabilità che entrambi i figli siano maschi corrisponde effettivamente (all'incirca) a $\frac{1}{3}$.

Il problema è che esistono diversi modi alternativi molto naturali in cui si può venire a conoscenza dell'informazione che almeno un figlio è maschio, che *non* possono essere tradotti con il semplice condizionamento. Ad esempio, data una coppia con due figli(e), immaginiamo che venga scelto uno dei due figli a caso e si scopra che è un maschio. Sulla base di questa informazione, qual è la probabilità che anche l'altro figlio sia maschio? Mostriamo che la risposta in questo caso vale $\frac{1}{2}$.

Ingrandiamo lo spazio campionario Ω in modo da descrivere anche quale figlio viene scelto, ponendo

$$\Omega' := \Omega \times \{1, 2\} = \{mm1, mm2, mf1, mf2, fm1, fm2, ff1, ff2\},$$

dove $mm1$ significa che il primogenito e il secondogenito sono maschi e viene scelto il primogenito, ecc. Si noti che gli eventi prima introdotti $A :=$ "il primogenito è maschio" e $B :=$ "il secondogenito è maschio" diventano ora

$$A = \{mm1, mm2, mf1, mf2\}, \quad B = \{mm1, mm2, fm1, fm2\}.$$

Per determinare la probabilità P' da mettere su Ω' , è naturale richiedere che

$$P'(\{mm1, mm2\}) = P'(\{mf1, mf2\}) = P'(\{fm1, fm2\}) = P'(\{ff1, ff2\}) = \frac{1}{4},$$

dal momento che le probabilità dei sessi dei figli presenti nella famiglia sono le stesse

di prima. Infine, visto che il figlio viene scelto “a caso”, imponiamo che

$$P'(\{mm1\}|\{mm1,mm2\}) = \frac{1}{2},$$

da cui segue che

$$P'(\{mm1\}) = P'(\{mm1,mm2\})P'(\{mm1\}|\{mm1,mm2\}) = \frac{1}{4} \frac{1}{2} = \frac{1}{8}.$$

Con analoghi argomenti si mostra che $P'(\{\omega\}) = \frac{1}{8}$ per ogni $\omega \in \Omega'$, dunque P' è la probabilità uniforme su Ω' . Introducendo l'evento

$$C := \text{“il figlio scelto è maschio”} = \{mm1,mm2,mf1,mf2\},$$

otteniamo infine la probabilità che entrambi i figli siano maschi, sapendo che quello scelto è maschio:

$$P'(A \cap B|C) = \frac{P'(A \cap B \cap C)}{P'(C)} = \frac{|A \cap B \cap C|}{|C|} = \frac{|A \cap B|}{|C|} = \frac{2}{4} = \frac{1}{2}.$$

L'informazione “un figlio scelto a caso risulta maschio” non corrisponde dunque al semplice condizionamento rispetto all'evento “almeno un figlio è maschio”. \square

Esempio 1.73 (Un caso giudiziario). Una donna venne assassinata, il marito era il principale sospettato. Nel corso delle indagini si scoprì che il marito aveva più volte picchiato la moglie. L'accusa affermò che questo fatto rappresentava un importante indizio per la colpevolezza. La difesa ribatté che, secondo i dati forniti dalla Polizia di Stato, tra gli uomini che picchiano le loro mogli, solo 1 su 10000 finisce poi per assassinarla. Pertanto tale dato contribuisce solo in modo molto marginale alla tesi di colpevolezza. In primo grado il giudice accolse la tesi della difesa.

In secondo grado l'accusa rilevò il seguente errore nell'argomento della difesa. La frazione $\frac{1}{10000}$ fornisce una stima della probabilità che una donna venga ammazzata dal marito condizionalmente al fatto che il marito la picchiasse. Ma noi non solo sappiamo che il marito la picchiava, ma anche che qualcuno ha effettivamente assassinato la donna! Quindi la quantità che occorre calcolare è la probabilità che “una donna venga ammazzata dal marito” condizionalmente all'evento “la donna veniva picchiata dal marito ed è stata ammazzata da qualcuno”.

Per formalizzare il problema, consideriamo una popolazione numerosa e sufficientemente omogenea di donne sposate, e consideriamo i seguenti eventi, relativi ad un individuo casualmente scelto in questa popolazione:

- $A =$ “la donna è stata picchiata dal marito”;
- $B =$ “la donna viene assassinata dal marito”;
- $C =$ “la donna viene assassinata da una persona diversa dal marito”.

Si noti che l'evento “la donna viene assassinata da qualcuno” corrisponde a $B \cup C$, pertanto la probabilità che desideriamo calcolare è $p := P(B|A \cap (B \cup C))$. Notando

che $B \cap C = \emptyset$, possiamo riscrivere questa probabilità nella forma più conveniente

$$\begin{aligned} p &= P(B|A \cap (B \cup C)) = \frac{P(B \cap A \cap (B \cup C))}{P(A \cap (B \cup C))} = \frac{P(B \cap A)}{P((B \cup C) \cap A)} \\ &= \frac{P(B|A)}{P(B \cup C|A)} = \frac{P(B|A)}{P(B|A) + P(C|A)} = \frac{1}{1 + \frac{P(C|A)}{P(B|A)}}. \end{aligned}$$

Questa relazione mostra che, per valutare la probabilità cercata p , non conta la grandezza assoluta di $P(B|A)$, ma la sua grandezza *relativa* rispetto a $P(C|A)$.

Sappiamo che $P(B|A) = \frac{1}{10000}$, mentre non conosciamo il valore di $P(C|A)$. Un'informazione utile è la seguente: secondo i dati della Polizia, nella totalità della popolazione circa una donna su 100000 viene assassinata (dal marito o da qualcun altro), cioè possiamo assumere che $P(B \cup C) = \frac{1}{100000}$. Questo dato permette di dare una stima di $P(C|A)$. Infatti è ragionevole assumere che tutte le donne, vengano o meno picchiate dal marito, abbiano la stessa probabilità di essere assassinate da una persona *diversa* dal marito, ossia $P(C|A) = P(C)$ (in altri termini, gli eventi A e C sono indipendenti). D'altro canto $P(C) \leq P(B \cup C)$, per cui mettendo insieme le precedenti considerazioni otteniamo la stima

$$\frac{P(C|A)}{P(B|A)} \leq \frac{P(B \cup C)}{P(B|A)} = \frac{\frac{1}{100000}}{\frac{1}{10000}} = \frac{1}{10} \implies p = \frac{1}{1 + \frac{P(C|A)}{P(B|A)}} \geq \frac{1}{1 + \frac{1}{10}} = \frac{10}{11}.$$

Quindi, con la sola informazione che avesse picchiato la moglie, il marito è l'assassino con probabilità $p \geq \frac{10}{11} \simeq 91\%$! Il marito fu poi condannato... \square

Esempio 1.74 (Un altro caso giudiziario). Dalle indagini relative ad un omicidio è emerso che il colpevole possiede un determinato set di caratteristiche (per es. capelli rossi, zoppicante, ecc.) che lo rendono piuttosto raro: si stima che una frazione $p \ll 1$ di popolazione possieda tali caratteristiche. La città in cui si è svolto l'attentato ha n abitanti, con $np \simeq 0.05$. Una ricerca su un database di individui schedati ha identificato un unico individuo che possiede il set di caratteristiche. Per affermar la colpevolezza di tale individuo, l'accusa argomenta come segue:

La probabilità che in città vi siano almeno due individui con il set di caratteristiche cercato è circa $(np)^2 = 0.0025$. Pertanto, l'individuo trovato è con probabilità $1 - (np)^2 = 0.9975$ l'unico con tale set di caratteristiche. La sua colpevolezza è accertata con probabilità 99.75%.

Questo argomento è sbagliato, in particolare la conclusione. Vediamo perché. Una ragionevole assunzione è che ogni individuo possieda il set di caratteristiche cercato con probabilità p , indipendentemente dagli altri. Pertanto, per quanto visto nel Paragrafo 1.3.4, se A_k denota l'evento "in città ci sono esattamente k individui con il set di caratteristiche", si ha

$$P(A_k) = \binom{n}{k} p^k (1-p)^{n-k}.$$

In particolare

$$P(A_0) = (1 - p)^n, \quad P(A_1) = np(1 - p)^{n-1}, \quad (1.65)$$

quindi la probabilità che vi siano almeno due individui con i requisiti richiesti è

$$P[(A_0 \cup A_1)^c] = 1 - (1 - p)^n - np(1 - p)^{n-1} \simeq \frac{n(n-1)}{2} p^2 \simeq \frac{1}{2} (np)^2, \quad (1.66)$$

dove la prima approssimazione si ottiene con uno sviluppo di Taylor al secondo ordine della funzione $f(p) = 1 - (1 - p)^n - np(1 - p)^{n-1}$ attorno a $p = 0$. La stima dell'accusa va dunque corretta di un fattore $\frac{1}{2}$. Ma non è certo questo l'errore più rilevante, anche perché la correzione rende l'argomento ancor più stringente.

Il punto fondamentale è che l'accusa non ha tenuto conto del fatto che un individuo con il set di caratteristiche dato è *già stato trovato*. La quantità probante non è dunque la “probabilità che vi siano in città almeno due individui con i requisiti richiesti”, bensì la “probabilità che vi siano in città almeno due individui con i requisiti richiesti *condizionalmente* al fatto che la ricerca nel database ne ha identificato uno”. Considerando gli eventi

$B :=$ “in città vi sono almeno due individui con i requisiti richiesti”,

$C :=$ “nel database c'è esattamente un individuo con i requisiti richiesti”,

occorre dunque calcolare $P(B|C)$. Notiamo che $B \cap C = B' \cap C$, dove

$B' :=$ “tra gli individui *non* schedati almeno uno ha le caratteristiche richieste”.

Possiamo inoltre assumere che gli eventi B' e C , riferendosi a gruppi distinti di individui, siano indipendenti. Ci riduciamo dunque a calcolare

$$P(B|C) = \frac{P(B \cap C)}{P(C)} = \frac{P(B' \cap C)}{P(C)} = \frac{P(B')P(C)}{P(C)} = P(B').$$

Se indichiamo con m il numero degli individui della città inseriti nel database, la prima formula in (1.65) con $n - m$ invece che n dà

$$P(B|C) = P(B') = 1 - (1 - p)^{n-m} \simeq (n - m)p \simeq np = 0.05, \quad (1.67)$$

dove abbiamo fatto l'ipotesi, molto verosimile, che $m \ll n$. Dunque, sulla base delle conoscenze acquisite, la probabilità che l'individuo trovato sia l'unico con le caratteristiche date, e quindi che sia colpevole, è

$$P(B^c|C) = 95\%,$$

probabilmente non sufficiente a fugare “ogni ragionevole dubbio”.

Per apprezzare la sottigliezza della questione, consideriamo il seguente quesito. Senza aver condotto alcuna ricerca su database, qual è la probabilità che l'autore dell'omicidio *non* sia l'unico individuo in città con le caratteristiche richieste? In questo caso, l'unica informazione disponibile è che “esiste almeno un individuo in città con

le caratteristiche richieste” (l’assassino), che corrisponde all’evento A_0^c . Dobbiamo pertanto calcolare $P(B|A_0^c)$. Osservando che $B = (A_0 \cup A_1)^c$, si ottiene

$$P(B|A_0^c) = P((A_0 \cup A_1)^c | A_0^c) = \frac{P((A_0 \cup A_1)^c)}{P(A_0^c)}.$$

Grazie alla relazione (1.66) si ha $P((A_0 \cup A_1)^c) \simeq \frac{1}{2}(np)^2$. Analogamente, grazie a (1.65), si ha

$$P(A_0^c) = 1 - (1 - p)^n \simeq np,$$

pertanto

$$P(B|A_0^c) \simeq \frac{1}{2}np = 0.025,$$

che differisce di un fattore $\frac{1}{2}$ dal risultato (1.67) trovato sopra!

Dunque, i risultati ottenuti mostrano che l’informazione “la ricerca nel database ha individuato un individuo con le caratteristiche cercate” non corrisponde al semplice condizionamento rispetto all’evento “in città esiste almeno un individuo con le caratteristiche cercate”. Il lettore attento potrà trovare analogie con quanto visto nell’Esempio 1.72 □

Esercizi

Esercizio 1.9. Si mostri, con degli esempi, che entrambe le disuguaglianze $P(A|B) > P(A)$ e $P(A|B) < P(A)$ sono possibili.

Esercizio 1.10. Siano A, B, C tre eventi in uno spazio di probabilità discreto (Ω, P) . Si assuma che A, B, C siano indipendenti. Si mostri che

- (i) $A \cap B$ è indipendente da C .
- (ii) $A \cup B$ è indipendente da C .

Esercizio 1.11. Siano A_1, A_2, \dots, A_n eventi indipendenti tali che $P(A_1 \cup A_2 \cup \dots \cup A_n) = 1$. Si mostri che esiste $k \in \{1, 2, \dots, n\}$ tale che $P(A_k) = 1$.

Esercizio 1.12. Siano assegnati tre numeri: $\alpha_1, \alpha_2 \in [0, 1]$ e $\beta \in (0, 1)$. Si mostri che esiste uno spazio di probabilità discreto (Ω, P) contenente due eventi A, B tali che

$$P(B) = \beta, \quad P(A|B) = \alpha_1, \quad P(A|B^c) = \alpha_2.$$

[Sugg. Si consideri $\Omega = \{ab, a\bar{b}, \bar{a}b, \bar{a}\bar{b}\} = \{a, \bar{a}\} \times \{b, \bar{b}\}$, definendo $A := \{ab, a\bar{b}\}$, $B := \{ab, \bar{a}b\}$ e mostrando che esiste un’unica probabilità P su Ω che soddisfa le specifiche richieste.]

Esercizio 1.13 (Paradosso dei tre prigionieri). Tre prigionieri (A, B, C) sono condannati all’impiccagione. Il sovrano decide di graziare uno dei tre scelto a caso, ma il

nome del fortunato verrà comunicato soltanto alla vigilia dell'esecuzione. Il prigioniero A si avvicina al secondino, che conosce il nome del graziato, e gli dice: "Per favore, comunicami un nome, tra B e C, che verrà sicuramente impiccato. È noto che almeno uno di loro due sarà impiccato, pertanto non mi fornisca alcuna informazione dicendomelo". Il secondino ci pensa, trova l'argomento sensato e risponde: "B verrà impiccato". A questo punto A esclama: "Evviva! Visto che B verrà impiccato, restiamo in gioco solo io e C, pertanto ho il 50% di probabilità di essere graziato, mentre in precedenza ne avevo solo $\frac{1}{3}$." Questo argomento è corretto?

Esercizio 1.14 (Paradosso delle tre carte). Infilo in una busta tre carte: una ha entrambe le facce rosse, una le ha entrambe nere, una ha una faccia rossa e una nera. Con gli occhi chiusi, pesco una carta a caso e la depongo sul tavolo su una faccia a caso, quindi apro gli occhi. Se la faccia che vedo è rossa, qual è la probabilità che anche l'altra faccia sia rossa?

1.4 Esercizi di riepilogo

Esercizio 1.15. Siano A, B eventi. Ricordando che $A \triangle B := (A \cup B) \setminus (A \cap B)$, si mostri che

$$P(A \triangle B) = P(A) + P(B) - 2P(A \cap B).$$

Siano ora A, B, C tre eventi. Si mostri che

$$P(A \triangle C) \leq P(A \triangle B) + P(B \triangle C).$$

Esercizio 1.16. Siano A e B due eventi arbitrari di uno spazio di probabilità (Ω, P) . Si dimostri la disuguaglianza

$$P(A \cap B) \geq P(A) + P(B) - 1.$$

Si mostri quindi per induzione che, per ogni $n \geq 2$ e per ogni scelta degli eventi A_1, A_2, \dots, A_n , si ha

$$P(A_1 \cap A_2 \cap \dots \cap A_n) \geq \sum_{i=1}^n P(A_i) - (n-1).$$

Esercizio 1.17. Da un mazzo di 52 carte da Poker si estraggono, a caso, tre carte. Si calcoli la probabilità che:

- (i) tra le carte estratte vi sia almeno un asso;
- (ii) le tre carte estratte siano di tre semi diversi;
- (iii) almeno due delle carte estratte abbiano lo stesso numero o figura.

Esercizio 1.18. Un mazzo di 52 carte da Poker viene diviso a metà. Si determini la probabilità che ognuna delle due parti contenga carte rosse e nere in egual numero.

Esercizio 1.19. Una lotteria emette n biglietti, di cui $m < n$ sono vincenti. Qual è la probabilità che un possessore di r biglietti ne abbia almeno uno di vincente?

Esercizio 1.20. Si mescolano n paia di guanti, che vengono poi distribuiti a caso a n persone, due guanti per ciascuno. Qual è la probabilità che ognuno riceva un guanto per la mano destra e uno per la sinistra?

Esercizio 1.21. Si esegua una permutazione casuale dei numeri $\{1, 2, \dots, n\}$. Qual è la probabilità che 1 e 2 siano successivi anche dopo la permutazione?

Esercizio 1.22. Sia S_n l'insieme delle permutazioni di $\{1, 2, \dots, n\}$. Dati $\sigma \in S_n$ e $I \subseteq \{1, 2, \dots, n\}$, diciamo che l'insieme I è *stabile* per σ se $\sigma(i) \in I$ per ogni $i \in I$. Denotiamo con $A_I \subseteq S_n$ l'insieme delle permutazioni per le quali I è stabile. Indicando con P la probabilità uniforme su S_n , si calcoli $P(A_I)$.

Esercizio 1.23. Si eseguano n estrazioni casuali *con reimmissione* da un'urna contenente $2n$ oggetti distinti. Si determini la probabilità p_n che gli n oggetti estratti siano tutti diversi. Usando la formula di Stirling (1.20), si determini quindi il comportamento asintotico di p_n per $n \rightarrow +\infty$, mostrando che $p_n \sim c\rho^n$ (nel senso che $\lim_{n \rightarrow \infty} p_n/(c\rho^n) = 1$) e calcolando i valori di c e ρ .

Esercizio 1.24. Da un'urna contenente n palline di cui k rosse e $n - k$ verdi, con $1 \leq k \leq n - 1$, si estrae una pallina e quindi, senza rimmetterla nell'urna, si estrae una seconda pallina. Si considerino gli eventi informalmente descritti da

$$\begin{aligned} A_1 &:= \text{"la prima pallina estratta è rossa"}, \\ A_2 &:= \text{"la seconda pallina estratta è rossa"} \end{aligned}$$

Si mostri che gli eventi A_1 e A_2 non sono indipendenti.

Esercizio 1.25. Si voglia illuminare una stanza con un certo numero di lampadine. Assumiamo che la probabilità che una lampadina sopravviva almeno n giorni vale p^n , con $p = 0.9$. Si può ritenere che le lampadine si comportino in modo indipendente. Quante lampadine occorre installare affinché, con probabilità almeno 0.99, dopo 10 giorni vi sia almeno una lampadina funzionante?

Esercizio 1.26. Si mostri la seguente *formula di disintegrazione per la probabilità condizionale*: dati tre eventi A, B, C tali che $P(B \cap C) > 0$ e $P(B \cap C^c) > 0$, si ha

$$P(A|B) = P(A|B \cap C)P(C|B) + P(A|B \cap C^c)P(C^c|B).$$

Esercizio 1.27. Una compagnia di assicurazioni offre una polizza che prevede il pagamento di una cifra forfettaria C a fronte di un danno subito dal cliente. La compagnia classifica gli assicurati in tre categorie: “basso rischio”, “medio rischio” e “alto rischio”. Dei suoi assicurati, il 75% sono a “basso rischio”, il 20% a “medio rischio” e il restante 5% ad “alto rischio”.

È noto che gli assicurati a “basso rischio” hanno una probabilità del 2% di subire un danno che prevede il pagamento dell'assicurazione, mentre tale probabilità è del 10% per gli assicurati a “medio rischio” e del 20% per quelli ad “alto rischio”.

- (i) Qual è la probabilità che un individuo scelto a caso tra gli assicurati reclami il pagamento dell'assicurazione?
- (ii) Se un individuo reclama il pagamento dell'assicurazione, qual è la probabilità che sia nella categoria ad "alto rischio"?

Esercizio 1.28. Durante la notte, un taxi ha causato un incidente. In città operano due compagnie di taxi, una con i taxi gialli, l'altra con i taxi bianchi. Un testimone ha dichiarato che il taxi coinvolto nell'incidente era giallo. Sappiamo che i taxi bianchi sono l'85% dei taxi in città. Inoltre, la probabilità che un testimone, di notte, identifichi correttamente il colore del taxi è pari a 0.8.

- (i) Sulla base di queste informazioni, qual è la probabilità che il taxi coinvolto nell'incidente fosse in realtà bianco?
- (ii) Supponiamo che un secondo testimone abbia dichiarato che il taxi era giallo, e che la correttezza dell'identificazione del colore da parte di questo testimone sia indipendente da quella del primo. Sulla base di questa ulteriore informazione, qual è ora la probabilità che il taxi coinvolto nell'incidente fosse in realtà bianco?

Esercizio 1.29. Il Ministero della Pubblica Istruzione vuole stimare la frazione $\alpha \in (0, 1)$ di studenti di terza media che hanno preparazione scarsa in matematica. A tal fine, sottopone a un grande numero di studenti un quesito con 10 possibili risposte, di cui una sola è corretta. Assumiamo che gli studenti con una buona preparazione in matematica rispondano correttamente al quesito, mentre quelli con preparazione scarsa diano una risposta scelta a caso (e non esistano altre possibilità). Sottoponendo ad una analisi più approfondita gli studenti che hanno risposto correttamente al quesito, si scopre che tra questi solo l'80% ha una buona preparazione in matematica. Sulla base di queste informazioni, si determini α .

Esercizio 1.30. Tre urne, etichettate con le lettere α, β, γ , contengono 10 palline ciascuna. Due urne contengono 5 palline rosse e 5 blu, mentre la terza contiene 3 palline rosse e 7 blu. Non sappiamo però quale sia l'urna con 3 palline rosse: in assenza di ulteriori informazioni, riteniamo che sia α, β o γ con la stessa probabilità.

Estraiamo ora 2 palline da ognuna delle tre urne. Se dall'urna α abbiamo estratto una pallina rossa e una blu, dall'urna β due palline rosse e dall'urna γ due palline blu, qual è la probabilità che l'urna γ sia quella contenente tre palline rosse?

Esercizio 1.31. Un'urna contiene M palline, di cui M_1 bianche.

- (i) Si effettuano n estrazioni successive, *con* reimmissione. Si considerino gli eventi

$B_j :=$ "la j -esima pallina estratta è bianca",

$A_m :=$ "delle n palline estratte esattamente m sono bianche",

dove $j, m \leq n$. Si calcoli $P(B_j | A_m)$.

- (ii) Si calcoli la probabilità condizionale del punto precedente nel caso di estrazioni *senza* reimmissione, supponendo che m sia tale che $P(A_m) > 0$.

Esercizio 1.32. Ho due dadi regolari: il dado α ha sei facce, su cui sono scritti i numeri da 1 a 6, mentre il dado β ha dodici facce, su cui sono scritti i numeri da 1 a 12. Scelgo uno dei due dadi a caso, con la stessa probabilità, e lo lancio per n volte, dove $n \in \mathbb{N}$ è un numero fissato.

- (i) Qual è la probabilità che tutti i lanci diano come risultato il numero 3?
- (ii) Qual è la probabilità che tutti i lanci diano come risultato lo stesso numero?
- (iii) Se tutti i lanci danno come risultato il numero 3, qual è la probabilità che il dado scelto sia stato α ? Si mostri che tale probabilità (condizionale) è sempre strettamente maggiore di $\frac{1}{2}$ e se ne studi il comportamento per $n \rightarrow \infty$.

Esercizio 1.33. Ho un'urna inizialmente vuota e un insieme di palline numerate coi numeri naturali. Il primo giorno inserisco nell'urna le palline numero 1 e 2, dopodiché ne estraggo una a caso (nell'urna rimane dunque una sola pallina). Il secondo giorno inserisco nell'urna le palline numero 3 e 4, dopodiché estraggo a caso una delle tre palline contenute nell'urna. Itero dunque la procedura: l' i -esimo giorno inserisco nell'urna le palline numero $2i - 1$ e $2i$, dopodiché estraggo a caso una delle $i + 1$ palline contenute nell'urna.

Si introduca per $i \in \mathbb{N}$ l'evento

$A_i :=$ “la pallina numero 1 è presente nell'urna alla fine dell' i -esimo giorno”.

- (i) Si spieghi perché vale l'inclusione $A_{i+1} \subseteq A_i$ per ogni $i \in \mathbb{N}$ e si deduca la formula

$$P(A_n) = P(A_1) \cdot P(A_2|A_1) \cdots P(A_n|A_{n-1}), \quad n \in \mathbb{N}.$$

- (ii) Si mostri che $P(A_n) = \frac{1}{n+1}$ per ogni $n \in \mathbb{N}$.

Esercizio 1.34. Si consideri il seguente modello per la distribuzione dei sessi dei figli in una famiglia:

- il primo figlio ha probabilità $\frac{1}{2}$ di essere maschio (o femmina);
- la probabilità che l' $(n + 1)$ -esimo figlio sia maschio, condizionalmente ai sessi dei figli precedenti, è $\frac{3}{5}$ se l' n -esimo figlio è maschio, $\frac{2}{5}$ se l' n -esimo figlio è femmina.

Si determini quindi:

- (i) la probabilità che il primo figlio sia maschio, condizionale al fatto che il secondo è maschio;
- (ii) la probabilità che il primo figlio sia maschio, condizionale al fatto che il terzo è maschio.

Esercizio 1.35. Il signor Bianchi da Roma e il signor Rossi da Milano decidono di incontrarsi a Roma. All'ultimo momento, Rossi, che è un tipo molto indeciso, rimette al caso la decisione di partire, lanciando una moneta. Successivamente, in caso di esito positivo, per scegliere quale dei 6 treni a sua disposizione prendere, tira un dado regolare a sei facce. Se Bianchi va in stazione e osserva che Rossi non è su nessuno dei primi 5 treni, qual è la probabilità che Rossi arrivi con l'ultimo treno?

Esercizio 1.36. Antonio e Berta si incontrano per una gara di scacchi. Convengono di fare due partite, assegnando un punto in caso di vittoria, zero punti in caso di sconfitta e mezzo punto in caso di pareggio o patta. Nel caso in cui dopo le due partite i due giocatori abbiano lo stesso punteggio, lanceranno una moneta equilibrata per determinare il vincitore della gara.

Antonio sa giocare con due diversi approcci, uno offensivo e uno difensivo, mentre Berta gioca sempre in maniera offensiva. Se Antonio gioca in maniera offensiva, vince con probabilità $p \in (0, 1]$ e perde con probabilità $1 - p$. Se invece gioca in maniera difensiva, pareggia con probabilità $q \in (0, 1]$ e perde con probabilità $1 - q$.

Antonio decide di adottare la seguente strategia. Gioca la prima partita in maniera offensiva. Se perde, gioca anche la seconda in maniera offensiva, mentre se vince gioca la seconda partita in maniera difensiva.

- (i) Si calcoli, in termini di p e q , la probabilità p_* che Antonio vinca la gara.
- (ii) Si assuma che $q = 0.9$. Per quali valori di p si ha $p_* > \frac{1}{2}$? È possibile che Berta sia la giocatrice più forte – nel senso che ha maggiore probabilità di vincere una partita rispetto ad Antonio – e ciononostante Antonio abbia maggiore probabilità di vincere la gara?

Esercizio 1.37. Una guida alpina organizza abitualmente salite alla cima del Monte Archimede. Talvolta, in presenza di cattive condizioni atmosferiche, la guida decide di tornare prima di aver raggiunto la cima, anche a seconda delle capacità delle persone accompagnate. In caso di pioggia senza raffiche di vento, la guida rinuncia a raggiungere la cima il 20% delle volte; in caso di raffiche di vento ma senza pioggia, la guida rinuncia il 30% delle volte; con pioggia e raffiche di vento, la guida rinuncia l'80% delle volte; infine, se non piove e non ci sono raffiche di vento, la cima viene sicuramente raggiunta. Assumiamo che gli eventi “si trova pioggia lungo il percorso” e “ci sono raffiche di vento lungo il percorso” siano indipendenti e abbiano probabilità rispettivamente 0.3 e 0.2.

Oggi un gruppo è partito con la guida, ma è tornato senza aver raggiunto la cima. Qual è la probabilità che abbia trovato pioggia?

Esercizio 1.38. (i) Si dimostri che, se $\alpha_1, \alpha_2, \dots, \alpha_n \geq 0$ e $\sum_{i=1}^n \alpha_i = 1$, allora per ogni scelta di $x_1, x_2, \dots, x_n \in \mathbb{R}$ si ha

$$\min_{i=1,2,\dots,n} x_i \leq \sum_{i=1}^n \alpha_i x_i \leq \max_{i=1,2,\dots,n} x_i.$$

- (ii) Siano ora A_1, A_2, \dots, A_n eventi *disgiunti* di uno spazio di probabilità (Ω, \mathcal{P}) , tali che $P(A_i) > 0$ per ogni $i = 1, \dots, n$. Si mostri che per ogni evento B

$$\min_{i=1,2,\dots,n} P(B|A_i) \leq P(B|A_1 \cup A_2 \cup \dots \cup A_n) \leq \max_{i=1,2,\dots,n} P(B|A_i).$$

Esercizio 1.39. Siano A e B due eventi con probabilità non nulla. Diciamo che A è *positivamente correlato a B* se

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

Si mostri che le seguenti tre affermazioni sono equivalenti.

- (i) A è *positivamente correlato a B* .
- (ii) B è *positivamente correlato a A* .
- (iii) A^c è *positivamente correlato a B^c* .

Esercizio 1.40. Un'urna contiene n palline, che possono essere di due colori, rosso e verde. Non conosciamo la composizione dell'urna e riteniamo che tutti i possibili valori $k = 0, 1, 2, \dots, n$ del numero di palline rosse siano equiprobabili.

- (i) Si estrae una pallina dall'urna, che si rivela essere rossa. Sapendo ciò, per quale valore di k la probabilità che nell'urna vi fossero k palline rosse è massimizzata?
- (ii) Si risponda alla medesima domanda, ma assumendo che dall'urna siano state estratte due palline, una rossa e una verde.

Esercizio 1.41. Si consideri il seguente modello di distribuzione dei figli nei nuclei familiari. La probabilità che un nucleo familiare scelto a caso abbia n figli, con $n \geq 0$, vale $e^{-\lambda} \frac{\lambda^n}{n!}$ (dove $\lambda > 0$ è un parametro fissato), e ciascun figlio è maschio con probabilità $1/2$, indipendentemente da tutti gli altri. Consideriamo l'evento

$A_k :=$ “il nucleo familiare scelto (a caso) ha esattamente k figli maschi” ,

per $k \geq 0$. Si mostri che $P(A_k) = e^{-\lambda/2} \frac{(\lambda/2)^k}{k!}$.

[Sugg. Si consideri l'evento $B_n =$ “il nucleo familiare scelto ha n figli”. Si determini innanzitutto $P(A_k|B_n)$ e poi si calcoli $P(A_k)$. Si ricordi la serie esponenziale (0.5).]

Esercizio 1.42. Sia $S = \{1, 2, \dots, n\}$, $\Omega := \mathcal{P}(S) \times \mathcal{P}(S)$, e P la probabilità uniforme su Ω . Dunque gli elementi di Ω sono coppie ordinate (A, B) , con $A, B \subseteq S$. Consideriamo l'evento

$$E := \{(A, B) \in \Omega : A \subseteq B\}.$$

Inoltre, per $B \subseteq S$, definiamo $F_B := \{(A', B') \in \Omega : B' = B\} = \{(A, B) : A \subseteq S\}$.

- (i) Si determini $P(E|F_B)$.
- (ii) Usando la formula di disintegrazione

$$P(E) = \sum_{B \subseteq S} P(E|F_B) P(F_B),$$

si mostri che $P(E) = (3/4)^n$.

[Sugg. Si ricordi il binomio di Newton (1.29) e il fatto che $|\mathcal{P}(S)| = 2^{|S|}$.]

Esercizio 1.43. È stato indetto un referendum in una popolazione di $n \geq 1$ individui (tutti aventi diritto al voto). Ciascun individuo andrà a votare con probabilità $\frac{1}{2}$,

indipendentemente dagli altri. Inoltre, se un individuo andrà a votare, voterà SÌ con probabilità $\frac{1}{2}$, indipendentemente dagli altri.

- (i) Qual è la probabilità p che un individuo scelto a caso vada a votare e voti SÌ?
- (ii) Qual è la probabilità che il numero di voti SÌ sia k , per $k \in \{0, \dots, n\}$?
- (iii) Assumendo che i voti SÌ siano k , si determini la probabilità (condizionale) che i votanti totali siano m , dove $m \in \{k, \dots, n\}$. Si mostri che tale probabilità vale

$$\binom{n-k}{m-k} \left(\frac{1}{3}\right)^{m-k} \left(\frac{2}{3}\right)^{n-m}.$$

1.5 Note bibliografiche

La teoria della probabilità su spazi finiti o numerabili è trattata da tutti i testi introduttivi di probabilità. Per approfondimenti, menzioniamo la monografia classica di P. Billingsley [6], che tratta anche gli sviluppi più avanzati del calcolo delle probabilità, a cui accenneremo nel Capitolo 5. Seppure con un approccio meno moderno, segnaliamo anche il bellissimo testo di W. Feller [23].

Il paradosso descritto nell'Esempio 1.71, detto di "Monty Hall", prende il nome dal presentatore del gioco a premi americano *Let's Make a Deal*, in cui un concorrente doveva scegliere una di tre porte chiuse, dietro le quali si celavano un'automobile e due capre. Questo paradosso, proposto originariamente da S. Selvin nel 1975 [50], raggiunse una certa notorietà in seguito alla soluzione pubblicata da M. vos Savant nel 1990 sulla rivista *Parade* [49]. Sebbene la soluzione fosse corretta, molti lettori convinti del contrario (inclusi alcuni matematici...) inviarono alla rivista lettere di protesta, con toni anche piuttosto feroci.

I problemi giudiziari descritti negli Esempi 1.73 e 1.74 sono liberamente tratti dal libro *Innumeracy* di J.A. Paulos [44].

Spazi di probabilità discreti: esempi e applicazioni

Sommario. In questo capitolo vediamo all'opera gli spazi di probabilità discreti, introdotti nel Capitolo 1, applicati a diversi esempi e modelli di grande rilevanza. Con l'eccezione di alcune osservazioni non essenziali, la nozione di indipendenza non viene usata prima del Paragrafo 2.5.

2.1 Permutazioni aleatorie

Come nel Paragrafo 1.2.4, denotiamo con S_n l'insieme delle funzioni biunivoche dall'insieme $\{1, 2, \dots, n\}$ in sé, dette permutazioni. Osserviamo che S_n è un gruppo rispetto alla composizione di applicazioni, che è non commutativo se $n \geq 3$, e ricordiamo che $|S_n| = n!$, dove $n!$ è stato definito in (1.19).

Indichiamo con P_n la *probabilità uniforme* sull'insieme S_n , pensato come spazio campionario. Lo spazio di probabilità (S_n, P_n) è un buon modello per l'esperimento aleatorio che consiste nel *mescolare* accuratamente n oggetti e quindi osservare l'ordinamento ottenuto. In questo paragrafo esaminiamo alcune proprietà interessanti dello spazio (S_n, P_n) , prendendo spunto da alcuni problemi. Per definizione di probabilità uniforme (si ricordi l'Esempio 1.13), si ha

$$P_n(A) = \frac{|A|}{|S_n|} = \frac{|A|}{n!}, \quad \text{per ogni } A \subseteq S_n,$$

quindi il calcolo di probabilità si riconduce a un calcolo di cardinalità. Per semplicità, quando ciò non causi confusione, ometteremo la dipendenza da n nella probabilità P_n , indicandola semplicemente con P .

Cicli

Cominciamo con un problema in cui compare in modo naturale il concetto di *ciclo* di una permutazione contenente un dato elemento.

Problema 2.1. Un gruppo di n amici affitta una casa per una vacanza. Dopo alcuni giorni tutti convergono che sia il caso di fare delle pulizie, ma si stenta a trovare dei volontari. Laura, che è volenterosa e bizzarra, avanza la seguente proposta. Ognuno scrive il proprio nome su una carta. Quindi le n carte vengono accuratamente mescolate e distribuite. Laura allora leggerà ad alta voce il nome sulla sua carta. Quindi la persona il cui nome è stato letto leggerà a sua volta il nome sulla sua carta; si prosegue così finché non viene letto il nome di Laura. A questo punto, le persone il cui nome è stato chiamato formeranno la squadra per le pulizie.

- (i) Qual è la probabilità che Laura si trovi a dover fare le pulizie da sola?
- (ii) Qual è la probabilità che tutti debbano fare le pulizie?
- (iii) Più in generale, qual è la probabilità che la squadra delle pulizie sia composta da m persone?

Soluzione. Etichettiamo gli n amici con i numeri $1, 2, \dots, n$, assegnando il numero 1 a Laura. L'esito del mescolamento delle n carte può allora descritto in modo naturale con una permutazione $\sigma \in S_n$: la carta in mano alla persona i ha il nome della persona $\sigma(i)$. Il fatto che le carte siano mescolate accuratamente corrisponde a considerare la probabilità uniforme P su S_n .

La squadra per le pulizie si ottiene applicando *ripetutamente* la permutazione σ all'elemento 1:

$$\sigma(1), \quad \sigma \circ \sigma(1) =: \sigma^2(1), \quad \dots \quad \sigma^{k-1}(1), \quad \sigma^k(1) = 1,$$

dove $k \geq 1$ è il più piccolo numero intero tale che $\sigma^k(1) = 1$. In questo modo, $1, \sigma(1), \sigma^2(1), \dots, \sigma^{k-1}(1)$ sono elementi *distinti* di $\{1, \dots, n\}$. La sequenza

$$(1, \sigma(1), \sigma^2(1), \dots, \sigma^{k-1}(1))$$

viene detta *ciclo contenente 1*, e l'intero k viene detto *lunghezza del ciclo*. Il quesito (iii), che contiene gli altri due come casi particolari, può essere pertanto riformulato come segue: qual è la probabilità che il ciclo contenente 1 abbia lunghezza m ? Introducendo per ogni $m \in \mathbb{N}$ l'evento $C_m \subseteq S_n$ definito da

$$C_m := \{\sigma \in S_n : \text{il ciclo contenente 1 ha lunghezza } m\},$$

dobbiamo dunque calcolare $P(C_m)$ per ogni $1 \leq m \leq n$.

Cominciamo con il quesito (i), che corrisponde a $m = 1$. Si noti che l'evento $C_1 = \{\sigma \in S_n : \sigma(1) = 1\}$ corrisponde l'insieme delle permutazioni di $\{1, \dots, n\}$ in cui 1 viene mandato in sé stesso. C'è pertanto una naturale corrispondenza biunivoca tra C_1 e l'insieme delle permutazioni di $\{2, 3, \dots, n\}$, da cui si deduce che

$$|C_1| = (n-1)! \implies P(C_1) = \frac{|C_1|}{n!} = \frac{1}{n}.$$

In altre parole, la probabilità che Laura si trovi da sola a fare le pulizie è pari a $\frac{1}{n}$.

Consideriamo ora la domanda (ii), cioè calcoliamo $P(C_n)$. Le permutazioni $\sigma \in C_n$ sono tali che il ciclo contenente 1 ha lunghezza n , ossia si può rappresentare nella forma

$$(1, \sigma(1), \sigma^2(1), \dots, \sigma^{n-1}(1)).$$

Osserviamo che la scrittura precedente è una n -upla in cui compaiono tutti e soli gli elementi di $\{1, \dots, n\}$ con 1 al primo posto. Tali n -uple possono essere determinate mediante le seguenti n scelte successive:

- si sceglie $\sigma(1)$ in $\{1, 2, \dots, n\} \setminus \{1\}$, per cui ci sono $n - 1$ esiti possibili;
- si sceglie $\sigma^2(1)$ in $\{1, 2, \dots, n\} \setminus \{1, \sigma(1)\}$, per cui ci sono $n - 2$ esiti possibili;
- e così via fino a $\sigma^{n-1}(1)$, per cui resta un solo esito possibile.

Di conseguenza, per il principio fondamentale del calcolo combinatorio,

$$|C_n| = (n-1)(n-2) \cdots 1 \cdot 1 = (n-1)! \implies P(C_n) = \frac{|C_n|}{n!} = \frac{1}{n}.$$

A questo punto abbiamo gli strumenti per calcolare $|C_m|$ per ogni valore di m , cioè per rispondere alla domanda (iii). Infatti, gli elementi $\sigma \in C_m$ possono essere determinati dalle seguenti tre scelte successive:

- si scelgono gli m elementi di $\{1, \dots, n\}$ che compongono il ciclo contenente 1, per cui ci sono $\binom{n-1}{m-1}$ esiti possibili (uno degli m elementi dev'essere 1);
- si sceglie uno dei cicli formati da questi m elementi: come abbiamo appena visto nella risposta alla domanda (ii), ci sono $(m-1)!$ tali cicli;
- si scelgono i valori di σ sui rimanenti $n-m$ elementi: dato che σ permuta in modo arbitrario tali elementi, per questa scelta ci sono $(n-m)!$ esiti possibili.

Per il principio fondamentale del calcolo combinatorio, si ottiene

$$|C_m| = \binom{n-1}{m-1} (m-1)! (n-m)! = (n-1)! \implies P(C_m) = \frac{1}{n}.$$

Concludendo, la probabilità che la squadra per le pulizie sia composta da m elementi è $\frac{1}{n}$, in particolare non dipende da m .

Nel problema precedente abbiamo introdotto, per una permutazione $\sigma \in S_n$ fissata, la nozione di ciclo contenente l'elemento 1. In modo analogo, si può costruire il ciclo contenente ogni altro elemento $i \in \{1, \dots, n\}$. È chiaro che, se i appartiene al ciclo contenente 1, il ciclo contenente i coincide con il ciclo contenente 1 (a meno di una traslazione dei suoi elementi); viceversa, se i non appartiene al ciclo contenente 1, il ciclo contenente i è disgiunto dal ciclo contenente 1. Di conseguenza, ogni permutazione $\sigma \in S_n$ individua una *partizione in cicli* di $\{1, 2, \dots, n\}$.

Problema 2.2. Lo stesso gruppo di n amici del Problema 2.1 decide di giocare a Trivial Pursuit. Decidono quindi di usare il metodo proposto da Laura per suddividersi in squadre, corrispondenti alla partizione in cicli determinata della permutazione.

(Il numero di squadre non è dunque fissato a priori, e le squadre non sono necessariamente della stessa numerosità.) Qual è la probabilità che si formi una squadra, necessariamente unica, con strettamente più di $n/2$ persone?

Soluzione. Muniamo ancora S_n della probabilità uniforme P . Introduciamo gli eventi D e D_m , per $m \in \{1, \dots, n\}$, definiti da

$$D := \{\sigma \in S_n : \sigma \text{ ha un ciclo di lunghezza strettamente maggiore di } n/2\},$$

$$D_m := \{\sigma \in S_n : \sigma \text{ ha un ciclo di lunghezza } m\}.$$

Il quesito del problema richiede di calcolare $P(D)$. Notiamo che D è l'unione degli eventi D_m per $m > n/2$, e che tali eventi sono disgiunti, pertanto

$$P(D) = \sum_{\frac{n}{2} < m \leq n} P(D_m), \quad (2.1)$$

e ci resta da determinare $P(D_m)$ per $m > n/2$.

L'osservazione fondamentale è che, se $m > n/2$, un ciclo di lunghezza m è necessariamente unico (mentre ciò non è vero se $m \leq n/2$). In effetti, se $m > n/2$, gli elementi $\sigma \in D_m$ possono essere determinati attraverso le seguenti scelte successive:

- si scelgono gli m elementi che compaiono nel ciclo “grande”, per cui ci sono $\binom{n}{m}$ esiti possibili;
- si sceglie uno dei possibili cicli formati da questi m elementi, per cui ci sono (come abbiamo visto nel Problema 2.1) $(m-1)!$ esiti possibili;
- si fissano in modo arbitrario i valori di σ sui rimanenti $n-m$ elementi, per cui ci sono $(n-m)!$ esiti possibili.

Pertanto

$$|D_m| = \binom{n}{m} (m-1)! (n-m)! = \frac{n!}{m} \implies P(D_m) = \frac{1}{m}.$$

(Per $m \leq \frac{n}{2}$, la possibile non unicità dei cicli di lunghezza m conduce a “contare più di una volta” la stessa permutazione, e quindi il precedente conteggio è scorretto.)

Ricordando (2.1), la risposta al quesito del problema è data da

$$P(D) = \sum_{\frac{n}{2} < m \leq n} \frac{1}{m} = \sum_{m=\lfloor \frac{n}{2} \rfloor + 1}^n \frac{1}{m} =: p_n,$$

dove abbiamo indicato con p_n la probabilità richiesta, per evidenziarne la dipendenza da n , e dove $\lfloor \frac{n}{2} \rfloor$ indica la parte intera di $\frac{n}{2}$. Abbiamo dunque ottenuto una formula esplicita, ma non molto trasparente. Cerchiamo di studiarne il comportamento per valori grandi di n . Sostituendo la somma con un integrale si ottiene

$$p_n \approx \int_{\lfloor n/2 \rfloor}^n \frac{1}{x} dx \approx \log n - \log(n/2) = \log 2.$$

Non è difficile rendere rigorosa questa stima: dato che la funzione $1/x$ è decrescente, per ogni $m \in \mathbb{N}$ possiamo scrivere

$$\int_m^{m+1} \frac{1}{x} dx \leq \frac{1}{m} \leq \int_{m-1}^m \frac{1}{x} dx,$$

e sommando su m da $\lfloor \frac{n}{2} \rfloor + 1$ a n si ottiene

$$\log \frac{n+1}{\lfloor \frac{n}{2} \rfloor + 1} = \int_{\lfloor \frac{n}{2} \rfloor + 1}^{n+1} \frac{1}{x} dx \leq p_n \leq \int_{\lfloor \frac{n}{2} \rfloor}^n \frac{1}{x} dx = \log \frac{n}{\lfloor \frac{n}{2} \rfloor}.$$

Dato che $\lim_{n \rightarrow \infty} (n+1)/(\lfloor \frac{n}{2} \rfloor + 1) = \lim_{n \rightarrow \infty} n/\lfloor \frac{n}{2} \rfloor = 2$, si ottiene infine

$$\lim_{n \rightarrow +\infty} p_n = \log 2.$$

In altre parole, per grandi valori di n , la probabilità che si formi una squadra con più di $n/2$ persone è approssimativamente $\log 2 \simeq 0.69$ e dunque (approssimativamente) *non dipende da n* , un risultato non evidente a priori. Per $n > 50$, $|p_n - \log 2| \leq 0.01$.

Il risultato appena ottenuto permette di trovare una soluzione al seguente difficile problema.

Problema 2.3. Il docente di un corso di probabilità frequentato da 100 studenti propone ai suoi allievi quanto segue. Si preparano 100 buste, numerate da 1 a 100, e 100 carte, su ciascuna delle quali è scritto il nome di uno studente del corso (senza ripetizioni, si escludano omonimie). Quindi le carte vengono inserite, casualmente, una in ogni busta. Le buste, chiuse ma non sigillate, vengono quindi disposte sulla cattedra di un aula. Gli studenti entrano nell'aula uno per volta. Ogni studente apre a suo piacimento 50 buste, una dopo l'altra, e comunica al docente se, tra le buste aperte, c'è quella con il proprio nome. Quindi le richiude ed esce dall'aula, senza poter in alcun modo comunicare con i colleghi che ancora devono entrare in aula.

Il docente alzerà il voto dell'esame di tre punti a tutti gli studenti solo nel caso in cui *ciascuno studente* trovi la busta contenente la carta con il proprio nome. Gli studenti non possono comunicare dopo l'inizio delle aperture, ma possono concordare una strategia *a priori*. Si determini una strategia che conduca al successo (cioè all'aumento di tre punti per tutti) con probabilità non trascurabile.

Prima di descrivere la soluzione, notiamo che è assolutamente non ovvio ottenere una probabilità di successo non trascurabile. Le strategie "banali" falliscono miseramente. Supponiamo che gli studenti non si accordino per nulla, ad esempio che ognuno di essi scelga a caso, indipendentemente dagli altri, le 50 buste da aprire. In questo caso, ognuno avrebbe probabilità $\frac{1}{2}$ di trovare il proprio nome e, vista l'indipendenza delle scelte, la probabilità che *tutti* trovino il proprio nome sarebbe $\frac{1}{2^{100}} \simeq 8 \cdot 10^{-31}$: irrisoria! Si può fare naturalmente di peggio: se fossero così sciocchi da accordarsi di aprire tutti le *stesse* 50 buste, la probabilità di successo sarebbe nulla. Quello che non è ovvio è se sia possibile fare meglio.

Soluzione del Problema 2.3. Poniamo $n := 100$ ed etichettiamo i cento nomi degli studenti con i numeri $1, 2, \dots, n$. Denotiamo inoltre con $\sigma(k)$, il numero (nome) all'interno della busta k . Tale σ è evidentemente un elemento di S_n e la probabilità uniforme su S_n corrisponde al fatto che i nomi nelle buste vengono inseriti a caso.

Lo scopo dello studente k è di aprire la busta che contiene al suo interno il numero k , ossia la busta j con $\sigma(j) = k$. Supponiamo che gli studenti si accordino per seguire la seguente strategia. Ogni studente k apre per prima la busta k e ne legge il contenuto $\sigma(k)$; quindi apre la busta $\sigma(k)$ leggendone il contenuto $\sigma^2(k)$, e così via. Se, nella permutazione σ , il ciclo contenente k ha lunghezza $m \leq \frac{n}{2}$, la m -esima busta aperta dallo studente k è la busta $\sigma^{m-1}(k)$, il cui contenuto è proprio $\sigma^m(k) = k$: questo significa che lo studente k trova la carta col proprio nome! Pertanto, *se tutti i cicli di σ hanno lunghezza minore o uguale a $n/2$, ogni studente troverà sicuramente la busta contenente il proprio nome.* Viceversa, se un ciclo di σ ha lunghezza strettamente maggiore di $n/2$, gli studenti il cui numero appartiene a quel ciclo *non* troveranno il loro nome tra le prime 50 buste aperte. Di conseguenza, indicando con p_n è la probabilità calcolata nel Problema 2.2,

$$\text{probabilità di successo della strategia} = 1 - p_n \simeq 1 - \log 2 \simeq 0.31,$$

una probabilità decisamente non trascurabile!

Sottolineiamo che il limite inferiore ottenuto alla probabilità di successo è approssimativamente indipendente da n , se n è abbastanza grande. (Per scrupolo, per $n = 100$, si calcola $p_n \simeq 0.688$ e dunque $1 - p_n \geq 0.31$.)

Per capire meglio la strategia, definiamo gli eventi

$$B_k := \{\text{lo studente numero } k \text{ trova la carta col proprio nome}\},$$

$$B := \{\text{tutti gli studenti trovano la carta col proprio nome}\} = \bigcap_{k=1}^n B_k.$$

Non è difficile convincersi del fatto che, qualunque sia la strategia seguita, ogni studente fissato ha probabilità $1/2$ di trovare la busta con il suo nome! In altri termini, $P(B_k) = 0.5$ per ogni $k = 1, \dots, n$, indipendentemente dalla strategia seguita dagli studenti. Dato che $B \subseteq B_k$, per ogni k , segue che $P(B) \leq 0.5$. Con la strategia proposta abbiamo mostrato che $P(B) = P(\bigcap_{k=1}^n B_k) \geq 0.31$. Questo significa che gli eventi $(B_k)_{1 \leq k \leq n}$ sono tutt'altro che indipendenti, come, d'altra parte, segue dal fatto che

$$\bigcap_{k=1}^{\lfloor n/2 \rfloor + 1} B_k = \bigcap_{k=1}^n B_k.$$

Infatti, se i primi $\lfloor n/2 \rfloor + 1$ studenti trovano tutti il proprio nome, significa che non c'è nessun ciclo di lunghezza strettamente maggiore di $n/2$ e dunque *tutti* gli studenti trovano il proprio nome!

Punti fissi

Consideriamo ora un problema in cui giocano un ruolo fondamentale i punti fissi di una permutazione.

Problema 2.4. Una comitiva di n turisti si sta imbarcando per un viaggio aereo. La loro guida ha tutte le carte d'imbarco (nominative), che deve distribuire ai turisti prima dell'imbarco. Per la fretta e la confusione le distribuisce a caso. Qual è la probabilità che qualcuno dei turisti riceva effettivamente la propria carta d'imbarco? Qual è la probabilità che esattamente m turisti ricevano la propria carta d'imbarco?

Soluzione. Etichettiamo con $\{1, 2, \dots, n\}$ gli n turisti e indichiamo con $\sigma(i)$ il numero (nome) sulla carta d'imbarco ricevuta dal turista i . Si noti che $\sigma \in S_n$. Come al solito, consideriamo su S_n la probabilità uniforme, che indichiamo con P_n .

Osserviamo che l' i -esimo turista riceve la propria carta d'imbarco se $\sigma(i) = i$, cioè se i è un *punto fisso* della permutazione σ . Dunque, i quesiti del problema si possono riformulare come segue: qual è la probabilità che una permutazione abbia almeno un punto fisso? E qual è la probabilità che abbia esattamente m punti fissi? Per $m = 0, 1, \dots, n$ e $i = 1, 2, \dots, n$, introduciamo gli eventi

$$A_m := \{\sigma \in S_n : \sigma \text{ ha esattamente } m \text{ punti fissi}\}, \quad C_i := \{\sigma \in S_n : \sigma(i) = i\},$$

così che ci resta da calcolare $P_n(A_0^c)$ e $P_n(A_m)$.

Cominciamo da $P_n(A_0^c)$. Notiamo che

$$A_0^c = C_1 \cup C_2 \cup \dots \cup C_n,$$

pertanto, per la formula di inclusione-esclusione (Proposizione 1.18),

$$P_n(A_0^c) = \sum_{k=1}^n \sum_{\substack{J \subseteq \{1, 2, \dots, n\} \\ \text{tali che } |J|=k}} (-1)^{k+1} P_n\left(\bigcap_{i \in J} C_i\right). \quad (2.2)$$

Fissiamo dunque $k \in \{1, 2, \dots, n\}$, e sia $J \subseteq \{1, 2, \dots, n\}$ tale che $|J| = k$. Si noti che

$$\bigcap_{i \in J} C_i = \{\sigma \in S_n : \sigma(i) = i \text{ per ogni } i \in J\},$$

ossia l'insieme delle permutazioni che lasciano fissi gli elementi di J . Queste sono in naturale corrispondenza biunivoca con le permutazioni di $\{1, 2, \dots, n\} \setminus J$, pertanto

$$\left| \bigcap_{i \in J} C_i \right| = (n-k)! \implies P_n\left(\bigcap_{i \in J} C_i\right) = \frac{(n-k)!}{n!}.$$

Poiché i sottoinsiemi J di $\{1, 2, \dots, n\}$ con k elementi, ossia le combinazioni di k elementi estratti da $\{1, 2, \dots, n\}$, sono $\binom{n}{k}$, si ha

$$\sum_{\substack{J \subseteq \{1, 2, \dots, n\} \\ \text{tali che } |J|=k}} P_n\left(\bigcap_{i \in J} C_i\right) = \binom{n}{k} \frac{(n-k)!}{n!} = \frac{1}{k!}.$$

Inserendo quest'ultima uguaglianza in (2.2) otteniamo

$$P_n(A_0^c) = \sum_{k=1}^n \frac{(-1)^{k+1}}{k!} = 1 - \sum_{k=0}^n \frac{(-1)^k}{k!}.$$

Segue dalla serie esponenziale (0.5) che $\sum_{k=0}^{\infty} \frac{(-1)^k}{k!} = e^{-1}$. Inoltre, è noto che il resto tra la somma parziale e la serie è maggiorato dal termine successivo della successione, ossia

$$\left| \sum_{k=0}^n \frac{(-1)^k}{k!} - e^{-1} \right| \leq \frac{1}{(n+1)!}, \quad \forall n \in \mathbb{N}.$$

Quindi $\lim_{n \rightarrow \infty} P_n(A_0^c) = 1 - e^{-1}$, e l'approssimazione

$$P_n(A_0^c) \simeq 1 - e^{-1} \simeq 0.632$$

è eccellente per valori non troppo piccoli di n (già per $n = 6$ i due numeri hanno le prime tre cifre decimali uguali). Dunque, la probabilità che almeno un passeggero riceva la sua carta di imbarco è “quasi” indipendente dal numero di passeggeri!

Resta da determinare $P_n(A_m)$ per $m \geq 1$. Notiamo che

$$A_m = \bigcup_{J \subseteq \{1, 2, \dots, n\}: |J|=m} B_J,$$

dove

$$B_J := \{ \sigma \in S_n : \sigma(j) = j \text{ per ogni } j \in J, \sigma(i) \neq i \text{ per ogni } i \notin J \}.$$

Dato che B_J e $B_{J'}$ sono disgiunti se $J \neq J'$, segue che

$$P_n(A_m) = \sum_{J \subseteq \{1, 2, \dots, n\}: |J|=m} P_n(B_J), \quad (2.3)$$

e ci resta da determinare $P_n(B_J)$. Introduciamo per $n \in \mathbb{N}$ la notazione

$$q_n := \sum_{k=0}^n \frac{(-1)^k}{k!},$$

che, come visto sopra, è la probabilità dell'insieme delle permutazioni di un insieme di n elementi che non hanno alcun punto fisso, cioè $P_n(A_0) = q_n$. Di conseguenza, il numero di tali permutazioni vale $n! q_n$. Dato che ogni elemento di B_J può essere identificato con una permutazione dell'insieme $\{1, 2, \dots, n\} \setminus J$ che non ha alcun punto fisso, segue che, se $|J| = m$,

$$|B_J| = (n-m)! q_{n-m} \implies P_n(B_J) = \frac{(n-m)!}{n!} q_{n-m}.$$

Ricordando (2.3), otteniamo

$$P_n(A_m) = \binom{n}{m} \frac{(n-m)!}{n!} q_{n-m} = \frac{q_{n-m}}{m!},$$

da cui segue che, per ogni $m \in \mathbb{N}$ fissato,

$$\lim_{n \rightarrow \infty} P_n(A_m) = \frac{e^{-1}}{m!} =: q(m). \quad (2.4)$$

Di conseguenza, se m non è troppo vicino a n , si ha l'eccellente approssimazione $P_n(A_m) \simeq q(m)$. Dunque, anche la probabilità che m passeggeri ricevano la propria carta di imbarco è “quasi” indipendente dal numero totale n di passeggeri.

Segue dalla serie esponenziale (0.5) che $q(\cdot)$ definita in (2.4) è una densità discreta su \mathbb{N}_0 , come abbiamo già osservato nell'Esempio 1.15. Tale densità discreta è detta di Poisson (di parametro 1) e ne studieremo le proprietà nel Paragrafo 3.5.5.

2.2 La passeggiata aleatoria semplice

In questo paragrafo studiamo un moto aleatorio sull'insieme dei numeri interi \mathbb{Z} che evolve a *tempo discreto*: più precisamente, l'insieme dei tempi è $\{0, 1, \dots, n\}$, dove l'istante finale $n \in \mathbb{N}$ è un parametro fissato. Indichiamo con s_k la posizione al tempo k e designiamo con il termine *cammino* il vettore delle posizioni (s_0, s_1, \dots, s_n) . Il moto avviene con le seguenti semplici regole:

- la posizione all'istante iniziale è il punto $0 \in \mathbb{Z}$, ossia $s_0 = 0$;
- se $x \in \mathbb{Z}$ è la posizione all'istante k , ossia se $s_k = x$, allora le posizioni possibili all'istante successivo $k+1$ sono $x+1$ e $x-1$, ossia $s_{k+1} \in \{x-1, x+1\}$;
- tutti i cammini *possibili* (s_0, s_1, \dots, s_n) sono equiprobabili.

L'insieme dei cammini possibili è dato dunque da

$$\tilde{\Omega}_n = \{(s_0, s_1, \dots, s_n) : s_0 = 0, |s_k - s_{k-1}| = 1 \text{ per ogni } k \in \{1, \dots, n\}\}.$$

La richiesta che tutti i cammini siano equiprobabili conduce a munire $\tilde{\Omega}_n$ della probabilità uniforme P . Lo spazio di probabilità $(\tilde{\Omega}_n, P)$ è detto *passeggiata aleatoria semplice e simmetrica su \mathbb{Z} di n passi* (l'aggettivo *semplice* sta ad indicare che gli incrementi $s_k - s_{k-1}$ possono assumere solo i valori ± 1). Si tratta del più semplice modello per un moto aleatorio, tuttavia di rilevanza teorica e applicativa fondamentale. Ci sono molte domande naturali su questo modello che, a dispetto della semplicità di formulazione, hanno risposte non banali e per certi versi sorprendenti.

Noi ci concentreremo sulla seguente domanda classica: *con quale probabilità la passeggiata ritorna al punto di partenza?* Prima di affrontare questo problema, anticipiamo alcune osservazioni importanti.

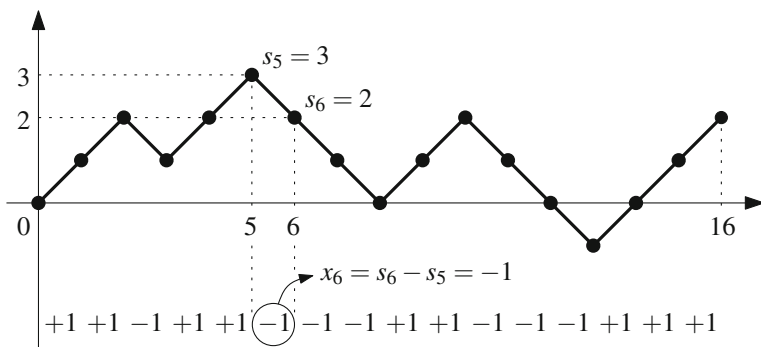


Figura 2.1 Un cammino (s_0, s_1, \dots, s_n) della passeggiata aleatoria semplice, con $n = 16$, sotto cui è riportata la sequenza $(x_1, x_2, \dots, x_n) \in \{-1, +1\}^n$ degli incrementi corrispondenti

Considerazioni preliminari

Introduciamo un nuovo spazio campionario Ω_n , definito da

$$\Omega_n := \{-1, +1\}^n = \{(x_1, \dots, x_n) : x_k \in \{-1, +1\} \text{ per ogni } k \in \{1, \dots, n\}\}.$$

Notiamo che s_k è definito per $k \in \{0, 1, \dots, n\}$ mentre x_k è definito per $k \in \{1, \dots, n\}$. Se le variabili s_k rappresentano le *posizioni* del cammino nei diversi istanti, le variabili x_k rappresentano gli *incrementi* del cammino. È chiaro che un cammino può essere equivalentemente descritto in termini delle posizioni o degli incrementi. Più formalmente, esiste una *corrispondenza biunivoca* $(s_0, s_1, \dots, s_n) \mapsto (x_1, x_2, \dots, x_n)$ tra gli insiemi $\tilde{\Omega}_n$ e Ω_n , definita da $x_k := s_k - s_{k-1}$, la cui inversa è data da $s_0 := 0$ e $s_k := x_1 + \dots + x_k$ per $k \geq 1$ (si veda la Figura 2.1). In particolare, $|\tilde{\Omega}_n| = |\Omega_n| = 2^n$, ossia ci sono 2^n cammini possibili che terminano all'istante n .

Se muniamo anche l'insieme Ω_n della probabilità uniforme P , è *indifferente lavorare con lo spazio di probabilità delle posizioni* $(\tilde{\Omega}_n, P)$ *o con quello degli incrementi* (Ω_n, P) : infatti, data la corrispondenza biunivoca appena citata, ogni evento di uno spazio è in corrispondenza biunivoca con un evento dell'altro spazio, ed eventi corrispondenti hanno la stessa probabilità. Come vedremo, per alcune questioni risulta più semplice lavorare con lo spazio degli incrementi (Ω_n, P) .

Osservazione 2.5. Il lettore attento ricorderà lo spazio di probabilità $(\{0, 1\}^n, P_{n,p})$ che descrive n prove ripetute e indipendenti con probabilità di successo p , introdotto nel Paragrafo 1.3.4. Nel caso speciale $p = \frac{1}{2}$ si ha che $P_{n,1/2}$ è la probabilità uniforme su $\{0, 1\}^n$, come segue dall'equazione (1.62). Pertanto, se “rinominiamo” $0 \rightarrow -1$, lo spazio di probabilità $(\{0, 1\}^n, P_{n,1/2})$ diventa lo spazio di probabilità (Ω_n, P) degli incrementi della passeggiata aleatoria semplice e simmetrica, introdotto sopra. Questo significa che gli incrementi della passeggiata aleatoria semplice e simmetrica costituiscono n prove ripetute e indipendenti con probabilità di successo $p = \frac{1}{2}$, dove con “successo” (risp. “insuccesso”) si intende che l'incremento valga $+1$ (risp. -1). Per inciso, l'aggettivo “simmetrica” si riferisce proprio al fatto che gli incrementi della passeggiata assumono i valori $+1$ e -1 con la stessa probabilità $p = \frac{1}{2}$. \square

Prima di analizzare in dettaglio la domanda sopra enunciata, c'è una questione che merita di essere approfondita. Consideriamo un evento che dipende solo dalle prime n posizioni della passeggiata aleatoria, o equivalentemente dai primi n incrementi, come ad esempio “la posizione s_n al tempo n è uguale a 0”. Per descrivere questo evento, è naturale considerare il sottoinsieme di Ω_n dato da

$$A_n := \{(x_1, \dots, x_n) \in \Omega_n : x_1 + x_2 + \dots + x_n = 0\}. \quad (2.5)$$

La scelta di Ω_n non è tuttavia obbligata: è altrettanto legittimo adottare come spazio campionario Ω_N , per un qualunque valore di $N \geq n$, e definire l'analogo sottoinsieme di Ω_N in termini delle prime n variabili x_1, \dots, x_n :

$$A_N := \{(x_1, \dots, x_N) \in \Omega_N : x_1 + x_2 + \dots + x_n = 0\}.$$

Questa “ambiguità” non crea problemi, perché *le probabilità degli eventi A_n e A_N sono le stesse*: più precisamente, indicando per chiarezza con P_n e P_N le probabilità (uniformi) su Ω_n e Ω_N rispettivamente, si ha $P_n(A_n) = P_N(A_N)$ per ogni $N \geq n$. La dimostrazione è semplice: dato che $A_N = A_n \times \{-1, 1\}^{N-n}$, si può scrivere

$$P_N(A_N) = \frac{|A_N|}{|\Omega_N|} = \frac{|A_n| 2^{N-n}}{2^N} = \frac{|A_n|}{2^n} = P_n(A_n).$$

Si noti che non si è usata in alcun modo la forma esplicita dell'evento A_n , data dall'equazione (2.5), ma solo il fatto che $A_n \subseteq \Omega_n$ e che $A_N = A_n \times \{-1, 1\}^{N-n}$.

Abbiamo dunque ottenuto un'importante conclusione: per calcolare la probabilità di un evento che dipende solo dai primi n incrementi – o, equivalentemente, dalle prime n posizioni – della passeggiata aleatoria, si può scegliere come spazio campionario Ω_N , per un qualunque valore di $N \geq n$.

Osservazione 2.6. Data questa arbitrarietà nella scelta dello spazio Ω_N , risulta *naturale* (almeno per un matematico...) considerare lo spazio campionario dato dai cammini di lunghezza infinita $\Omega_\infty := \{-1, 1\}^{\mathbb{N}}$, che contiene in modo canonico Ω_N per ogni $N \in \mathbb{N}$. Il problema è di definire la “giusta” probabilità su Ω_∞ , che estenda (in un senso da precisare) la probabilità uniforme su Ω_n . Si noti che lo spazio Ω_∞ è infinito (più che numerabile), dunque la probabilità uniforme non ha senso. Più in generale, nessuna probabilità discreta su Ω_∞ va bene, come mostreremo nell'Osservazione 3.42. Questo problema ammette una soluzione positiva, che però richiede una nozione più generale di spazio di probabilità, e sarà affrontato nel Capitolo 5. \square

Il problema della ricorrenza

Studiamo finalmente la probabilità che la passeggiata aleatoria ritorni al punto di partenza. Più precisamente, per $n \in \mathbb{N}$ poniamo

$$r_n := P(s_k = 0 \text{ per qualche } k \leq n). \quad (2.6)$$

Il nostro obiettivo è di studiare il comportamento di r_n per $n \rightarrow \infty$.

Sottolineiamo che gli eventi che descriviamo informalmente “a parole” corrispondono sempre a sottoinsiemi dello spazio $\tilde{\Omega}_n$, o equivalentemente di Ω_n , per un opportuno $n \in \mathbb{N}$. Per esempio, nel membro destro in (2.6) compare l’evento costituito dai vettori di incrementi $(x_1, \dots, x_n) \in \Omega_n$ per cui i cammini corrispondenti (s_0, s_1, \dots, s_n) sono tali che $s_k = 0$ per qualche $k \leq n$.

Osserviamo che la successione r_n è crescente, perché vale l’inclusione di eventi

$$\{s_k = 0 \text{ per qualche } k \leq n\} \subseteq \{s_k = 0 \text{ per qualche } k \leq n+1\}, \quad \forall n \in \mathbb{N}.$$

Esiste dunque il limite $r := \lim_{n \rightarrow \infty} r_n$ e inoltre $r \leq 1$, perché $r_n \leq 1$ per ogni $n \in \mathbb{N}$ (r_n una probabilità!). Non è chiaro a priori se $r = 1$ oppure se $r < 1$. Il risultato principale di questo paragrafo è che $r = 1$, una proprietà che viene detta *ricorrenza*.

Teorema 2.7. *La passeggiata aleatoria semplice e simmetrica su \mathbb{Z} è ricorrente:*

$$r := \lim_{n \rightarrow \infty} r_n = \lim_{n \rightarrow \infty} P(s_k = 0 \text{ per qualche } k \leq n) = 1.$$

Verrebbe voglia di affermare che “la probabilità che la passeggiata aleatoria semplice e simmetrica su \mathbb{Z} non ritorni mai all’origine vale 0”, ma per formalizzare rigorosamente questo enunciato servirebbe la nozione di spazio campionario delle passeggiate aleatorie di lunghezza infinita. Per il momento, possiamo affermare che “la probabilità che la passeggiata aleatoria semplice e simmetrica su \mathbb{Z} non ritorni all’origine nei primi n passi tende a 0 per $n \rightarrow \infty$ ”.

La dimostrazione del Teorema 2.7 si articola in diversi passi, alcuni dei quali costituiscono risultati interessanti di per sé. Osserviamo innanzitutto che $r_{2n} = r_{2n+1}$ per ogni $n \in \mathbb{N}$, perché la passeggiata può ritornare all’origine solo dopo un numero pari di passi, pertanto è sufficiente considerare r_{2n} .

Definiamo per $n \in \mathbb{N}$ le quantità

$$\begin{aligned} u_{2n} &:= P(s_{2n} = 0), \\ f_{2n} &:= P(s_2 \neq 0, \dots, s_{2(n-1)} \neq 0, s_{2n} = 0), \end{aligned} \tag{2.7}$$

ossia u_{2n} è la probabilità che la passeggiata aleatoria valga 0 al passo $2n$, mentre f_{2n} è la probabilità che la passeggiata aleatoria ritorni a 0 *per la prima volta* al passo $2n$. Le quantità u_{2n} giocheranno un ruolo fondamentale tra poco. Per il momento, mostriamo come esprimere r_{2n} in funzione delle f_{2k} . Dire che la passeggiata aleatoria visita zero in un qualche passo $k \leq 2n$ è equivalente a dire che il primo ritorno a zero avviene prima di $2n$ passi: si ha dunque l’uguaglianza di eventi

$$\{s_k = 0 \text{ per qualche } k \leq 2n\} = \bigcup_{m=1}^n \{s_2 \neq 0, \dots, s_{2(m-1)} \neq 0, s_{2m} = 0\},$$

e inoltre gli eventi che appaiono nell'unione sono disgiunti (perché?). Per l'additività della probabilità, si ha pertanto l'uguaglianza

$$r_{2n} = \sum_{k=1}^n f_{2k}. \quad (2.8)$$

Per inciso, questa relazione conferma che r_{2n} è una successione crescente, perché somma di termini positivi.

Grazie alla relazione (2.8), per dimostrare il Teorema 2.7 resta da mostrare che $\sum_{k \in \mathbb{N}} f_{2k} = 1$. Saranno fondamentale i due lemmi seguenti.

Lemma 2.8. *Per ogni $n > 0$*

$$u_{2n} = \sum_{k=1}^n f_{2k} u_{2(n-k)}.$$

Dimostrazione. Sia $A := \{s_{2n} = 0\}$, per cui $u_{2n} = P(A)$. L'evento A può essere decomposto secondo il primo istante in cui la passeggiata aleatoria ritorna a zero, ossia A si può scrivere come unione dei seguenti n eventi disgiunti:

$$A = \bigcup_{k=1}^n A_k, \quad A_k := \{s_2 \neq 0, \dots, s_{2k-2} \neq 0, s_{2k} = 0, s_{2n} = 0\}.$$

Contiamo i cammini in A_k . La cardinalità di A_k è uguale al *numero* di cammini di lunghezza $2k$ che ritornano a 0 la prima volta dopo $2k$ passi, ossia $2^{2k} f_{2k}$, moltiplicato il *numero* di cammini di lunghezza $2n - 2k$ che terminano in 0, ossia $2^{2(n-k)} u_{2(n-k)}$. Pertanto

$$P(A_k) = \frac{|A_k|}{2^{2n}} = \frac{2^{2k} f_{2k} 2^{2(n-k)} u_{2(n-k)}}{2^{2n}} = f_{2k} u_{2(n-k)}.$$

Essendo $P(A) = \sum_{k=1}^n P(A_k)$, la conclusione segue facilmente. \square

Lemma 2.9. *Siano $(a_n)_{n \geq 0}$, $(b_n)_{n \geq 1}$ due successioni di numeri reali positivi tali che*

$$a_0 := 1, \quad a_n = \sum_{k=1}^n b_k a_{n-k}, \quad \forall n \geq 1. \quad (2.9)$$

Allora

$$\sum_{k=1}^{+\infty} b_k \geq 1 \quad \Longleftrightarrow \quad s := \sum_{n=1}^{+\infty} a_n = +\infty.$$

Dimostrazione. Ricordiamo che, per la somma a blocchi (0.11), è lecito permutare l'ordine degli addendi di somme infinite a termini positivi. Pertanto

$$\begin{aligned} s &= \sum_{n=1}^{+\infty} a_n = \sum_{n=1}^{+\infty} \left(\sum_{k=1}^n b_k a_{n-k} \right) = \sum_{k=1}^{\infty} b_k \left(\sum_{n=k}^{+\infty} a_{n-k} \right) \\ &= \sum_{k=1}^{+\infty} b_k \left(\sum_{m=0}^{+\infty} a_m \right) = (1+s) \sum_{k=1}^{+\infty} b_k, \end{aligned}$$

avendo usato le relazioni in (2.9). Di conseguenza, se $s < +\infty$, allora

$$\sum_{k=1}^{+\infty} b_k = \frac{s}{1+s} < 1.$$

Resta da dimostrare che, se $s = +\infty$, si ha $\sum_{k=1}^{+\infty} b_k \geq 1$. Per ogni $N \in \mathbb{N}$ si ha

$$\begin{aligned} \sum_{n=1}^N a_n &= \sum_{n=1}^N \left(\sum_{k=1}^n b_k a_{n-k} \right) = \sum_{k=1}^N b_k \left(\sum_{n=k}^N a_{n-k} \right) = \sum_{k=1}^N b_k \left(\sum_{m=0}^{N-k} a_m \right) \\ &\leq \sum_{k=1}^N b_k \left(\sum_{m=0}^N a_m \right) = \sum_{k=1}^N b_k \left(1 + \sum_{m=1}^N a_m \right), \end{aligned}$$

pertanto

$$\sum_{k=1}^N b_k \geq \frac{\sum_{n=1}^N a_n}{1 + \sum_{n=1}^N a_n}, \quad \forall N \in \mathbb{N}.$$

Passando al limite per $N \rightarrow +\infty$, se $s = +\infty$ si ottiene $\sum_{k=1}^{+\infty} b_k \geq 1$. \square

Grazie al Lemma 2.8, possiamo applicare il Lemma 2.9 con $a_n = u_{2n}$ e $b_n = f_{2n}$. Nel nostro caso, ricordando (2.8) e il fatto che $r_n \leq 1$ per ogni $n \in \mathbb{N}$, già sappiamo che $r = \lim_{n \rightarrow \infty} r_{2n} = \sum_{k=1}^{\infty} f_{2k} \leq 1$, pertanto

$$\lim_{n \rightarrow +\infty} r_{2n} = \sum_{k=1}^{+\infty} f_{2k} = 1 \quad \Longleftrightarrow \quad \sum_{n=1}^{\infty} u_{2n} = +\infty.$$

Resta solo da mostrare la divergenza della serie $\sum_{n=1}^{\infty} u_{2n}$ e avremo completato la dimostrazione del Teorema 2.7. Vale la seguente espressione esplicita per u_{2n} :

$$u_{2n} = P(s_{2n} = 0) = \frac{1}{2^{2n}} \binom{2n}{n}. \quad (2.10)$$

La dimostrazione è semplice: un vettore di incrementi $(x_1, \dots, x_{2n}) \in \Omega_{2n}$ determina un cammino con $s_{2n} = 0$ se e solo se esattamente n incrementi x_i valgono $+1$ (e dunque gli altri n valgono -1); di conseguenza, tali vettori sono tanti quante le combinazioni di n elementi estratti da un insieme che ne contiene $2n$, ossia $\binom{2n}{n}$.

Grazie alla formula (2.10), mostriamo infine che

$$\sum_{n=1}^{\infty} u_{2n} = +\infty. \quad (2.11)$$

Usando la formula di Stirling, per semplicità nella versione (1.23), abbiamo

$$\begin{aligned} u_{2n} &= \frac{1}{2^{2n}} \binom{2n}{n} = \frac{1}{2^{2n}} \frac{(2n)!}{(n!)^2} = \frac{1}{2^{2n}} \frac{(2n)^{2n} e^{-2n} \sqrt{2n} e^{\frac{\theta(2n)}{24n}}}{\sqrt{2\pi} n^{2n} e^{-2n} n e^{\frac{\theta(n)}{6n}}} \\ &= \frac{1}{\sqrt{\pi n}} \exp \left[\frac{\theta(2n)}{24n} - \frac{\theta(n)}{6n} \right], \end{aligned}$$

da cui segue che, essendo $0 \leq \theta(\cdot) \leq 1$,

$$\lim_{n \rightarrow +\infty} \frac{u_{2n}}{1/\sqrt{n}} = \frac{1}{\sqrt{\pi}}. \quad (2.12)$$

Per il criterio del confronto asintotico tra serie, ricordando la relazione (0.2), si ricava (2.11), e questo conclude la dimostrazione del Teorema 2.7.

Il caso multidimensionale

Concludiamo questo paragrafo con una generalizzazione multidimensionale, ossia su \mathbb{Z}^d , della passeggiata aleatoria semplice e simmetrica. Per $d \geq 1$, consideriamo i cammini di lunghezza n “generati” dallo spazio di incrementi

$$\Omega_n^d = (\{-1, +1\}^d)^n = \{\mathbf{x} = (x_1, x_2, \dots, x_n) : x_i \in \{-1, 1\}^d \text{ per } i = 1, 2, \dots, n\}.$$

Ciò significa che consideriamo cammini uscenti dall'origine di \mathbb{Z}^d e la cui posizione al tempo k è $s_k := x_1 + x_2 + \dots + x_k \in \mathbb{Z}^d$, dove $x_i \in \{-1, 1\}^d$. Tutti questi cammini si assumono equiprobabili, cioè la probabilità P su Ω_n^d è quella uniforme. Si noti che se $A \subseteq \Omega_n = \{-1, 1\}^n$ allora $A^d = A \times A \times \dots \times A \subseteq \Omega_n^d$ e vale la formula

$$|A^d| = |A|^d.$$

Consideriamo allora l'evento $A_{2n}^{(d)} := \{s_{2n} = 0\}$ (in Ω_{2n}^d , con $N \geq 2n$), dove 0 denota l'origine di \mathbb{Z}^d , e denotiamo con $u_{2n}^{(d)}$ la sua probabilità. In particolare, $u_{2n}^{(1)} = u_{2n}$ è la quantità già introdotta in (2.7) e studiata sopra. Poiché $s_{2n} = 0$ se e solo se tutte le d componenti sono uguali a zero, si ha

$$A_{2n}^{(d)} = (A_{2n}^{(1)})^d,$$

pertanto

$$u_{2n}^{(d)} = P(A_{2n}^{(d)}) = \frac{|A_{2n}^{(d)}|}{|\Omega_N^d|} = \frac{|A_{2n}^{(1)}|^d}{|\Omega_N^d|^d} = (u_{2n})^d. \quad (2.13)$$

In modo analogo al caso unidimensionale, possiamo definire come in (2.7)

$$f_{2k}^{(d)} := P(s_2 \neq 0, \dots, s_{2(k-1)} \neq 0, s_{2k} = 0).$$

Allora la quantità $r_{2n}^{(d)}$, definita da

$$r_{2n}^{(d)} := \sum_{k=1}^n f_{2k}^{(d)}$$

è la probabilità che la passeggiata aleatoria torni all'origine entro $2n$ passi, esattamente come in (2.8). Anche la relazione

$$u_{2n}^{(d)} = \sum_{k=1}^n f_{2k}^{(d)} u_{2(n-k)}^{(d)}$$

continua a valere e si dimostra esattamente come nel caso $d = 1$.

Applicando il Lemma 2.9, abbiamo che *la passeggiata aleatoria d -dimensionale è ricorrente*, cioè $\lim_{n \rightarrow \infty} r_{2n}^{(d)} = 1$, se e solo se $\sum_{n=1}^{\infty} u_{2n}^{(d)} = +\infty$. D'altra parte, dalle relazioni (2.13) e (2.12) segue immediatamente che $u_{2n}^{(d)}$ è asintoticamente equivalente a $n^{-d/2}$. Dato che $\sum_{n \in \mathbb{N}} n^{-d/2} < \infty$ se e solo se $d > 2$, per la relazione (0.2), possiamo concludere quanto segue.

Teorema 2.10. *La passeggiata aleatoria semplice e simmetrica in dimensione d è ricorrente per $d = 1, 2$ e non è ricorrente per $d \geq 3$.*

In analogia con le considerazioni espresse dopo il Teorema 2.7, un modo suggestivo per formulare il Teorema 2.10 consiste nell'affermare che “la passeggiata aleatoria in dimensione ≥ 3 ha una probabilità strettamente positiva di non ritornare mai all'origine”. Tuttavia, non avendo introdotto lo spazio di probabilità delle passeggiate aleatorie di lunghezza infinita, ci dobbiamo accontentare della seguente descrizione: “se $d \geq 3$, esiste una costante $\varepsilon > 0$ tale che la probabilità che la passeggiata in dimensione ≥ 3 non sia mai tornata all'origine *nei primi n passi* è maggiore di ε , per ogni $n \in \mathbb{N}$ ”.

2.3 Statistiche classiche e quantistiche

Coefficienti multinomiali

Dato un insieme A , ricordiamo che una partizione di A è una famiglia ordinata (A_1, \dots, A_r) di sottoinsiemi di A tali che $A_i \cap A_j = \emptyset$, per ogni $i \neq j$, e $\bigcup_{i=1}^r A_i = A$. Sottolineiamo che l'ordine dei sottoinsiemi è importante: (A_1, A_2, A_3) e (A_2, A_1, A_3) sono due partizioni distinte. Supponiamo ora che $|A| = n \in \mathbb{N}$ e siano assegnati $r \in \mathbb{N}$ e $k_1, \dots, k_r \in \mathbb{N}_0$ tali che $k_1 + \dots + k_r = n$: ci proponiamo di calcolare il numero di partizioni (A_1, \dots, A_r) di A tali che $|A_1| = k_1, \dots, |A_r| = k_r$.

Il caso $r = 1$ è banale: esiste infatti un'unica partizione di A in un solo sottoinsieme, che è naturalmente A stesso. Il caso $r = 2$ è strettamente connesso con le combinazioni: infatti una partizione in due sottoinsiemi (A_1, A_2) è identificata da A_1 , poiché

$A_2 = A_1^c$; dovendo essere $|A_1| = k_1$, il numero di tali partizioni coincide con il numero di combinazioni di k_1 elementi estratti da A , pari a $\binom{n}{k_1} = \frac{n!}{k_1!(n-k_1)!} = \frac{n!}{k_1!k_2!}$.

A questo punto non è difficile indovinare la formula nel caso generale: il numero di partizioni (A_1, \dots, A_r) di A tali che $|A_1| = k_1, \dots, |A_r| = k_r$ (dove $k_1 + \dots + k_r = n$) è dato dall'espressione

$$\frac{n!}{k_1!k_2! \cdots k_r!}. \quad (2.14)$$

La dimostrazione è una applicazione del principio fondamentale del calcolo combinatorio. Una partizione (A_1, \dots, A_r) può essere infatti determinata mediante le seguenti scelte successive:

- si sceglie A_1 come sottoinsieme di A di cardinalità k_1 , in $\binom{n}{k_1}$ modi possibili;
- si sceglie A_2 come sottoinsieme di $A \setminus A_1$ di cardinalità k_2 , in $\binom{n-k_1}{k_2}$ modi possibili;
- si procede analogamente per i sottoinsiemi successivi, fino alla scelta di A_{r-1} , per cui ci sono $\binom{n-(k_1+k_2+\dots+k_{r-2})}{k_{r-1}}$ possibilità;
- una volta scelto A_{r-1} , l'ultimo elemento della partizione è automaticamente determinato dalla relazione $A_r = A \setminus (A_1 \cup \dots \cup A_{r-1})$.

Applicando il Teorema 1.26, il numero di partizioni cercato è dato dunque da

$$\binom{n}{k_1} \cdot \binom{n-k_1}{k_2} \cdot \binom{n-(k_1+k_2)}{k_3} \cdots \binom{n-(k_1+k_2+\dots+k_{r-2})}{k_{r-1}},$$

che è una riscrittura di (2.14), come si verifica con qualche semplificazione.

La quantità in (2.14), per $r \in \mathbb{N}$ e $n, k_1, \dots, k_r \in \mathbb{N}_0$ con $k_1 + \dots + k_r = n$, è detta *coefficiente multinomiale*: può essere infatti vista come un'estensione del coefficiente binomiale, a cui si riduce per $r = 2$. I coefficienti multinomiali ricorrono frequentemente nei cosiddetti problemi di occupazione. Un caso molto importante è descritto nell'esempio seguente.

Esempio 2.11. Riprendiamo l'Esempio 1.41, estendendone l'analisi. Supponiamo di disporre casualmente n oggetti in r cassetti, dove $n \in \mathbb{N}_0$, $r \in \mathbb{N}$ (ogni cassetto può contenere un numero qualunque di oggetti). Chiediamoci qual è la probabilità che il primo cassetto contenga k_1 oggetti, il secondo ne contenga k_2 , ..., l' r -esimo ne contenga k_r , dove naturalmente $k_1, \dots, k_r \in \mathbb{N}_0$ sono tali che $k_1 + \dots + k_r = n$.

Lo spazio di probabilità naturale per questo esperimento aleatorio è l'insieme $\Omega_{n,r}$ delle disposizioni con ripetizione di n elementi presi dall'insieme $\{1, \dots, r\}$ (che rappresenta i cassetti), vale a dire

$$\Omega_{n,r} := \{f : \{1, \dots, n\} \rightarrow \{1, \dots, r\}\}, \quad (2.15)$$

munito della probabilità P uniforme. L'evento che ci interessa (che il primo cassetto contenga k_1 oggetti, il secondo ne contenga k_2 , ..., l' r -esimo ne contenga k_r) è

$$O_{k_1, \dots, k_r} := \{f \in \Omega_{n,r} : |f^{-1}(\{1\})| = k_1, \dots, |f^{-1}(\{r\})| = k_r\}.$$

La cardinalità $|O_{k_1, \dots, k_r}|$, cioè il numero di disposizioni di oggetti nei cassetti tali che il primo ne contenga k_1 , il secondo ne contenga k_2, \dots , l' r -esimo ne contenga k_r , è data proprio dal coefficiente multinomiale (2.14). Infatti, identificando con $A := \{1, \dots, n\}$ l'insieme degli oggetti e con $A_i := f^{-1}(\{i\}) \subseteq A$ gli oggetti contenuti nel cassetto i -esimo, le disposizioni $f \in O_{k_1, \dots, k_r}$ sono in corrispondenza biunivoca con le partizioni (A_1, \dots, A_r) di A tali che $|A_1| = k_1, \dots, |A_r| = k_r$. Di conseguenza, per ogni scelta di $k_1, \dots, k_r \in \mathbb{N}_0$ con $k_1 + \dots + k_r = n$, si ha

$$P(O_{k_1, \dots, k_r}) = \frac{|O_{k_1, \dots, k_r}|}{|\Omega_{n,r}|} = \frac{1}{r^n} \frac{n!}{k_1! \dots k_r!}, \quad (2.16)$$

poiché sappiamo che $|\Omega_{n,r}| = r^n$. \square

Maxwell-Boltzmann, Bose-Einstein, Fermi-Dirac

Nell'Esempio 2.11 abbiamo parlato di n oggetti disposti in r cassetti, ma lo stesso schema si può applicare a contesti completamente diversi, come ad esempio:

- (i) gli esiti del lancio di n dadi regolari (o di n lanci di uno stesso dado regolare) a r facce;
- (ii) la disposizione di n passeggeri su un treno composto da r vagoni;
- (iii) la distribuzione di n incidenti in un certo periodo di r giorni;
- (iv) le configurazioni di un sistema di n particelle, ciascuna delle quali può assumere r stati distinti.

In tutti questi casi, spesso non si è interessati a conoscere *quali*, ma soltanto *quanti* “oggetti” ci siano in ciascun “cassetto”. Se indichiamo con k_i il numero di “oggetti” nel “cassetto” i -esimo, ci si concentra cioè sui numeri $k_1, \dots, k_r \in \mathbb{N}_0$, che sono detti *numeri di occupazione*. Per descrivere esperimenti aleatori che coinvolgono tali numeri, risulta dunque naturale introdurre lo spazio campionario

$$\hat{\Omega}_{n,r} := \{(k_1, \dots, k_r) \in (\mathbb{N}_0)^r : k_1 + \dots + k_r = n\}. \quad (2.17)$$

Per quanto visto nell'Esempio 2.11, una probabilità P_{MB} “naturale” sullo spazio $\hat{\Omega}_{n,r}$ è quella corrispondente alla densità p_{MB} definita da

$$p_{MB}((k_1, \dots, k_r)) = P_{MB}(\{(k_1, \dots, k_r)\}) := \frac{1}{r^n} \frac{n!}{k_1! \dots k_r!}, \quad (2.18)$$

in accordo con (2.16). Per ragioni storiche legate alla meccanica statistica, la probabilità P_{MB} è detta *statistica di Maxwell-Boltzmann* ed è efficacemente utilizzata in svariate situazioni, quali ad esempio quelle descritte nei punti (i), (ii), (iii) in alto. L'applicazione della probabilità P_{MB} alla situazione (iv) è invece delicata e merita una discussione più approfondita.

Supponiamo che un sistema fisico sia composto da n particelle, dove ogni particella può assumere r stati distinti. Facciamo inoltre l'ipotesi che, qualunque siano gli

stati delle n particelle, l'energia totale del sistema sia sempre la stessa. Allora, quando il sistema è in equilibrio termico, secondo i principi della meccanica statistica *tutte le configurazioni del sistema sono equiprobabili*.

Il punto fondamentale è decidere che cosa si debba intendere con *configurazione del sistema*. A priori sembra naturale identificare la configurazione del sistema con la n -upla degli stati assunti dalle particelle che lo compongono. In questa ottica, affermare che tutte le configurazioni sono equiprobabili significa considerare la probabilità uniforme sullo spazio $\Omega_{n,r}$ definito in (2.15), cioè sulle disposizioni con ripetizione di n elementi estratti dall'insieme $\{1, \dots, r\}$ (che rappresenta i possibili stati). Di conseguenza, se siamo interessati ai numeri di occupazione, cioè a quante (e non quali) particelle assumano un certo stato, la probabilità “giusta” sullo spazio $\hat{\Omega}_{n,r}$ dei numeri di occupazione, definito in (2.17), sembrerebbe proprio essere la statistica di Maxwell-Boltzmann P_{MB} , introdotta in (2.18).

Tuttavia, se le n particelle che compongono il sistema sono *indistinguibili*, la statistica di Maxwell-Boltzmann conduce a predizioni errate. La ragione è che, in base ai principi della meccanica quantistica, la scelta dello spazio $\Omega_{n,r}$ definito in (2.15) per descrivere le configurazioni del sistema è fondamentalmente sbagliata: infatti, essendo le particelle indistinguibili, non esiste alcun esperimento che permetta di distinguere due elementi di $\Omega_{n,r}$ con gli stessi numeri di occupazione. Di conseguenza, le configurazioni del sistema devono essere *identificate* con le r -uple dei numeri di occupazione, cioè con gli elementi di $\hat{\Omega}_{n,r}$: infatti, non ha fisicamente senso chiedersi *quali*, ma soltanto *quante* particelle assumano un certo stato. L'affermazione che tutte le configurazioni sono equiprobabili conduce dunque a munire $\hat{\Omega}_{n,r}$ della *probabilità uniforme*. Tale probabilità, che indicheremo con P_{BE} , è detta *statistica di Bose-Einstein* ed è dunque definita dalla densità

$$P_{BE}((k_1, \dots, k_r)) = P_{BE}(\{(k_1, \dots, k_r)\}) := \frac{1}{|\hat{\Omega}_{n,r}|}. \quad (2.19)$$

Gli esperimenti hanno effettivamente mostrato che la statistica di Bose-Einstein si applica con successo a sistemi composti da particelle (o atomi) indistinguibili a spin intero, che vengono dette *bosoni* (un esempio è costituito dai fotoni).

Per sistemi composti da particelle a spin semi-intero, dette *fermioni* (ne sono un esempio protoni, neutroni, elettroni, neutrini, ...), il principio di esclusione di Pauli impone l'ulteriore restrizione che *due particelle non possono assumere lo stesso stato*. Questo significa che i numeri di occupazione possono assumere soltanto i valori 0 e 1: in particolare, si deve avere $n \leq r$. Di conseguenza, le configurazioni del sistema sono descritte dall'insieme ristretto

$$\hat{\Omega}'_{n,r} := \{(k_1, \dots, k_r) \in \{0, 1\}^r : k_1 + \dots + k_r = n\}. \quad (2.20)$$

Dovendo essere tutte le configurazioni equiprobabili, la probabilità “corretta” su $\hat{\Omega}'_{n,r}$ è quella uniforme, detta *statistica di Fermi-Dirac* e indicata con P_{FD} :

$$P_{FD}((k_1, \dots, k_r)) = P_{FD}(\{(k_1, \dots, k_r)\}) := \frac{1}{|\hat{\Omega}'_{n,r}|}. \quad (2.21)$$

Anche in questo caso, gli esperimenti hanno mostrato che la probabilità P_{FD} descrive correttamente il comportamento di sistemi composti da fermioni.

Osserviamo che i valori di $|\hat{\Omega}_{n,r}|$ e $|\hat{\Omega}'_{n,r}|$, che compaiono nelle definizioni (2.19) e (2.21), possono essere esplicitati:

$$|\hat{\Omega}_{n,r}| = \binom{n+r-1}{n}, \quad |\hat{\Omega}'_{n,r}| = \binom{r}{n}. \quad (2.22)$$

La seconda relazione segue immediatamente dal fatto che $\hat{\Omega}'_{n,r}$ è in corrispondenza biunivoca naturale con l'insieme delle combinazioni di n elementi estratti da $\{1, \dots, r\}$: la corrispondenza è quella che a $(k_1, \dots, k_r) \in \hat{\Omega}'_{n,r}$ associa il sottoinsieme degli indici $i \in \{1, \dots, r\}$ per cui $k_i = 1$ (si ricordi che $k_i \in \{0, 1\}$).

Per dimostrare la prima relazione in (2.22), mostriamo che $\hat{\Omega}_{n,r}$ è in corrispondenza biunivoca con l'insieme delle combinazioni di $(r-1)$ elementi estratti da $\{1, \dots, n+r-1\}$, che sappiamo avere cardinalità pari a $\binom{n+r-1}{r-1} = \binom{n+r-1}{n}$. Un sottoinsieme di $(r-1)$ elementi di $\{1, \dots, n+r-1\}$ può essere indicato con $\{t_1, \dots, t_{r-1}\}$, con $1 \leq t_1 < t_2 < \dots < t_{r-1} \leq n+r-1$. Se ora definiamo

$$k_1 := t_1 - 1, \quad k_i := t_i - t_{i-1} - 1, \text{ per } 1 < i < r, \quad k_r := (n+r-1) - t_{r-1},$$

si ha per costruzione $k_i \in \mathbb{N}_0$ e $k_1 + \dots + k_r = n$, cioè $(k_1, \dots, k_r) \in \hat{\Omega}_{n,r}$. La corrispondenza ora descritta è biunivoca: la corrispondenza inversa è quella che a (k_1, \dots, k_r) associa il sottoinsieme $\{t_1, \dots, t_{r-1}\}$ di $\{1, \dots, n+r-1\}$ definito da

$$t_1 := k_1 + 1, \quad t_i := t_{i-1} + k_i + 1, \text{ per } 2 \leq i \leq r-1.$$

Una rappresentazione grafica di questa corrispondenza si ottiene disegnando i numeri di occupazione mediante palline separate da sbarrette: per esempio, per $n=7$ e $r=4$, all'elemento $(1, 4, 0, 2) \in \hat{\Omega}_{7,4}$ corrisponde la stringa “ $\circ \mid \circ \circ \circ \circ \mid \mid \circ \circ$ ” composta da $n+r-1=10$ simboli. Le sbarrette identificano allora un sottoinsieme di $\{1, \dots, 10\}$ di cardinalità $r-1=4$.

Le statistiche di Bose-Einstein e Fermi-Dirac, definite dalle relazioni (2.19) e (2.21), sono dette *statistiche quantistiche*, mentre la statistica di Maxwell-Boltzmann, definita in (2.18), è talvolta indicata come *statistica classica*.

Esempio 2.12. Si consideri un sistema di n particelle non interagenti, ciascuna delle quali può assumere r stati differenti, tutti con la stessa energia. Qual è la probabilità $q_{n,r}$ che gli stati assunti dalle particelle siano tutti diversi? Per ipotesi l'energia totale del sistema è indipendente dagli stati delle singole particelle, per cui si può applicare l'analisi svolta in questo paragrafo. Osserviamo inoltre che se $n > r$ si ha ovviamente $q_{n,r} = 0$, per cui supporremo d'ora in avanti che $n \leq r$.

In termini di numeri di occupazione (k_1, \dots, k_r) , dire che gli stati assunti dalle particelle siano tutti diversi significa imporre $k_i \in \{0, 1\}$ per ogni $1 \leq i \leq r$. In altri termini, $q_{n,r}$ è la probabilità dell'insieme $\hat{\Omega}'_{n,r}$ definito in (2.20). Il valore di $q_{n,r}$ dipende naturalmente dal tipo di particelle considerate. Se le particelle sono fermioni

indistinguibili, sappiamo che $\widehat{\Omega}'_{n,r}$ è l'intero spazio di configurazioni ammissibili, per cui ovviamente $q_{n,r}^{FD} = 1$, in accordo col principio di esclusione di Pauli. Se invece le n particelle sono bosoni indistinguibili, occorre usare la statistica di Bose-Einstein definita in (2.19), da cui si ottiene

$$q_{n,r}^{BE} = P_{BE}(\widehat{\Omega}'_{n,r}) = \frac{|\widehat{\Omega}'_{n,r}|}{|\widehat{\Omega}_{n,r}|} = \frac{\binom{r}{n}}{\binom{n+r-1}{n}} = \prod_{i=0}^{n-1} \frac{r-i}{r+i} = \prod_{i=0}^{n-1} \frac{1-\frac{i}{r}}{1+\frac{i}{r}}, \quad (2.23)$$

avendo usato le relazioni in (2.22). Infine, se le particelle sono distinguibili, usando la statistica di Maxwell-Boltzmann (2.18) si ottiene

$$q_{n,r}^{MB} = P_{MB}(\widehat{\Omega}'_{n,r}) = \sum_{(k_1, \dots, k_r) \in \widehat{\Omega}'_{n,r}} P_{MB}((k_1, \dots, k_r)) = \frac{n!}{r^n} |\widehat{\Omega}'_{n,r}| = \prod_{i=0}^{n-1} \left(1 - \frac{i}{r}\right), \quad (2.24)$$

dove si è usata la seconda relazione in (2.22) e il fatto che $P_{MB}((k_1, \dots, k_r)) = n!/r^n$ per ogni $(k_1, \dots, k_r) \in \widehat{\Omega}'_{n,r}$, poiché $k_i \in \{0, 1\}$ per ogni $1 \leq i \leq r$. Dalle relazioni (2.23) e (2.24) segue in particolare che $q_{n,r}^{BE} < q_{n,r}^{MB}$ per ogni $n \geq 1$.

Per avere un'idea più precisa dei risultati ottenuti, supponiamo che r sia grande e che $n \ll r^{2/3}$. Più precisamente, fissiamo un'arbitraria successione positiva $(\varepsilon_r)_{r \in \mathbb{N}}$ tale che $\lim_{r \rightarrow \infty} \varepsilon_r = 0$ e studiamo il comportamento asintotico di $q_{n,r}^{MB}$ e $q_{n,r}^{BE}$ nel limite $r \rightarrow \infty$, assumendo che $n \leq \varepsilon_r r^{2/3}$. Dato che $(1+x) = \exp(x + O(x^2))$ per $x \rightarrow 0$, da (2.24) si ottiene

$$q_{n,r}^{MB} = \exp \left(- \sum_{i=0}^{n-1} \frac{i}{r} + \sum_{i=0}^{n-1} O\left(\frac{i^2}{r^2}\right) \right) = \exp \left(-\frac{1}{2} \frac{n^2}{r} + o(1) \right), \quad (2.25)$$

dove abbiamo usato il fatto che $\sum_{i=0}^{n-1} i = \frac{1}{2}n(n-1)$ e $\sum_{i=0}^{n-1} i^2 = O(n^3) = o(n^2)$, per le relazioni in (0.7). Analogamente, da (2.23) si ottiene:

$$q_{n,r}^{BE} = \exp \left(-\frac{n^2}{r} + o(1) \right). \quad (2.26)$$

Le due relazioni (2.25) e (2.26) da un lato ci dicono che, sia per la statistica di Maxwell-Boltzmann sia per quella di Bose-Einstein, il regime “critico” per la probabilità $q_{n,r}$ è $n \approx \sqrt{r}$: in altri termini, $q_{n,r} \approx 1$ se $n \ll \sqrt{r}$ mentre $q_{n,r} \approx 0$ se $n \gg \sqrt{r}$. Dall'altro lato, se già sapevamo che $q_{n,r}^{BE} < q_{n,r}^{MB}$ per ogni $n \geq 1$, le relazioni (2.25) e (2.26) mostrano che nel regime $n \sim c\sqrt{r}$, con $c \in (0, \infty)$, la differenza tra $q_{n,r}^{MB}$ e $q_{n,r}^{BE}$ resta positiva anche nel limite di r grande: si ha infatti $q_{n,r}^{BE} \rightarrow e^{-c}$ e $q_{n,r}^{MB} \rightarrow e^{-c/2}$.

Ricordiamo infine che la statistica di Maxwell-Boltzmann, adottata per descrivere un sistema di “particelle distinguibili”, si può applicare anche a sistemi “macroscopici” come quelli descritti nell'Esempio 2.11 e nei punti (i), (ii) e (iii) all'inizio del paragrafo. In particolare, $q_{n,r}^{MB}$ coincide con la probabilità che, se n passeggeri salgono su un treno composto da r vagoni e si dispongono a caso, nessun vagone contenga più di un passeggero. \square

2.4 Il modello di Ising in meccanica statistica

Descriviamo ora un celebre modello in meccanica statistica, il *modello di Ising* per un materiale ferromagnetico. Sia Λ un sottoinsieme finito di \mathbb{Z}^d . I punti di Λ vanno interpretati come i *nodi* (detti anche *siti*) di un cristallo regolare. Ogni nodo è occupato da un atomo, il moto dei cui elettroni produce un campo magnetico. In questo modello semplificato, si assume che tale campo magnetico, che chiameremo *spin*, assuma solo due valori, $+1$ e -1 . Una *configurazione* per tale sistema è dunque descritta da

$$\sigma = (\sigma_x)_{x \in \Lambda},$$

dove $\sigma_x = \pm 1$ è lo spin nel nodo $x \in \Lambda$. In altre parole $\Omega = \{-1, +1\}^\Lambda$ è l'insieme di tutte le configurazioni.

L'energia associata ad una configurazione è dovuta all'interazione tra gli spin dei nodi in Λ e all'interazione con l'esterno. In questa presentazione, assumiamo che l'interazione sia *locale*: l'interazione tra gli spin in Λ avviene solo tra siti *primi vicini*, la cui distanza è pari a 1, mentre l'interazione con l'esterno riguarda solo i nodi del “bordo” di Λ , cioè $\partial\Lambda = \{x \in \Lambda : \exists y \in \Lambda^c \text{ tale che } |x - y| = 1\}$, che interagiscono con i siti in Λ^c a distanza uno. Più precisamente, l'energia (o Hamiltoniana) di una configurazione σ è data da

$$H_\Lambda^\tau(\sigma) := -\frac{1}{2} \sum_{\substack{x \in \Lambda, y \in \Lambda \\ |x-y|=1}} \sigma_x \sigma_y - \sum_{\substack{x \in \partial\Lambda, y \in \Lambda^c \\ |x-y|=1}} \sigma_x \tau_y. \quad (2.27)$$

Si noti che il primo termine in H_Λ^τ descrive l'interazione tra gli spin primi vicini in Λ , mentre il secondo termine descrive l'interazione tra gli spin σ_x con $x \in \partial\Lambda$ e gli spin τ_y dell'esterno (il fattore $\frac{1}{2}$ è dovuto semplicemente al fatto che ogni sottoinsieme di due siti primi vicini $\{x, y\} \subseteq \Lambda$ è contato due volte nella prima somma). Assumiamo per semplicità che anche $\tau_y = \pm 1$, cioè $\tau \in \{-1, 1\}^{\Lambda^c}$.

Si noti che se fosse $\tau_y \equiv 0$, l'energia H_Λ^τ avrebbe esattamente due minimi assoluti, dati rispettivamente da $\sigma_x \equiv 1$ e $\sigma_x \equiv -1$; più in generale, l'energia di una configurazione σ sarebbe uguale a quella della configurazione $-\sigma$. La presenza di un campo magnetico al bordo $\tau \in \{-1, 1\}^{\partial\Lambda}$, rompe tale simmetria: in particolare, se $\tau_x \equiv +1$, l'unico minimo di H_Λ^τ è la configurazione con $\sigma_x \equiv +1$. In ogni caso, una configurazione ha un valore tanto più basso dell'energia quanto più gli spin della configurazione sono allineati tra di loro.

Una quantità che gioca un ruolo fisico fondamentale è la temperatura. Il moto termico degli atomi si traduce in un “disturbo aleatorio” sugli spin: il sistema ha una “preferenza” per le configurazioni a bassa energia, ma tale preferenza è tanto più debole tanto più è alta la temperatura. Queste considerazioni intuitive hanno una traduzione precisa in meccanica statistica considerando la *misura di Gibbs* associata all'energia H_Λ , descritta nell'Esempio 1.14, che ora riprendiamo in dettaglio.

Se T è la temperatura assoluta, indichiamo con $\beta = 1/(k_B T)$ la temperatura inversa, dove k_B è la costante di Boltzmann. È conveniente semplificare le notazioni

ponendo $k_B = 1$ (il che equivale a misurare la temperatura in unità di k_B), in modo che $\beta = 1/T$. Secondo l'ipotesi di Gibbs, se il sistema è in equilibrio ad una temperatura inversa $\beta > 0$ con campo magnetico al bordo τ , la probabilità di osservare una configurazione di spin σ è data da

$$p_{\Lambda, \beta}^{\tau}(\sigma) := \frac{1}{Z_{\Lambda}^{\tau}} \exp[-\beta H_{\Lambda}^{\tau}(\sigma)],$$

dove

$$Z_{\Lambda, \beta}^{\tau} := \sum_{\sigma \in \Omega} \exp[-\beta H_{\Lambda}^{\tau}(\sigma)].$$

In questo modo $\sum_{\sigma \in \Omega} p_{\Lambda, \beta}^{\tau}(\sigma) = 1$. Dunque $p_{\Lambda, \beta}^{\tau}$ è la densità discreta di una probabilità $P_{\Lambda, \beta}^{\tau}$ su Ω definita, per $A \subset \Omega$, da

$$P_{\Lambda, \beta}^{\tau}(A) := \sum_{\sigma \in A} p_{\Lambda, \beta}^{\tau}(\sigma).$$

Come abbiamo visto nell'Esempio 1.14, per ogni $\sigma \in \Omega$ si ha che

$$\lim_{\beta \rightarrow 0} P_{\Lambda, \beta}^{\tau}(\{\sigma\}) = \frac{1}{|\Omega|},$$

cioè, nel limite di temperatura infinita, tutte le configurazioni diventano equiprobabili. Inoltre, per ogni $\sigma \in \Omega$ che non sia un minimo assoluto di H_{Λ}^{τ} ,

$$\lim_{\beta \rightarrow +\infty} P_{\Lambda}^{\tau}(\{\sigma\}) = 0,$$

cioè, nel limite di temperatura zero, il sistema tende a “congelarsi” nelle configurazioni che minimizzano l'energia.

Supponiamo ora di fissare il reticolo $\Lambda = \Lambda_n = \{-n, -n+1, \dots, 0, \dots, n-1, n\}^d$ e le condizioni al bordo $\tau_x \equiv 1$. Poniamo $\Omega_n := \{-1, 1\}^{\Lambda_n}$ e scriveremo $P_{n, \beta}^+$ in luogo di $P_{\Lambda_n, \beta}^{\tau}$, $Z_{n, \beta}^+$ in luogo di $Z_{\Lambda_n, \beta}^{\tau}$ e H_n^+ in luogo di $H_{\Lambda_n}^{\tau}$. Introduciamo l'evento

$$A := \{\sigma \in \Omega_n : \sigma_0 = +1\}, \quad (2.28)$$

dove $0 = (0, 0, \dots, 0)$ indica l'origine in \mathbb{Z}^d . Quindi $P_{n, \beta}^+(A)$ è la probabilità che lo spin nell'origine sia positivo. Come suggerito in precedenza, la presenza al bordo di un campo magnetico positivo “favorisce” gli spin positivi rispetto a quelli negativi: di conseguenza, è intuitivamente plausibile che si abbia

$$P_{n, \beta}^+(A) > 1/2.$$

Questa disuguaglianza è effettivamente vera, per ogni valore fissato di $n \in \mathbb{N}$ e $\beta > 0$, ma ne omettiamo per brevità la dimostrazione. Essa esprime il fatto che lo spin nell'origine “risente” del campo magnetico al bordo, e, con probabilità $> 1/2$, si allinea allo stesso. Facendo crescere n , aumenta la distanza tra l'origine e il campo magne-

tico al bordo, la cui influenza, si può congetturare, diventi sempre meno rilevante. In altre parole, potrebbe accadere che

$$\lim_{n \rightarrow +\infty} P_{n,\beta}^+(A) = 1/2. \quad (2.29)$$

Viceversa, potrebbe accadere che l'influenza del campo magnetico al bordo sull'origine sia rilevante anche per n grande, ossia che esista $\varepsilon > 0$ tale che, *per ogni* n ,

$$P_{n,\beta}^+(A) > \frac{1}{2} + \varepsilon. \quad (2.30)$$

Se si verifica la relazione (2.30), si dice che (per il valore di β dato) si ha *magnetizzazione spontanea*.

Per i ferromagneti reali, la magnetizzazione spontanea è un fenomeno effettivamente osservato, purché la temperatura sia non troppo elevata. Il problema che ci poniamo è di stabilire se il modello di Ising, almeno per questo aspetto, è un buon modello per un ferromagnete reale. Il risultato interessante è che la risposta dipende dalla dimensione d dello spazio: per $d \geq 2$ si ha magnetizzazione spontanea a basse temperature, mentre per $d = 1$ non si ha magnetizzazione spontanea per nessun valore della temperatura. Non dovrebbe sorprendere il fatto che l'analisi del modello diventa via via più difficile al crescere della dimensione d . Benché la dimensione "fisica" sia $d = 3$, ci occuperemo per semplicità solo dei casi $d = 1$ e $d = 2$.

Il caso unidimensionale

Per $d = 1$ si ha $\Lambda_n = \{-n, -n+1, \dots, n-1, n\}$ e $\Omega_n = \{-1, 1\}^{\Lambda_n}$. Con un calcolo esplicito possiamo scrivere:

$$\begin{aligned} P_{n,\beta}^+(A) &= \sum_{\sigma \in A} p_{n,\beta}^+(\sigma) = \frac{1}{Z_{n,\beta}^+} \sum_{\sigma \in \Omega_n: \sigma_0=1} e^{\beta(\sigma_{-n} + \sum_{k=-n}^{n-1} \sigma_k \sigma_{k+1} + \sigma_n)} \\ &= \frac{1}{Z_{n,\beta}^+} \sum_{\substack{\sigma_{-n}, \dots, \sigma_{-1} \\ \sigma_1, \dots, \sigma_n}} e^{\beta(\sigma_{-n} + \sum_{k=-n}^{-2} \sigma_k \sigma_{k+1} + \sigma_{-1})} e^{\beta(\sigma_1 + \sum_{k=1}^{n-1} \sigma_k \sigma_{k+1} + \sigma_n)} \\ &= \frac{1}{Z_{n,\beta}^+} \left(\sum_{\sigma_{-n}, \dots, \sigma_{-1}} e^{\beta(\sigma_{-n} + \sum_{k=-(n-1)}^{-1} \sigma_{k-1} \sigma_k + \sigma_{-1})} \right) \\ &\quad \times \left(\sum_{\sigma_1, \dots, \sigma_n} e^{\beta(\sigma_1 + \sum_{k=1}^{n-1} \sigma_k \sigma_{k+1} + \sigma_n)} \right). \end{aligned}$$

Si noti ora che le due somme contenute in quest'ultima espressione sono uguali, cambiando solo i nomi delle variabili sommate ($\sigma_i \leftrightarrow \sigma_{-i}$). Dunque:

$$P_{n,\beta}^+(A) = \frac{1}{Z_{n,\beta}^+} \left\{ \sum_{\sigma_1, \dots, \sigma_n} e^{\beta(\sigma_1 + \sum_{k=1}^{n-1} \sigma_k \sigma_{k+1} + \sigma_n)} \right\}^2. \quad (2.31)$$

Per semplificare questa espressione, introduciamo un operatore lineare T , che agisce sullo spazio vettoriale delle funzioni f da $\{+1, -1\}$ in \mathbb{R} nel modo seguente: la funzione Tf , sempre da $\{+1, -1\}$ in \mathbb{R} , è definita da

$$(Tf)(s) := \sum_{s'=\pm 1} e^{\beta s s'} f(s') = e^{\beta s} f(+1) + e^{-\beta s} f(-1).$$

Una funzione $f : \{+1, -1\} \rightarrow \mathbb{R}$ può essere identificata con il vettore colonna $\begin{pmatrix} f(+1) \\ f(-1) \end{pmatrix}$. In questo modo, la trasformazione $f \rightarrow Tf$ corrisponde alla trasformazione lineare sui vettori di dimensione due data dalla matrice $T_{s,s'} := e^{\beta s s'}$, cioè

$$T = \begin{pmatrix} e^{\beta} & e^{-\beta} \\ e^{-\beta} & e^{\beta} \end{pmatrix}. \quad (2.32)$$

Posto $\varphi(s) = e^{\beta s}$, possiamo riscrivere la relazione (2.31) come

$$P_{n,\beta}^+(A) = \frac{1}{Z_{n,\beta}^+} \left[\sum_{\sigma_1, \dots, \sigma_n} T_{1,\sigma_1} \cdot T_{\sigma_1,\sigma_2} \cdots T_{\sigma_{n-1},\sigma_n} \cdot \varphi(\sigma_n) \right]^2 = \frac{1}{Z_{n,\beta}^+} [(T^n \varphi)(+1)]^2.$$

Si osservi che, se avessimo voluto calcolare $P_n^+(A^c)$, avremmo dovuto sommare sulle configurazioni per le quali $\sigma_0 = -1$. L'unica differenza, rispetto all'espressione in (2.31), è che l'addendo σ_1 nell'esponentiale sarebbe stato sostituito da $-\sigma_1$. Usando le notazioni or ora introdotte, possiamo concludere che

$$P_{n,\beta}^+(A^c) = \frac{1}{Z_{n,\beta}^+} [(T^n \varphi)(-1)]^2.$$

Dato che $P_{n,\beta}^+(A) + P_{n,\beta}^+(A^c) = 1$, otteniamo

$$Z_{n,\beta}^+ = [(T^n \varphi)(+1)]^2 + [(T^n \varphi)(-1)]^2.$$

Mettendo tutto assieme:

$$P_{n,\beta}^+(A) = \frac{[(T^n \varphi)(+1)]^2}{[(T^n \varphi)(+1)]^2 + [(T^n \varphi)(-1)]^2}. \quad (2.33)$$

Usiamo ora un po' di algebra lineare. La matrice T definita in (2.32) ha come autovalori $\lambda_1 = 2 \cosh(\beta)$, $\lambda_2 = 2 \sinh(\beta)$, corrispondenti agli autovettori $v_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ e $v_2 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$. Identificando vettori e funzioni come sopra indicato, possiamo esprimere la funzione $\varphi = \begin{pmatrix} e^{\beta} \\ e^{-\beta} \end{pmatrix}$ come

$$\varphi = \cosh(\beta) v_1 + \sinh(\beta) v_2,$$

da cui, usando la linearità di T , si ottiene

$$(T^n \varphi)(\pm 1) = 2^n \cosh^{n+1}(\beta) v_1 \pm 2^n \sinh^{n+1}(\beta) v_2.$$

In conclusione, abbiamo calcolato l'espressione *esatta* di $P_{n,\beta}^+(A)$:

$$P_{n,\beta}^+(A) = \frac{[\cosh^{n+1}(\beta) + \sinh^{n+1}(\beta)]^2}{[\cosh^{n+1}(\beta) + \sinh^{n+1}(\beta)]^2 + [\cosh^{n+1}(\beta) - \sinh^{n+1}(\beta)]^2}.$$

Si noti che effettivamente $P_{n,\beta}^+(A) > \frac{1}{2}$, per ogni valore fissato di $n \in \mathbb{N}$ e $\beta > 0$ (mentre $P_{n,\beta}^+(A) = \frac{1}{2}$ per ogni $n \in \mathbb{N}$, se $\beta = 0$). Dato che $\cosh(\beta) > \sinh(\beta) > 0$ per ogni $\beta > 0$, lasciamo al lettore il compito di dedurre dalla formula precedente che, per ogni $\beta > 0$, si ha

$$\lim_{n \rightarrow +\infty} P_{n,\beta}^+(A) = \frac{1}{2}.$$

Quindi in dimensione 1 non c'è magnetizzazione spontanea per nessun $\beta > 0$.

Il caso bidimensionale

In dimensione due non tenteremo di effettuare calcoli esatti con il modello di Ising. In realtà, molti calcoli esatti sono possibili: la loro complessità va però al di là del livello di questo libro. Dimosteremo l'esistenza di magnetizzazione spontanea in $d = 2$ (a basse temperature) mediante un argomento geometrico-combinatorio semplice ed efficace, generalizzabile a molti modelli più complessi: si tratta del celebre *argomento di Peierls*.

In questo caso il reticolo $\Lambda_n = \{-n, -n+1, \dots, n-1, n\}^2$ è formato dai punti a coordinate intere del quadrato di lato $2n$ avente l'origine al centro. Poniamo $\Omega_n := \{-1, 1\}^{\Lambda_n}$. Per comodità di calcolo, conviene modificare leggermente la definizione dell'energia, ponendo

$$\tilde{H}_n^+(\sigma) := -\frac{1}{2} \sum_{\substack{x \in \Lambda_n, y \in \Lambda_n \\ |x-y|=1}} (\sigma_x \sigma_y - 1) - \sum_{\substack{x \in \partial \Lambda_n, y \in \Lambda_n^c \\ |x-y|=1}} (\sigma_x - 1). \quad (2.34)$$

Si noti che, con riferimento all'energia originale H_n^+ definita in (2.27), per ogni $\sigma \in \Omega$ si ha $\tilde{H}_n^+(\sigma) = H_n^+(\sigma) + c_n$, dove

$$c_n := \frac{1}{2} |\{(x, y) \in \Lambda_n \times \Lambda_n : |x-y|=1\}| + |\{(x, y) \in \partial \Lambda_n \times \Lambda_n^c : |x-y|=1\}|$$

è una costante che non dipende da σ . Di conseguenza possiamo scrivere

$$P_{n,\beta}^+(\sigma) = \frac{1}{Z_{n,\beta}^+} \exp[-\beta H_n^+(\sigma)] = \frac{1}{\tilde{Z}_{n,\beta}^+} \exp[-\beta \tilde{H}_n^+(\sigma)],$$

dove

$$\tilde{Z}_{n,\beta}^+ := \sum_{\sigma \in \Omega_n} \exp \left[-\beta \tilde{H}_n^+(\sigma) \right]. \quad (2.35)$$

In altre parole, la nuova energia \tilde{H}_n^+ è fisicamente equivalente a quella originale, cioè determina *la stessa misura di Gibbs*. La ragione per introdurre l'energia \tilde{H}_n^+ è che essa si può riscrivere come

$$\begin{aligned} \tilde{H}_n^+(\sigma) = & 2|\{(x, y) \in \Lambda_n \times \Lambda_n : |x - y| = 1 \text{ e } \sigma_x \neq \sigma_y\}| \\ & + 2|\{(x, y) \in \partial\Lambda_n \times \Lambda_n^c : |x - y| = 1 \text{ e } \sigma_x = -1\}|. \end{aligned} \quad (2.36)$$

Vedremo tra poco l'utilità di tale espressione.

Data una configurazione di spin σ su Λ_n , completiamola ad una configurazione di spin su tutto \mathbb{Z}^2 assegnando spin $+1$ a tutti i punti esterni a Λ_n (in realtà, per la costruzione che segue, è sufficiente assegnare spin $+1$ ai punti di Λ_n^c che distano 1 da qualche punto di Λ_n). Per ogni coppia di punti $x, y \in \mathbb{Z}^2$ tali che $|x - y| = 1$ e $\sigma_x \neq \sigma_y$, disegniamo nel piano un segmento di lunghezza 1, ortogonale al segmento congiungente x e y , il cui punto medio sia $\frac{x+y}{2}$. La figura ottenuta dall'unione di tutti i segmenti disegnati è detta *contour* associato alla configurazione di spin (si veda la Figura 2.2 per un esempio). Si noti che l'insieme Ξ_n dei contour ammissibili è in corrispondenza biunivoca con lo spazio delle configurazioni Ω_n , perché, a partire da un contour, si può ricostruire la configurazione di spin.[†] Pertanto, d'ora in avanti identificheremo ogni configurazione σ di spin con il contour \mathcal{C} corrispondente, prendendoci la libertà di scrivere $p_{n,\beta}^+(\mathcal{C})$, $\tilde{H}_n^+(\mathcal{C})$, ecc. (L'insieme Ξ_n può essere caratterizzato esplicitamente, ma non ne avremo bisogno.)

Denotiamo ora con $\ell(\mathcal{C})$ la lunghezza geometrica del contour \mathcal{C} , ossia il numero dei segmenti di lunghezza 1 che lo compongono. In questo modo, ricordando la relazione (2.36), si ottiene la rappresentazione basilare

$$\tilde{H}_n^+(\mathcal{C}) = 2\ell(\mathcal{C}),$$

che permette quindi di scrivere

$$P_{n,\beta}^+(\{\mathcal{C}\}) = \frac{1}{\tilde{Z}_{n,\beta}^+} e^{-2\beta\ell(\mathcal{C})}, \quad \tilde{Z}_{n,\beta}^+ = \sum_{\mathcal{C}' \in \Xi_n} e^{-2\beta\ell(\mathcal{C}')}. \quad (2.37)$$

Osserviamo che, per come è costruito un contour, da ogni punto $(x \pm \frac{1}{2}, y \pm \frac{1}{2})$ con $(x, y) \in \Lambda_n$ partono necessariamente 0, 2 oppure 4 segmenti del contour. Nel caso in cui non ci sia nessun punto da cui partono 4 segmenti, ossia se non ci sono “auto-

[†] Infatti gli spin “esterni” al contour hanno necessariamente valore $+1$, per le condizioni al bordo positive; procedendo verso l'interno, gli spin che “confinano” con spin di valore $+1$ hanno valore -1 ; iterando la procedura, risulta univocamente determinata la configurazione di spin.

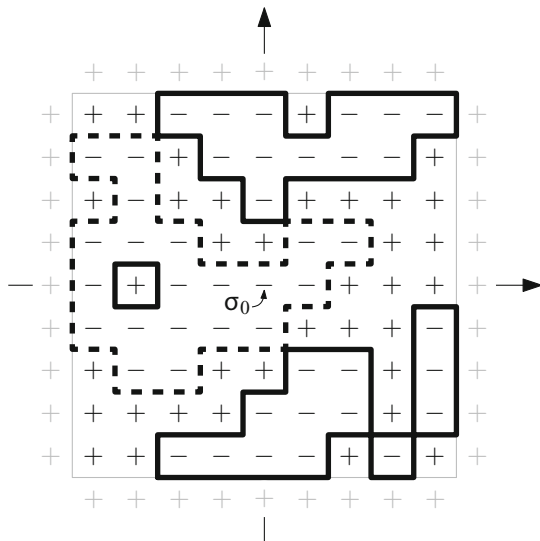


Figura 2.2 Una configurazione di spin σ per il modello di Ising nel piano, sul reticolo $\Lambda_n = \{-n, \dots, n\}^2$ con $n = 4$, con condizioni al bordo positive. Il contour \mathcal{C} corrispondente è descritto dalla linea spessa (sia quella piena sia quella tratteggiata) e ha lunghezza $\ell(\mathcal{C}) = 84$. La parte di contour tratteggiata è un loop che racchiude l'origine

intersezioni”, il contour sarà detto *loop*. Un esempio di loop è dato dalla poligonale tratteggiata in Figura 2.2.[†]

Sia ora \mathcal{C} un contour corrispondente a una configurazione di spin σ in cui $\sigma_0 = -1$. L'osservazione cruciale è la seguente: dato che l'“isola” di spin -1 a cui l'origine appartiene è separata dal “mare” di spin $+1$ che ricopre tutto Λ_n^c , il contour \mathcal{C} *deve necessariamente contenere un loop γ che racchiude l'origine* (si veda la Figura 2.2); in questo caso si può scrivere $\mathcal{C} = \gamma \cup \mathcal{C}'$, dove \mathcal{C}' è un contour che interseca γ in un numero finito di punti, ossia $|\gamma \cap \mathcal{C}'| < \infty$.[‡] Esprimendo l'evento A in (2.28) in termini di contour, si ha dunque la seguente inclusione:

$$A^c \subseteq \{\mathcal{C} \in \Xi_n : \mathcal{C} = \gamma \cup \mathcal{C}', \gamma \text{ loop che racchiude } 0, \mathcal{C}' \in \Xi_n, |\gamma \cap \mathcal{C}'| < \infty\}.$$

[†] Più concretamente, un loop è dato dall'unione $\bigcup_{i=1}^{k-1} \overline{P_i P_{i+1}}$ dei segmenti che congiungono in successione k punti P_1, \dots, P_k del piano, dove $P_i = (x_i \pm \frac{1}{2}, y_i \pm \frac{1}{2})$ con $(x_i, y_i) \in \Lambda_n$, tali che

$$|P_{i+1} - P_i| = 1, \quad P_k = P_1, \quad P_i \neq P_j \text{ se } \{i, j\} \neq \{1, k\}.$$

Sottolineiamo che un loop (e, più in generale, un contour) è un *sottoinsieme* del piano, non una collezione ordinata di segmenti.

[‡] La scelta del loop γ non è univoca, come mostra la Figura 2.2, ma questo non costituisce un problema. Fissato il loop γ , il contour “complementare” \mathcal{C}' si ottiene semplicemente “flippando” (ossia cambiando segno a) tutti gli spin interni al loop.

Dato che $|\gamma \cap \mathcal{C}'| < \infty$, si ha $\ell(\gamma \cup \mathcal{C}') = \ell(\gamma) + \ell(\mathcal{C}')$ e possiamo scrivere

$$\begin{aligned} P_{n,\beta}^+(A^c) &= \frac{1}{\bar{Z}_{n,\beta}^+} \sum_{\mathcal{C} \in A^c} e^{-2\beta \ell(\mathcal{C})} \leq \frac{1}{\bar{Z}_{n,\beta}^+} \sum_{\substack{\gamma \text{ loop} \\ \text{che racchiude 0}}} e^{-2\beta \ell(\gamma)} \sum_{\substack{\mathcal{C}' \in \Xi_n \\ |\gamma \cap \mathcal{C}'| < \infty}} e^{-2\beta \ell(\mathcal{C}')} \\ &\leq \sum_{\substack{\gamma \text{ loop} \\ \text{che racchiude 0}}} e^{-2\beta \ell(\gamma)} \left(\frac{1}{\bar{Z}_{n,\beta}^+} \sum_{\mathcal{C}' \in \Xi_n} e^{-2\beta \ell(\mathcal{C}')} \right) = \sum_{\substack{\gamma \text{ loop} \\ \text{che racchiude 0}}} e^{-2\beta \ell(\gamma)}, \end{aligned}$$

dove la prima disuguaglianza è dovuta alla possibile non unicità della decomposizione $\mathcal{C} = \gamma \cup \mathcal{C}'$, la seconda disuguaglianza è data dalla rimozione del vincolo $|\gamma \cap \mathcal{C}'| < \infty$, mentre per l'ultima uguaglianza si è usata la seconda relazione in (2.37). Indicando con K_m il numero di loop γ che racchiudono l'origine di lunghezza $\ell(\gamma) = m$, per ogni $m \in \mathbb{N}$, otteniamo

$$P_n^+(A^c) \leq \sum_{m=1}^{\infty} K_m e^{-2\beta m}.$$

Ci resta da stimare K_m . Notiamo che ogni loop di lunghezza m , che racchiude 0 al suo interno, è necessariamente contenuto nel quadrato $Q := [-\frac{m}{2}, \frac{m}{2}] \times [-\frac{m}{2}, \frac{m}{2}]$. Costruiamo ora una generica poligonale di lunghezza m nel modo seguente:

- scegliamo un punto in Q della forma $(x \pm \frac{1}{2}, y \pm \frac{1}{2})$, con x, y interi (per questa scelta abbiamo al massimo $(\frac{m}{2} + \frac{m}{2} + 1)^2 = (m+1)^2$ possibilità);
- scegliamo una delle quattro direzioni possibili e tracciamo un segmento di lunghezza uno in quella direzione;
- a questo punto, per il tratto successivo, scegliamo una delle tre direzioni che non ci fanno tornare al punto da cui proveniamo;
- iteriamo la procedura per m passi.

Tra le poligonali costruite in questo modo ci sono in particolare tutti i possibili loop di lunghezza m che intersecano Q , quindi

$$K_m \leq (m+1)^2 \cdot 4 \cdot 3^{m-1} \leq 4 \cdot 9^m,$$

avendo usato per la seconda disuguaglianza la rozza stima $x^2 \leq 3^x$, valida per ogni $x \in \mathbb{N}$. In questo modo, ponendo $c(\beta) := 2\beta - \log 9$, otteniamo

$$P_{n,\beta}^+(A^c) \leq 4 \sum_{m=1}^{+\infty} 9^m e^{-2\beta m} = 4 \sum_{m=1}^{+\infty} e^{-c(\beta)m} = 4 \frac{e^{-c(\beta)}}{1 - e^{-c(\beta)}}, \quad (2.38)$$

dove l'ultima uguaglianza è semplicemente la somma della serie geometrica (0.4). Si noti che il membro destro in (2.38) *non dipende da n* e tende a zero per $\beta \rightarrow +\infty$, dal momento che $\lim_{\beta \rightarrow \infty} c(\beta) = +\infty$. Quindi per ogni $\eta > 0$ esiste $\beta_0 \in (0, \infty)$ tale che per ogni $\beta > \beta_0$ e per ogni $n \in \mathbb{N}$ si ha $P_{n,\beta}^+(A^c) \leq \eta$, ovvero

$$P_{n,\beta}^+(A) \geq 1 - \eta,$$

che è una relazione più forte di (2.30).

In definitiva, abbiamo mostrato che, per grandi valori di β , cioè a temperatura sufficientemente bassa, nel modello di Ising in dimensione 2 ha luogo il fenomeno della magnetizzazione spontanea. È possibile mostrare (non lo faremo) che, al contrario, per valori piccoli di β non si ha magnetizzazione spontanea.

2.5 Il modello di Hardy-Weinberg in genetica

In questo paragrafo studieremo un semplice modello per la trasmissione ereditaria dei caratteri, detto *modello di Hardy-Weinberg*. A differenza dei paragrafi precedenti di questo capitolo, qui è richiesta la nozione di probabilità condizionale.

Consideriamo una popolazione di una determinata specie, e supponiamo di essere interessati ad uno specifico *carattere*. Assumiamo che questo carattere si manifesti solo in due modi, che chiameremo *dominante* e *recessivo*, e che esso sia caratterizzato da una coppia (non ordinata) di *geni*, ognuno dei quali può assumere uno dei due valori a, A . Dunque, ogni individuo della specie possiede una delle tre possibili coppie di geni AA, aA e aa . Nel terzo caso il carattere manifestato sarà quello recessivo, negli altri due quello dominante. Le tre coppie AA, aA e aa sono chiamate *genotipi*.

Il modello ha per oggetto l'evoluzione della frequenza dei genotipi in popolazioni sessuate. In un determinato istante consideriamo gli individui della popolazione, che chiameremo di *generazione* 0. Assumiamo che tale popolazione sia numerosa e usiamo le seguenti notazioni:

- $u \in (0, 1)$ è la *frazione* di individui della generazione 0 di genotipo AA ;
- $2v \in (0, 1)$ è la *frazione* di individui della generazione 0 di genotipo aA ;
- $w \in (0, 1)$ è la *frazione* di individui della generazione 0 di genotipo aa .

Evidentemente $u + 2v + w = 1$. Tali numeri hanno un'evidente interpretazione probabilistica: se si sceglie a caso un individuo della generazione 0, la probabilità che esso sia di genotipo AA (risp. aA, aa) è u (risp. $2v, w$). (La ragione del fattore 2 è che permette di semplificare alcune formule nel seguito.)

Gli individui che sono stati generati da una coppia della generazione 0 costituiscono la *generazione* 1. Supponiamo di scegliere a caso un individuo della generazione 1, e consideriamo, per $g \in \{AA, aA, aa\}$, gli eventi

$$D_g := \text{"l'individuo è di genotipo } g\text{"}. \quad (2.39)$$

Siano

$$u_1 := P(D_{AA}), \quad 2v_1 := P(D_{aA}), \quad w_1 := P(D_{aa}).$$

In altre parole, $u_1, 2v_1$ e w_1 sono le frequenze dei tre genotipi nella generazione 1. Introduciamo anche i seguenti eventi, per $g \in \{AA, aA, aa\}$ e $x \in \{A, a\}$:

$$F_{g,x} := \text{"l'individuo scelto ha padre di genotipo } g, \text{ e ha da lui ereditato il gene } x\text{"};$$

$$M_{g,x} := \text{"l'individuo scelto ha madre di genotipo } g, \text{ e ha da lei ereditato il gene } x\text{"}.$$

Inoltre poniamo

$$F_g := \text{“l'individuo scelto ha padre di genotipo } g\text{”} = F_{g,A} \cup F_{g,a},$$

e analogamente per M_g .

Il modello è definito dalle seguenti ipotesi:

- (i) per ogni scelta di $g, h \in \{AA, aA, aa\}$ e $x, y \in \{A, a\}$, gli eventi $F_{g,x}$ e $M_{h,y}$ sono indipendenti;
- (ii) $P(F_{AA}) = P(M_{AA}) = u$, $P(F_{aA}) = P(M_{aA}) = 2v$, $P(F_{aa}) = P(M_{aa}) = w$;
- (iii) $P(F_{AA,A} | F_{AA}) = P(M_{AA,A} | M_{AA}) = 1$,
 $P(F_{aA,A} | F_{aA}) = P(M_{aA,A} | M_{aA}) = \frac{1}{2}$, $P(F_{aa,A}) = 0$.

Le condizioni (i) e (ii) sono dette di *accoppiamento casuale*, e implicano in particolare che la fertilità è indipendente dal sesso e dal genotipo, e che la frequenza dei genotipi non dipende dal sesso. La condizione (iii) è invece una semplice istanza delle ben note *leggi di Mendel*.

Lo scopo di quanto segue è quello di mostrare che le ipotesi precedenti implicano una relazione funzionale tra le frequenze dei genotipi nella generazione 0 e quelle nella generazione 1. Si noti anzitutto che, ricordando la definizione (2.39), si ha

$$D_{AA} = [F_{AA,A} \cap M_{AA,A}] \cup [F_{aA,A} \cap M_{AA,A}] \cup [F_{AA,A} \cap M_{aA,A}] \cup [F_{aA,A} \cap M_{aA,A}].$$

Evidentemente, le unioni precedenti sono tra eventi disgiunti; inoltre, in ognuna delle intersezioni gli eventi sono, per l'ipotesi (i), indipendenti. Pertanto:

$$\begin{aligned} u_1 := P(D_{AA}) &= P(F_{AA,A})P(M_{AA,A}) + P(F_{aA,A})P(M_{AA,A}) \\ &\quad + P(F_{AA,A})P(M_{aA,A}) + P(F_{aA,A})P(M_{aA,A}). \end{aligned} \quad (2.40)$$

Inoltre, per l'ipotesi (iii)

$$P(F_{AA,A}) = P(M_{AA,A}) = P(F_{AA}) = u, \quad P(F_{aA,A}) = P(M_{aA,A}) = \frac{1}{2} P(F_{aA}) = v.$$

Perciò

$$u_1 = u^2 + 2uv + v^2 = (u + v)^2.$$

Per simmetria, con analoghi calcoli, si ottiene

$$w_1 = (v + w)^2.$$

Essendo

$$\begin{aligned} u_1 + w_1 &= (u + v)^2 + (v + w)^2 = (u + 2v + w)^2 - 2(u + v)(v + w) \\ &= 1 - 2(u + v)(v + w), \end{aligned}$$

si ha

$$2v_1 = 1 - u_1 - w_1 = 2(u + v)(v + w).$$

Quindi, se definiamo la funzione di tre variabili e a valori in \mathbb{R}^3

$$T(u, v, w) = ((u+v)^2, (u+v)(v+w), (v+w)^2),$$

abbiamo mostrato che

$$(u_1, v_1, w_1) = T(u, v, w).$$

In altre parole, la mappa T fornisce la relazione tra le frequenze dei genotipi di una generazione e quelle della successiva.

A questo punto possiamo affermare che le frequenze dei genotipi nella generazione successiva alla generazione 1, diciamo la generazione 2, sono date da

$$(u_2, v_2, w_2) = T(u_1, v_1, w_1) = T(T(u, v, w)).$$

Si osservi che la prima componente di $T(T(u, v, w))$ è

$$[(u+v)^2 + (u+v)(v+w)]^2 = [(u+v)(u+2v+w)]^2 = (u+v)^2,$$

essendo $u+2v+w=1$, e quindi è uguale alla prima componente di $T(u, v, w)$. In modo analogo si mostra che la terza componente di $T(T(u, v, w))$ coincide con la terza componente di $T(u, v, w)$. Pertanto, essendo $u_2+2v_2+w_2=u_1+2v_1+w_1=1$, possiamo concludere che

$$T(T(u, v, w)) = T(u, v, w).$$

Questo implica che *la frequenza dei genotipi dalla generazione 1 in poi rimane costante*: la popolazione raggiunge un equilibrio dopo una sola generazione!

Quello che abbiamo fin qui descritto è la versione standard del modello di Hardy-Weinberg. Ne sono state proposte numerose modifiche e generalizzazioni, per tener conto di fenomeni quali la dipendenza dal sesso dei caratteri e la selezione naturale. Vedremo ora brevemente una di queste versioni modificate, che rappresenta un semplice modello per la *selezione naturale*. Questo modello si ottiene dal precedente modificando l'ipotesi (ii) come segue:

$$(ii)' \quad P(F_{AA}) = P(M_{AA}) = \frac{u}{u+2v}, \quad P(F_{Aa}) = P(M_{Aa}) = \frac{2v}{u+2v},$$

$$P(F_{aa}) = P(M_{aa}) = 0.$$

Il significato di (ii)' è evidente: gli individui che manifestano il carattere recessivo, cioè di genotipo aa non si riproducono.

Lasciamo al lettore il compito di controllare che la relazione in (2.40) diventa:

$$u_1 = \left(\frac{u}{u+2v}\right)^2 + 2\left(\frac{u}{u+2v}\right)\left(\frac{v}{u+2v}\right) + \left(\frac{v}{u+2v}\right)^2 = \left(\frac{u+v}{u+2v}\right)^2.$$

Analogamente si ottiene

$$w_1 = \left(\frac{v}{u+2v}\right)^2, \quad v_1 = \frac{v(u+v)}{(u+2v)^2}.$$

Posto, come prima, $(u_1, v_1, w_1) =: T(u, v, w)$, possiamo definire induttivamente

$$(u_n, v_n, w_n) = T(u_{n-1}, v_{n-1}, w_{n-1}). \quad (2.41)$$

Così $u_n, 2v_n$ e w_n sono le frequenze dei tre genotipi nell' n -esima generazione. Con un calcolo paziente ma elementare, si verifica per induzione che

$$u_n = \left(\frac{u + nv}{u + (n+1)v} \right)^2, \quad v_n = \frac{v(u + nv)}{(u + (n+1)v)^2}, \quad w_n = \frac{v^2}{(u + (n+1)v)^2}.$$

Si noti che che per $n \rightarrow \infty$

$$v_n \sim \frac{1}{n}, \quad w_n \sim \frac{1}{n^2}, \quad \implies \quad u_n = 1 - 2v_n - w_n = 1 - O\left(\frac{1}{n}\right) \longrightarrow 1.$$

Abbiamo dunque quantificato, come effetto della selezione naturale, la progressiva diminuzione dei genotipi che contengono il gene a .

2.6 Note bibliografiche

Il Problema 2.3 è ispirato all'articolo [52] di M. Warshauer e E. Curtin, in cui si mostra che la strategia proposta nella soluzione è ottimale.

Ciò che abbiamo presentato in questo capitolo riguardo alle passeggiate aleatorie e al modello di Hardy-Weinberg è basato sul testo [23] di W. Feller. Consigliamo in particolare la lettura di tale testo per un'introduzione elementare alle passeggiate aleatorie.

I Paragrafi 2.3 e 2.4 contengono alcuni modelli, elementari ma fondamentali, di meccanica statistica. Rimandiamo il lettore alla monografia [27] di G. Gallavotti per un'introduzione, profonda e impegnativa, alla meccanica statistica classica. Le statistiche di Fermi-Dirac e Bose-Einstein si applicano a sistemi quantistici; consigliamo la lettura del testo [29] di K. Hannabuss per un'approccio accessibile, ma matematicamente rigoroso, alla meccanica quantistica, contenente cenni di meccanica statistica quantistica.

L'argomento geometrico-combinatorio che abbiamo utilizzato per mostrare la magnetizzazione spontanea nel modello di Ising in dimensione $d = 2$ è dovuto a R. Peierls [45]. Un'analisi più esplicita, in particolare con il calcolo della *temperatura critica* al di sotto della quale avviene la magnetizzazione spontanea, venne eseguita qualche anno dopo da L. Onsager [42]. Per una panoramica su questo e altri modelli esplicitamente risolubili della meccanica statistica si veda [3].

Il modello di Hardy-Weinberg è uno dei numerosi esempi di modelli matematici motivati dalla genetica; il lettore interessato potrà apprezzare la monografia di W.J. Ewens [22].

Variabili aleatorie discrete: teoria

Sommario. Questo capitolo è interamente dedicato alle variabili aleatorie discrete. Dopo averne introdotto la distribuzione e la densità discreta, definiamo le nozioni di indipendenza e di valor medio e ne studiamo in dettaglio le proprietà. Le principali distribuzioni notevoli discrete su \mathbb{R} , argomento di grande rilevanza applicativa (ma non solo), sono presentate nei paragrafi finali del capitolo, per ragioni di praticità; tuttavia, diverse parti sono accessibili da molto prima, una volta che sia acquisita la nozione di indipendenza per variabili aleatorie.

3.1 Variabili aleatorie e distribuzioni

3.1.1 Considerazioni introduttive

La nozione matematica di *variabile aleatoria* (o *casuale*) formalizza l'idea intuitiva di una quantità che dipende dal caso, o più precisamente di una quantità che è *funzione dell'esito di un esperimento aleatorio*. Riprendendo l'Esempio 1.1 del Capitolo 1, sono esempi di variabili aleatorie:

- (i) il numero X che si ottiene lanciando un dado ordinario a sei facce;
- (ii) il numero di voti Y ottenuti dal politico Cetto alle elezioni del prossimo anno;
- (iii) il tempo di emissione Z di una particella radioattiva da un determinato atomo.

In fatto interessante è che *molte eventi possono essere espressi in termini di variabili aleatorie*, ad esempio

- (i) $A := \text{"Esce un numero pari"} = \{X \in \{2, 4, 6\}\};$
- (ii) $D = \text{"Cetto non ottiene alcun voto alle elezioni del prossimo anno"} = \{Y = 0\}.$
- (iii) $F = \text{"La particella viene emessa dopo più di un anno"} = \{Z > 31\,536\,000\}.$

Le scritture $\{X \in \{2, 4, 6\}\}$, $\{Y = 0\}$ e $\{Z > 31\,536\,000\}$ sono per il momento delle notazioni informali, a cui daremo a breve un senso matematicamente preciso.

Caravenna F., Dai Pra P.: *Probabilità. Un'introduzione attraverso modelli e applicazioni*.

Unitext – La Matematica per il 3+2 67

DOI 10.1007/978-88-470-2595-0_3, © Springer-Verlag Italia 2013

Sottolineiamo la differenza tra *eventi* e *variabili aleatorie*: i primi corrispondono ad *affermazioni* sull'esito di un esperimento aleatorio, mentre le seconde corrispondono a *quantità* che sono funzioni di tale esito. Per un evento A ha senso chiedersi “ A si è verificato?”, e la risposta è “sì” oppure “no”, mentre per una variabile aleatoria X ha senso chiedersi “che valore ha assunto X ?”.

3.1.2 Definizioni

Motivati dalle considerazioni sopra esposte, diamo la seguente definizione.

Definizione 3.1 (Variabile aleatoria discreta). Sia (Ω, P) uno spazio di probabilità discreto, e E un insieme generico. Si dice *variabile aleatoria* (o variabile casuale) *discreta* ogni funzione $X : \Omega \rightarrow E$ definita su Ω a valori in E .

Se $E = \mathbb{R}$, si dice che X è una variabile aleatoria *reale* (o scalare). Se $E = \mathbb{R}^n$ si dice che X è una variabile aleatoria *vettoriale*, o un *vettore aleatorio* di dimensione n .

Questa definizione formalizza matematicamente il concetto di variabile aleatoria come quantità che dipende dal caso: in effetti, se descriviamo un esperimento aleatorio mediante uno spazio di probabilità discreto (Ω, P) , una variabile aleatoria X definita su Ω a valori in E può essere pensata come un elemento $X(\omega)$ di E che è funzione dell'esito $\omega \in \Omega$ dell'esperimento aleatorio (ossia del “caso”).

Con un piccolo abuso di notazione, scriveremo talvolta che una variabile aleatoria X è definita sullo spazio di probabilità discreto (Ω, P) , anche se il dominio della funzione X è l'insieme Ω e non la coppia (Ω, P) .

Esempio 3.2 (Variabili aleatorie costanti). Un caso “banale” di variabile aleatoria è dato da una funzione X costante: $X(\omega) = c$ per ogni $\omega \in \Omega$, dove $c \in E$ è un elemento fissato. Una tale variabile aleatoria è definita su qualunque spazio di probabilità discreto (Ω, P) e verrà identificata con l'elemento $c \in E$. Essa descrive una quantità deterministica, che assume sempre lo stesso valore c , qualunque sia l'esito ω dell'esperimento aleatorio. Si tratta insomma di una variabile non molto casuale... \square

Esempio 3.3 (Variabili aleatorie indicatrici). Sia (Ω, P) uno spazio di probabilità discreto. Per ogni evento fissato $A \subseteq \Omega$, la funzione indicatrice $X = \mathbb{1}_A : \Omega \rightarrow \mathbb{R}$ definisce una variabile aleatoria reale, detta *indicatrice dell'evento* (si ricordi la definizione (0.1)). Infatti X assume solo due valori: $X(\omega) = 1$ se $\omega \in A$, ossia “se l'evento A si verifica”, mentre $X(\omega) = 0$ se $\omega \notin A$, ossia “se l'evento A non si verifica”.

Viceversa, ogni variabile aleatoria reale X che assume solo i valori 0 e 1 è necessariamente un'indicatrice: si ha infatti $X = \mathbb{1}_A$, dove $A := \{\omega \in \Omega : X(\omega) = 1\}$. A dispetto della loro semplicità, le variabili aleatorie indicatrici, dette *di Bernoulli*, giocano un ruolo importante in probabilità. Le incontreremo spesso nel seguito, e ne studieremo sistematicamente le proprietà nel Paragrafo 3.5.2. \square

Vediamo qualche esempio di come gli eventi possano essere espressi in funzioni delle variabili aleatorie.

Esempio 3.4 (Prove ripetute e indipendenti). Consideriamo uno schema di n prove ripetute e indipendenti con probabilità di successo p , ossia uno spazio di probabilità discreto (Ω, \mathcal{P}) in cui sono definiti n eventi C_1, \dots, C_n indipendenti e con la stessa probabilità $p = P(C_i)$ (si ricordi il Paragrafo 1.3.4). Avevamo studiato gli eventi

$A_k :=$ “esattamente k prove hanno successo”,

$B_\ell :=$ “il primo successo si verifica nella ℓ -esima prova”,

per $0 \leq k \leq n$ e $1 \leq \ell \leq n$. Introduciamo ora due variabili aleatorie

$S :=$ “numero di successi nelle n prove”,

$T :=$ “prova in cui si ha il primo successo”,

definite su Ω a valori in $\mathbb{N} \cup \{+\infty\}$ mediante

$$S(\omega) := \sum_{i=1}^n \mathbb{1}_{C_i}(\omega), \quad T(\omega) := \min\{i \in \{1, \dots, n\} : \omega \in C_i\}, \quad (3.1)$$

con la convenzione $\min \emptyset := +\infty$. Possiamo allora esprimere gli eventi A_k e B_ℓ nel modo seguente:

$$A_k = \{\omega \in \Omega : S(\omega) = n\}, \quad B_\ell = \{\omega \in \Omega : T(\omega) = \ell\}. \quad (3.2)$$

Questo mostra che gli eventi A_k e B_ℓ possono essere definiti in modo naturale in termini delle variabili aleatorie S e T , specificandone un sottoinsieme di valori. \square

Definizione 3.5 (Eventi generati da una variabile aleatoria). Se $X : \Omega \rightarrow E$ è una variabile aleatoria, per ogni $A \subseteq E$ indichiamo con $\{X \in A\}$ la controimmagine di A tramite X , ossia

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

In altri termini, $\{X \in A\}$ è il sottoinsieme di Ω (ossia l'evento) costituito da tutti e soli gli esiti ω dell'esperimento aleatorio in esame per cui $X(\omega) \in A$.

Gli eventi del tipo $\{X \in A\}$, per $A \subseteq E$, si dicono *eventi generati da X* .

L'evento $\{X \in A\}$ si può descrivere informalmente come “ X prende valori in A ”. Sebbene si tratti in fondo soltanto di una notazione, la scrittura $\{X \in A\}$ è molto utile perché permette di tradurre in linguaggio matematico affermazioni espresse nel linguaggio ordinario. Ad esempio, se $X : \Omega \rightarrow \mathbb{R}$ è una variabile aleatoria che descrive il numero che si ottiene lanciando un dado a sei facce, l'affermazione “il numero che si ottiene è pari” viene tradotta con la scrittura $\{X \in \{2, 4, 6\}\}$, che definisce un sottoinsieme di Ω , ossia un evento. Questo giustifica le notazioni usate negli esempi introduttivi discussi poco sopra, nel Paragrafo 3.1.1.

Osservazione 3.6. È prassi scrivere $\{X = x\}$ invece di $\{X \in \{x\}\}$ e analogamente, se X è una variabile aleatoria reale, $\{X > a\}$ invece di $\{X \in (a, \infty)\}$, ecc. Allo stesso modo, se X, Y sono variabili aleatorie reali definite sullo stesso spazio di probabilità, scriveremo $\{X = Y\}$ invece di $\{X - Y = 0\}$, ecc. \square

3.1.3 Probabilità discrete su spazi generali

Finora abbiamo concentrato la nostra attenzione su insiemi Ω finiti o numerabili e abbiamo definito probabilità ogni funzione $P: \mathcal{P}(\Omega) \rightarrow \mathbb{R}$ che soddisfa le proprietà (P1) e (P2) della Definizione 1.5. Come vedremo nel Capitolo 5, quando Ω è infinito più che numerabile è opportuno dare una definizione più ampia, permettendo a una probabilità di non essere definita per *tutti* i sottoinsiemi dello spazio, ma soltanto per una opportuna sottoclasse.

Per il momento, introduciamo una *classe speciale* di probabilità su insiemi generali, definite su tutti i sottoinsiemi dello spazio, sfruttando la Proposizione 1.10 (che è stata dimostrata per insiemi generici, anche infiniti più che numerabili). Per comodità di notazione, chiameremo E invece di Ω lo spazio.

Definizione 3.7 (Probabilità discrete su insiemi generali). Sia E un generico insieme non vuoto (anche infinito più che numerabile) e sia q una *densità discreta* su E , ossia una funzione $q: E \rightarrow \mathbb{R}$ tale che

$$q(x) \geq 0, \quad \forall x \in E, \quad \sum_{x \in E} q(x) = 1. \quad (3.3)$$

La funzione $\mu: \mathcal{P}(E) \rightarrow \mathbb{R}$ definita da

$$\mu(A) = \sum_{x \in A} q(x), \quad \forall A \subseteq E, \quad (3.4)$$

sarà detta *probabilità discreta su E* (associata alla densità discreta q). Essa soddisfa le proprietà le proprietà (P1) e (P2) della Definizione 1.5, ossia:

(P1) $\mu(E) = 1$.

(P2) Per ogni successione $(A_n)_{n \in \mathbb{N}}$ di sottoinsiemi di E disgiunti, tali cioè che $A_n \cap A_m = \emptyset$ se $n \neq m$, si ha

$$\mu\left(\bigcup_{n=1}^{+\infty} A_n\right) = \sum_{n=1}^{+\infty} \mu(A_n).$$

Per il Teorema 1.11, se E è un insieme finito o numerabile, tutte le probabilità su E sono discrete, secondo la Definizione 3.7. Nel Capitolo 5 daremo una nozione più generale di probabilità e vedremo che, in spazi infiniti più che numerabili, vi sono probabilità non discrete, ossia *non esprimibili in termini di una densità discreta*.

Osservazione 3.8. Continueremo a chiamare *spazio di probabilità discreto* una coppia (Ω, P) dove Ω è un insieme *finito o numerabile* e P è una probabilità su Ω . Per il

momento non diamo un nome alla coppia (Ω, \mathbf{P}) quando Ω è infinito più che numerabile e \mathbf{P} è una probabilità discreta su Ω . \square

Vale la pena di sottolineare che un insieme infinito anche più che numerabile, munito di una probabilità discreta, si può sempre facilmente *riconduire a uno spazio di probabilità discreto*, come spieghiamo nell'osservazione seguente. In questo senso, l'estensione data dalla Definizione 3.7 è più formale che sostanziale.

Osservazione 3.9. Sia μ una probabilità discreta su un insieme E infinito più che numerabile, associata alla densità discreta q mediante la relazione (3.4). Dato che la famiglia di numeri reali $(q(x))_{x \in E}$ ha somma finita – si ricordi la relazione (3.3) – il sottoinsieme

$$\tilde{E} := \{x \in E : q(x) > 0\} \subseteq E,$$

che consiste dei punti in cui la densità è non nulla, è necessariamente *finito o numerabile*, per quanto visto in (0.9). Inoltre, essendo $q(x) = 0$ se $x \notin \tilde{E}$, si ha

$$\mu(\tilde{E}) = \sum_{x \in \tilde{E}} q(x) = \sum_{x \in E} q(x) = 1,$$

per le relazioni (3.4) e (3.3).

In altri termini, anche se l'insieme E è infinito più che numerabile, *ogni probabilità discreta μ su E è necessariamente “concentrata” su un sottoinsieme finito o numerabile $\tilde{E} \subseteq E$* . In particolare, non si perde niente restringendo l'attenzione allo spazio di probabilità discreto $(\tilde{E}, \mu)^\dagger$ dal momento che $\mu(A) = \mu(A \cap \tilde{E})$ per ogni sottoinsieme $A \subseteq E$ (si veda l'Esercizio 1.3). \square

3.1.4 Distribuzione e densità discreta

A ogni variabile aleatoria $X : \Omega \rightarrow E$ è associato un oggetto di fondamentale importanza, la *distribuzione* μ_X di X , che è una probabilità sullo spazio di arrivo E . Dato che l'insieme E può anche essere più che numerabile, ci dobbiamo riferire alla nozione di *probabilità discreta* introdotta nella Definizione 3.7.

Definizione 3.10 (Distribuzione e densità discreta). Sia $X : \Omega \rightarrow E$ una variabile aleatoria, definita su uno spazio di probabilità discreto (Ω, \mathbf{P}) a valori in un insieme arbitrario E .

- Si dice *distribuzione* (o *legge*) di X l'applicazione $\mu_X : \mathcal{P}(E) \rightarrow [0, 1]$ definita da

$$\mu_X(A) := \mathbf{P}(X \in A), \quad \forall A \subseteq E. \quad (3.5)$$

- Si dice *densità discreta* di X l'applicazione $p_X : E \rightarrow [0, 1]$ definita da

$$p_X(x) := \mathbf{P}(X = x) = \mu_X(\{x\}), \quad \forall x \in E. \quad (3.6)$$

† Per non appesantire le notazioni, indichiamo con μ anche la restrizione di μ a \tilde{E} .

Si noti che scriviamo $P(X \in A)$ e $P(X = x)$ invece di $P(\{X \in A\})$ e $P(\{X = x\})$, per alleggerire le notazioni.

Proposizione 3.11. *Sia $X : \Omega \rightarrow E$ una variabile aleatoria. La funzione p_X è una densità discreta su E , ossia*

$$p_X(x) \geq 0, \quad \forall x \in E; \quad \sum_{x \in E} p_X(x) = 1, \quad (3.7)$$

e la distribuzione μ_X è la probabilità discreta su E ad essa associata, ossia

$$\mu_X(A) = \sum_{x \in A} p_X(x), \quad \forall A \subseteq E. \quad (3.8)$$

Dimostrazione. È chiaro dalla definizione (3.6) che $p_X(x) \geq 0$ per ogni $x \in E$, quindi la prima relazione in (3.7) è verificata. Indichiamo con $X(\Omega) \subseteq E$ l'immagine di X :

$$X(\Omega) := \{X(\omega) : \omega \in \Omega\} = \{x \in E : \exists \omega \in \Omega \text{ tale che } X(\omega) = x\}. \quad (3.9)$$

Notiamo che si tratta di un insieme numerabile, perché è l'immagine della funzione X , il cui dominio Ω è numerabile. Per ogni $A \subseteq E$ possiamo scrivere

$$\{X \in A\} = \bigcup_{x \in A \cap X(\Omega)} \{X = x\}, \quad (3.10)$$

perché, per la Definizione 3.5,

$$\omega \in \{X \in A\} \iff X(\omega) \in A \iff \exists x \in A : X(\omega) = x,$$

e nell'ultima espressione x deve necessariamente appartenere all'immagine $X(\Omega)$. L'unione nel membro destro di (3.10) è dunque al più numerabile – perché l'insieme $X(\Omega)$ è numerabile – e *disgiunta*: infatti, se esiste $\omega \in \{X = x\} \cap \{X = y\}$, significa che $X(\omega) = x$ e $X(\omega) = y$, dunque $x = y$. Di conseguenza, per la σ -addittività di P , ricordando le definizioni (3.5) e (3.6),

$$\mu_X(A) = P(X \in A) = \sum_{x \in A \cap X(\Omega)} P(X = x) = \sum_{x \in A \cap X(\Omega)} p_X(x) = \sum_{x \in A} p_X(x),$$

dove l'ultima uguaglianza è dovuta al fatto che $p_X(x) = P(X = x) = 0$ se $x \notin X(\Omega)$, perché in tal caso $\{X = x\} = \emptyset$. Abbiamo dunque dimostrato la relazione (3.8). Scegliendo $A = E$, otteniamo anche la seconda relazione in (3.7). \square

L'importanza della distribuzione μ_X di una variabile causale X risiede nel fatto che essa descrive *quali valori vengono assunti da X e con quali probabilità*. Se si vuole calcolare la probabilità “che X faccia qualcosa” (ossia di un evento generato da X , nel senso della Definizione 3.5) non è necessario conoscere in dettaglio come è definita la funzione X , ma ci basta sapere qual è la sua distribuzione.

Esempio 3.12. Consideriamo il lancio di un dado regolare a sei facce. Lo spazio di probabilità naturale per questo esperimento è $\Omega = \{1, 2, 3, 4, 5, 6\}$ munito della pro-

babilità uniforme P . Il numero ottenuto lanciando il dado è rappresentato dalla variabile aleatoria $X : \Omega \rightarrow \mathbb{R}$, data semplicemente da $X(\omega) = \omega$ per ogni $\omega \in \Omega$. (Naturalmente si potrebbe considerare che X prenda valori in \mathbb{N} , o anche in $\{1, 2, 3, 4, 5, 6\}$: c'è sempre una certa arbitrarietà nella scelta dello spazio di arrivo.)

Supponiamo ora che il lancio del dado sia parte di un esperimento aleatorio più ampio, che comprende anche il lancio di una moneta equilibrata. Per questo nuovo esperimento, scegliamo lo spazio di probabilità (Ω', P') dove

$$\Omega' = \{1, 2, 3, 4, 5, 6\} \times \{T, C\} = \{1T, 2T, 3T, 4T, 5T, 6T, 1C, 2C, 3C, 4C, 5C, 6C\}$$

e P' è la probabilità uniforme su Ω' . La variabile aleatoria che corrisponde al numero ottenuto dal lancio del dado è ora rappresentata da una *funzione diversa* $X' : \Omega' \rightarrow \mathbb{R}$, dato che il dominio di definizione è diverso: si ha infatti $X'(ia) := i$ per ogni $\omega' = ia \in \Omega'$ (con $1 \leq i \leq 6$ e $a \in \{T, C\}$).

Si noti che le variabili aleatorie X e X' , definite sugli spazi diversi Ω e Ω' , prendono valori nello stesso spazio \mathbb{R} e hanno la stessa distribuzione, ossia $\mu_X = \mu_{X'}$. Questo è intuitivamente chiaro: X e X' rappresentano entrambe il numero ottenuto dal lancio di un dado regolare, dunque necessariamente devono assumere gli stessi valori con le stesse probabilità. La verifica formale è facile: basta mostrare che X e X' hanno la stessa densità discreta $p_X = p_{X'}$. Chiaramente $p_X(x) = P(X = x) = 0$ se $x \notin \{1, 2, 3, 4, 5, 6\}$, perché $\{X = x\} = \emptyset$ (non c'è nessun $\omega \in \Omega$ tale che $X(\omega) = x$) e analogamente $p_{X'}(x) = 0$. D'altro canto, per $x \in \{1, 2, 3, 4, 5, 6\}$

$$p_X(x) = P(X = x) = P(\{x\}) = \frac{|\{x\}|}{|\Omega|} = \frac{1}{6},$$

perché $\{X = x\} = \{\omega \in \Omega : X(\omega) = x\} = \{x\}$, e analogamente

$$p_{X'}(x) = P'(X' = x) = P'(\{xT, xC\}) = \frac{|\{xT, xC\}|}{|\Omega'|} = \frac{2}{12} = \frac{1}{6}.$$

In definitiva, X e X' hanno la stessa distribuzione, ossia la probabilità discreta μ su \mathbb{R} associata alla densità discreta $q(x) = p_X(x) = p_{X'}(x) = \frac{1}{6} \mathbb{1}_{\{1, 2, 3, 4, 5, 6\}}(x)$. La conoscenza della distribuzione μ permette di calcolare le probabilità di interesse senza lavorare direttamente con le variabili aleatorie X e X' . Per esempio, la probabilità che il numero ottenuto lanciando il dado sia pari vale (ovviamente)

$$P(X \in \{2, 4, 6\}) = P'(X' \in \{2, 4, 6\}) = \mu(\{2, 4, 6\}) = q(2) + q(4) + q(6) = \frac{1}{2}.$$

Questo mostra tra l'altro che, al fine di calcolare probabilità che coinvolgono la singola variabile aleatoria X o X' , è indifferente lavorare usare una o l'altra: avendo la stessa distribuzione, le corrispondenti probabilità sono uguali. \square

Nel prossimo paragrafo vedremo diversi esempi interessanti di determinazione della distribuzione di variabili aleatorie. Mostriamo ora due risultati teorici di cui si fa uso frequente (talvolta in modo implicito). Cominciamo con un'utile definizione.

Definizione 3.13. Sia (Ω, \mathcal{P}) uno spazio di probabilità discreto. Due variabili aleatorie X, X' , definite su Ω a valori nello stesso insieme E , si dicono *quasi certamente uguali* se $P(X = X') = P(\{\omega \in \Omega : X(\omega) = X'(\omega)\}) = 1$.

In altre parole, due variabili aleatorie X e X' sono quasi certamente uguali se differiscono su un insieme di probabilità nulla: $P(X \neq X') = 0$, o equivalentemente (perché?) se $P(\{\omega\}) = 0$ per ogni $\omega \in \Omega$ per cui $X(\omega) \neq X'(\omega)$. Si ha allora:

Proposizione 3.14. Due variabili aleatorie X, X' quasi certamente uguali hanno la stessa distribuzione (e dunque la stessa densità discreta).

Dimostrazione. Siano (Ω, \mathcal{P}) e E lo spazio di probabilità e lo spazio di arrivo rispettivamente di X e X' . Per ipotesi l'evento $C := \{X = X'\} = \{\omega \in \Omega : X(\omega) = X'(\omega)\}$ ha probabilità uno. Notiamo che, se C è un generico evento di probabilità uno, vale che

$$P(B) = P(B \cap C) \quad \text{per ogni evento } B \subseteq \Omega,$$

come si vede scrivendo $P(B) = P(B \cap C) + P(B \cap C^c)$ e notando che $P(B \cap C^c) = 0$, perché $B \cap C^c \subseteq C^c$ e $P(C^c) = 1 - P(C) = 0$. In particolare, scegliendo B della forma $\{X \in A\}$ e $\{X' \in A\}$, con $A \subseteq E$ generico (si ricordi la Definizione 3.5), si ottiene

$$P(X \in A) = P(\{X \in A\} \cap C), \quad P(X' \in A) = P(\{X' \in A\} \cap C). \quad (3.11)$$

Ma $\{X \in A\} \cap C$ e $\{X' \in A\} \cap C$ sono lo stesso evento, dal momento che

$$\begin{aligned} \{X \in A\} \cap C &= \{\omega \in \Omega : X(\omega) \in A \text{ e } X(\omega) = X'(\omega)\} \\ &= \{\omega \in \Omega : X'(\omega) \in A \text{ e } X(\omega) = X'(\omega)\} = \{X' \in A\} \cap C. \end{aligned}$$

Pertanto, ricordando (3.11) e la definizione (3.5) di distribuzione,

$$\mu_X(A) = P(X \in A) = P(X' \in A) = \mu_{X'}(A), \quad \forall A \subseteq E,$$

ossia $\mu_X = \mu_{X'}$, come volevasi dimostrare. \square

Mostriamo infine che la distribuzione si conserva per composizione. Date due funzioni $X : \Omega \rightarrow E$ e $f : E \rightarrow F$, indichiamo la funzione composta $\omega \mapsto f(X(\omega))$ da Ω in F con il simbolo $f(X)$, invece di $f \circ X$.

Proposizione 3.15 (Conservazione della distribuzione). Siano X e X' due variabili aleatorie a valori nello stesso insieme E e con la stessa distribuzione: $\mu_X = \mu_{X'}$. Sia F un insieme generico e sia $f : E \rightarrow F$ una funzione qualunque. Allora le variabili aleatorie $f(X)$ e $f(X')$ hanno la stessa distribuzione (e dunque la stessa densità discreta).

Dimostrazione. Dato che si richiede solo che le variabili aleatorie X e X' abbiano la stessa distribuzione, esse possono essere definite su spazi di probabilità discreti (Ω, \mathcal{P}) e (Ω', \mathcal{P}')

non necessariamente uguali. Le variabili aleatorie $f(X)$ e $f(X')$ sono allora definite rispettivamente su Ω e Ω' , entrambe a valori in F . Notiamo che, per ogni sottoinsieme $A \subseteq F$, si ha l'uguaglianza di eventi

$$\{f(X) \in A\} = \{X \in f^{-1}(A)\},$$

dove $f^{-1}(A) \subseteq E$ indica la controimmagine del sottoinsieme A attraverso f . Infatti

$$\begin{aligned} \omega \in \{f(X) \in A\} &\iff f(X(\omega)) \in A \iff X(\omega) \in f^{-1}(A) \\ &\iff \omega \in \{X \in f^{-1}(A)\}. \end{aligned}$$

Pertanto, ricordando la definizione (3.5) di distribuzione, per ogni $A \subseteq F$ si ha

$$\begin{aligned} \mu_{f(X)}(A) &= P(f(X) \in A) = P(X \in f^{-1}(A)) = \mu_X(f^{-1}(A)) = \mu_{X'}(f^{-1}(A)) \\ &= P'(X' \in f^{-1}(A)) = P'(f(X') \in A) = \mu_{f(X')}(A), \end{aligned}$$

e dunque $\mu_{f(X)} = \mu_{f(X')}$. □

3.1.5 Osservazioni ed esempi

Studiare la distribuzione di una variabile aleatoria è un problema di grande rilevanza. Nella pratica, risulta spesso conveniente lavorare con la densità discreta, che caratterizza la distribuzione ed è un oggetto più semplice e intuitivo.[†]

Se $X : \Omega \rightarrow E$ è una variabile aleatoria, indichiamone con $X(\Omega) \subseteq E$ l'immagine, definita in (3.9). Ricordiamo che $X(\Omega)$ è sempre un insieme finito o numerabile e osserviamo che la densità discreta è nulla al di fuori di $X(\Omega)$:

$$p_X(x) = P(X = x) = 0 \quad \text{per ogni } x \notin X(\Omega), \quad (3.12)$$

perché $\{X = x\} = X^{-1}(\{x\}) = \emptyset$ se $x \notin X(\Omega)$. Pertanto, per determinare la densità discreta p_X , tipicamente si comincia identificando l'insieme finito o numerabile $X(\Omega) \subseteq E$, dopodiché si calcola $p_X(x)$ per ogni $x \in X(\Omega)$.

Osservazione 3.16. La relazione (3.12) mostra che se $p_X(x) > 0$, allora $x \in X(\Omega)$. Il viceversa non è necessariamente vero, perché ci possono essere valori $x \in E$ assunti da X , ma con probabilità nulla. In altre parole, la densità discreta p_X identifica l'insieme $\{x \in E : p_X(x) > 0\}$ dei valori assunti da X con probabilità positiva, mentre l'immagine $X(\Omega) \subseteq E$ comprende anche gli eventuali valori assunti da X con probabilità nulla. Tuttavia, dato che tali valori non giocano alcun ruolo in un problema probabilistico, si identifica spesso $X(\Omega)$ con l'insieme $\{x \in E : p_X(x) > 0\}$, con un abuso di notazione. □

Vediamo ora alcuni esempi concreti di determinazione di densità discrete.

Esempio 3.17 (Variabili aleatorie (quasi certamente) costanti). Come abbiamo visto, l'esempio più semplice di variabile aleatoria è dato da una costante: $X(\omega) = c$

[†] Formuleremo tuttavia molti risultati teorici in termini della distribuzione, perché essa, a differenza della densità discreta, può essere definita per variabili aleatorie generali, come vedremo.

per ogni $\omega \in \Omega$, dove $c \in E$ è un elemento fissato dello spazio di arrivo. In questo caso, chiaramente, $X(\Omega) = \{c\}$ e $p_X(c) = P(X = c) = 1$. Ricordando la definizione (0.1) della funzione indicatrice, la densità discreta di X è data dunque da

$$p_X(x) = \mathbb{1}_{\{c\}}(x), \quad \forall x \in E. \quad (3.13)$$

Viceversa, se X è una variabile aleatoria con densità discreta $p_X(\cdot)$ data da (3.13), allora $P(X = c) = 1$, dunque X è quasi certamente uguale alla costante c (si ricordi la Definizione 3.13). Una tale variabile aleatoria X si dice *quasi certamente costante*.

In definitiva, per la Proposizione 3.14, una variabile aleatoria X è quasi certamente costante se e solo se la sua densità discreta è della forma (3.13), con $c \in E$. \square

Esempio 3.18 (Variabili aleatorie indicatrici). Sia $A \subseteq \Omega$ un evento di uno spazio di probabilità discreto (Ω, P) . Consideriamo la variabile aleatoria (reale) $X := \mathbb{1}_A$ indicatrice dell'evento A . Chiaramente $X(\Omega) = \{0, 1\}$ e inoltre $P(X = 1) = P(A)$, dunque $P(X = 0) = 1 - P(A)$. La densità discreta di X è data dunque da

$$p_X(x) = P(A) \mathbb{1}_{\{1\}}(x) + (1 - P(A)) \mathbb{1}_{\{0\}}(x), \quad \forall x \in \mathbb{R}.$$

Ritorniamo su queste variabili aleatorie, dette *di Bernoulli*, nel Paragrafo 3.5.2. \square

Esempio 3.19 (Prove ripetute e indipendenti). Riprendiamo l'Esempio 3.4. Date n prove ripetute e indipendenti con probabilità di successo p , abbiamo definito in (3.1)

$$\begin{aligned} S &:= \text{“numero di successi nelle } n \text{ prove”}, \\ T &:= \text{“prova in cui si ha il primo successo”}, \end{aligned}$$

che sono variabili aleatorie a valori in $E = \mathbb{N}_0 \cup \{+\infty\}$ (con la convenzione $T = +\infty$ se nessun successo si verifica nelle n prove). È chiaro che

$$S(\Omega) = \{0, 1, \dots, n\}, \quad T(\Omega) = \{1, 2, \dots, n\} \cup \{+\infty\},$$

e restano da determinare le densità discrete, che in realtà abbiamo già calcolato. Infatti, ricordando le relazioni (3.2) e (1.56), (1.52), si ha

$$\begin{aligned} p_S(k) &= \binom{n}{k} p^k (1-p)^{n-k}, \quad \forall k \in S(\Omega) = \{0, 1, \dots, n\}, \\ p_T(k) &= p(1-p)^{k-1} \quad \forall k \in \{1, \dots, n\}, \quad p_T(+\infty) = (1-p)^n, \end{aligned} \quad (3.14)$$

dove per l'ultima relazione basta notare che $\{T = +\infty\} = \{S = 0\}$.

La distribuzione di S è detta *binomiale* e ne studieremo in dettaglio le proprietà nel Paragrafo 3.5.3 (in particolare, la Figura 3.2 a pagina 168 mostra il grafico della densità discreta di S , per alcuni valori di n e p). Come vedremo nel Paragrafo 3.5.6, un'estensione della distribuzione di T è detta *geometrica*. \square

Esempio 3.20 (Size bias). 120 studenti sono suddivisi in 3 gruppi, detti A, B e C, di 36, 40 e 44 studenti rispettivamente. Consideriamo i seguenti esperimenti aleatori, in cui “a caso” significa “uniformemente tra tutte le possibilità”.

- (i) Viene scelto un gruppo a caso e si indica con X il numero di studenti nel gruppo scelto.
- (ii) Viene scelto uno studente a caso e si indica con Y il numero di studenti nel gruppo dello studente scelto.

Determiniamo le distribuzioni di X e Y , o equivalentemente le loro densità discrete.

Notiamo innanzitutto che $X(\Omega) = Y(\Omega) = \{36, 40, 44\}$. Dato che la variabile aleatoria X è determinata scegliendo un gruppo a caso tra i 3 possibili, dovrebbe essere chiaro che

$$p_X(36) = p_X(40) = p_X(44) = \frac{1}{3}.$$

Infatti $p_X(36) = P(X = 36)$ e l'evento $\{X = 36\}$ coincide con l'evento "viene scelto il gruppo A", che ha per ipotesi probabilità $\frac{1}{3}$; con analoghi ragionamenti si determinano $p_X(40)$ e $p_X(44)$.

La densità discreta di Y è diversa. Infatti l'evento $\{Y = 36\}$ corrisponde all'evento "viene scelto uno studente del primo gruppo", e in questo caso lo studente (e non il gruppo) viene scelto uniformemente, tra i 120 possibili. Di conseguenza

$$p_Y(36) = \frac{36}{120} = \frac{3}{10}, \quad p_Y(40) = \frac{40}{120} = \frac{1}{3}, \quad p_Y(44) = \frac{44}{120} = \frac{11}{30}.$$

Si noti che abbiamo determinato le distribuzioni delle variabili aleatorie X e Y senza scrivere esplicitamente lo spazio di probabilità su cui esse sono definite. Volendo esplicitare lo spazio, per la variabile X si può scegliere ad esempio lo spazio campionario $\Omega = \{A, B, C\}$ munito della probabilità uniforme P , che rappresenta la scelta di un gruppo a caso, e definire $X : \Omega \rightarrow \mathbb{R}$ ponendo

$$X(A) := 36, \quad X(B) := 40, \quad X(C) := 44.$$

Per quanto riguarda la variabile Y , la scelta più naturale di spazio campionario è data da $\Omega = \{1, 2, \dots, 120\}$ munito della probabilità uniforme P , che rappresenta la scelta di uno studente a caso, e definire $Y : \Omega \rightarrow \mathbb{R}$ ponendo

$$Y(\omega) := \begin{cases} 36 & \text{se } \omega \in \{1, \dots, 36\} \\ 40 & \text{se } \omega \in \{37, \dots, 77\} \\ 44 & \text{se } \omega \in \{78, \dots, 120\} \end{cases}.$$

È una facile verifica che le variabili X e Y così definite (su spazi di probabilità diversi!) hanno effettivamente le distribuzioni sopra determinate. \square

Esempio 3.21 (Strategia del raddoppio). La roulette è composta dai numeri 0, 1, 2, ..., 37, di cui 18 sono "rossi", 18 sono "neri" e uno (lo zero) è "verde". Assumendo che la pallina si fermi su un numero scelto uniformemente, la probabilità che in una giocata esca il rosso (ossia un numero "rosso") vale dunque $\frac{18}{37}$. Puntando una somma di denaro sull'uscita del rosso, si riceve il doppio della somma se esce il rosso, mentre si perde la somma giocata se non esce il rosso.

Si consideri la seguente classica strategia di gioco. Un giocatore possiede un capitale iniziale pari a 1023 euro. Alla prima giocata punta un euro sul rosso: se esce

il rosso, il giocatore riceve due euro, si ritira e il gioco finisce. Se invece alla prima giocata non esce il rosso, il giocatore punta due euro sul rosso alla seconda giocata: se esce il rosso, il giocatore riceve quattro euro, si ritira e il gioco finisce. Se invece anche alla seconda giocata non esce il rosso, il giocatore continua raddoppiando la posta alla giocata successiva, e così via. Si osservi che, se il rosso non esce in nessuna delle prime 10 giocate, il capitale puntato dal giocatore, pari a $1 + 2 + 4 + \dots + 2^9 = 2^{10} - 1 = 1023$ euro (si ricordi la serie geometrica (0.3) per $x = 2$ e $N = 9$) si è esaurito e il giocatore si ritira. Indichiamo con X la differenza tra il capitale del giocatore alla fine e all'inizio del gioco, ossia il guadagno (con segno!). Determiniamo la distribuzione della variabile aleatoria X .

Sebbene a priori il problema possa sembrare complicato, la soluzione è in realtà abbastanza semplice. Notiamo innanzitutto che l'esperimento aleatorio in questione è costituito da 10 giocate di roulette, che costituiscono *10 prove ripetute e indipendenti* con probabilità di successo $p = \frac{18}{37}$, dove con "successo" intendiamo "esce il rosso" (si ricordi il Paragrafo 1.3.4). Infatti la variabile aleatoria X è *una funzione dell'esito delle prime 10 giocate*, dal momento che conoscendo l'esito di tali giocate il valore di X è univocamente determinato. Possiamo pertanto considerare lo spazio campionario $\Omega = \{0, 1\}^{10}$ munito della probabilità $P = P_{10,p}$ definita in (1.62), ma in realtà non ci sarà bisogno di scendere in dettagli eccessivi.

Per ottenere la distribuzione di X , è sufficiente determinarne i valori assunti, ossia l'insieme $X(\Omega)$, e la probabilità con cui tali valori sono assunti, ossia la densità discreta $p_X(x)$ per ogni $x \in X(\Omega)$.

- Se esce rosso alla prima giocata, il giocatore ha puntato 1 euro e ne riceve 2, pertanto il suo guadagno vale $X = 2 - 1 = 1$ euro;
- se il rosso non esce alla prima giocata, ma esce alla seconda, il giocatore ha puntato complessivamente $1 + 2 = 3$ euro e ne riceve $2 \cdot 2 = 4$, pertanto il suo guadagno vale $X = 4 - 3 = 1$ euro;
- passando al caso generale, se il rosso non esce nelle prime $n - 1$ giocate, ma esce all' n -esima (con $n = 1, 2, \dots, 10$), il giocatore ha puntato complessivamente $\sum_{i=1}^n 2^{i-1} = 1 + 2 + \dots + 2^{n-1} = 2^n - 1$ euro e ne riceve $2 \cdot 2^{n-1} = 2^n$; il suo guadagno vale dunque $X = 2^n - (2^n - 1) = 1$ euro;
- infine, se non esce mai rosso (nelle 10 giocate), il giocatore non riceve niente e dunque $X = 0 - (1 + 2 + 4 + \dots + 2^9) = -(2^{10} - 1) = -1023$ euro.

In definitiva, se esce rosso almeno una volta nelle prime 10 giocate, il guadagno del giocatore vale 1, mentre se non esce mai rosso il guadagno vale -1023 . I valori assunti dalla variabile aleatoria X sono dunque $X(\Omega) = \{-1023, 1\}$.

Resta solo da determinare $p_X(-1023)$, da cui si ricava $p_X(1) = 1 - p_X(-1023)$ (perché?). Si noti che l'evento $\{X = -1023\}$ coincide con l'evento "nelle prime 10 giocate non esce mai il rosso". Ricordando la formula (1.56), o più semplicemente notando che ad ogni giocata la probabilità che non esca il rosso vale $1 - \frac{18}{37} = \frac{19}{37}$ e

che le giocate sono indipendenti, si ottiene dunque

$$p_X(-1023) = \left(\frac{19}{37}\right)^{10} \simeq 0.0013, \quad p_X(1) = 1 - \left(\frac{19}{37}\right)^{10} \simeq 0.9987. \quad \square$$

Osservazione 3.22. L'esempio precedente mostra che, se si dispone di un capitale di 1023 euro, *esiste una strategia di gioco alla roulette che permette di vincere con probabilità $\simeq 99.87\%$* . Prima di lasciarsi andare a facili entusiasmi, conviene dare uno sguardo più attento alla strategia. Infatti:

- la probabilità di vincere è elevatissima, ma si vince solo 1 euro;
- la probabilità di perdere è bassissima, ma si perdono ben 1023 euro!

In definitiva, è una buona idea giocare usando questa strategia? La domanda è vaga e non esiste una “risposta corretta”, ma si possono fare alcune considerazioni.

Se ci si chiede se sia una buona idea adottare la strategia *una sola volta*, occorre soppesare la soddisfazione che si ha dalla vincita di 1 euro (che è un evento di probabilità $\simeq 99.87\%$) con l'insoddisfazione provocata dalla perdita di 1023 euro (che è un evento di probabilità $\simeq 0.13\%$). È verosimile che la maggior parte delle persone non accetterebbe il rischio di perdere 1023 euro a fronte di una vincita di solo 1 euro, anche se la probabilità di perdita fosse molto più bassa di 0.13%. Si tratta in ogni caso di considerazioni soggettive.

Se invece l'intenzione fosse di adottare la strategia *ripetutamente*, un numero molto elevato di volte (in modo da accumulare i probabili guadagni da 1 euro), ci sono elementi oggettivi per sconsigliare questa pratica: come discuteremo nell'Esempio 3.50, la probabilità di finire in perdita è in questo caso molto elevata. \square

Osservazione 3.23. A scanso di grossolani equivoci, sottolineiamo che *non si deve confondere, o addirittura identificare, una variabile aleatoria con la sua densità discreta*: si tratta di oggetti diversi! Se X è una funzione da Ω a valori in E , la sua densità discreta p_X è una funzione da E a valori in \mathbb{R} . Variabili aleatorie con la stessa densità possono essere definite in *diversi* spazi di probabilità, e/o avere relazioni diverse con altre variabili aleatorie (si vedano gli Esempi 3.25, 3.26 e 3.27). \square

3.1.6 Costruzione canonica di una variabile aleatoria

Abbiamo visto che a ogni variabile aleatoria X è associata la sua distribuzione μ_X , che è una probabilità discreta sullo spazio di arrivo di X , associata alla densità discreta p_X . È naturale chiedersi se valga il viceversa, cioè se ogni probabilità discreta è la distribuzione di una variabile aleatoria. Più precisamente: fissati arbitrariamente un insieme E e una probabilità discreta μ (o equivalentemente la sua densità discreta p) su E , esistono uno spazio di probabilità discreto (Ω, P) e una variabile aleatoria discreta $X : \Omega \rightarrow E$ che ha distribuzione $\mu_X = \mu$ (o equivalentemente $p_X = p$)?

La risposta è affermativa. Nel caso in cui E sia finito o numerabile, notiamo che (E, μ) è uno spazio di probabilità discreto, pertanto è sufficiente considerare $(\Omega, P) = (E, \mu)$ e definire X come l'identità, ossia $X(\omega) := \omega$ per ogni $\omega \in \Omega$. In questo modo

si ha per costruzione $\{X \in A\} = X^{-1}(A) = A$ per ogni $A \subseteq E$, quindi

$$\mu_X(A) := P(X \in A) = P(A) = \mu(A),$$

ossia X ha distribuzione $\mu_X = \mu$ come cercato. In altri termini, abbiamo scelto come spazio campionario lo spazio dei valori assunti dalla variabile aleatoria cercata, come probabilità la distribuzione assegnata e come variabile aleatoria la funzione identità. Questa procedura è nota come *costruzione canonica di una variabile aleatoria con distribuzione assegnata*.

Nel caso in cui E sia infinito più che numerabile, la procedura appena descritta non può essere applicata se vogliamo lavorare con spazi di probabilità discreti, in cui Ω è per definizione finito o numerabile. Tuttavia è sufficiente una piccola modifica. Infatti, ricordando l'Osservazione 3.9, per ogni probabilità discreta μ su E esiste un sottoinsieme $\tilde{E} \subseteq E$ finito o numerabile tale che $\mu(\tilde{E}) = 1$, e la restrizione di μ a \tilde{E} è una probabilità che soddisfa $\mu(A) = \mu(A \cap \tilde{E})$ per ogni $A \subseteq E$. Possiamo pertanto considerare $(\Omega, P) = (\tilde{E}, \mu)$ e definire ancora $X : \Omega \rightarrow E$ come l'identità – o, più precisamente, l'inclusione – ossia $X(\omega) := \omega$ per ogni $\omega \in \Omega$. Dato che $\{X \in A\} = X^{-1}(A) = A \cap \tilde{E}$ per ogni $A \subseteq E$, segue che

$$\mu_X(A) := P(X \in A) = \mu(A \cap \tilde{E}) = \mu(A),$$

ossia X ha distribuzione μ .

Esercizi

Esercizio 3.1. Sia X una variabile aleatoria, definita su (Ω, P) a valori in E , quasi certamente costante, ossia esiste $c \in E$ tale che $P(X = c) = 1$. Si mostri che esiste un *unico* elemento $c \in E$ con tale proprietà.

Esercizio 3.2. Siano A e B due eventi in uno spazio di probabilità discreto (Ω, P) , e consideriamo la variabile aleatoria

$$X := \lambda \mathbb{1}_A + \mu \mathbb{1}_B,$$

dove $\lambda, \mu \in \mathbb{R}$. Si determini la densità discreta di X .

[Sugg. Si consideri innanzitutto il caso in cui i quattro numeri $0, \lambda, \mu, \lambda + \mu$ sono distinti.]

Esercizio 3.3. Sia X una variabile aleatoria definita su uno spazio di probabilità discreto (Ω, P) a valori in un insieme E . Indichiamo con $\sigma(X)$ la famiglia degli eventi generati da X , secondo la Definizione 3.5:

$$\sigma(X) = \{C \subseteq \Omega : \exists A \subseteq E \text{ tale che } C = \{X \in A\}\}.$$

Si mostri che un evento $D \subseteq \Omega$ appartiene a $\sigma(X)$ se e solo se verifica la seguente proprietà: per ogni coppia di elementi $\omega, \omega' \in \Omega$ tali che $X(\omega) = X(\omega')$, o entrambi gli elementi appartengono a D o entrambi non vi appartengono, ossia non si può avere che $\omega \in D$ e $\omega' \notin D$, o viceversa che $\omega \notin D$ e $\omega' \in D$.

3.2 Indipendenza di variabili aleatorie

3.2.1 Distribuzioni congiunte e marginali

In molti casi si è interessati a due (o più) variabili aleatorie che sono funzioni dello stesso esperimento aleatorio. Ad esempio, lanciando due dadi, si può essere interessati sia al numero X del primo dado, sia alla somma Y dei numeri dei due dadi. In termini matematici, ciò significa che *sullo stesso spazio di probabilità* discreto (Ω, \mathcal{P}) sono definite due variabili aleatorie $X : \Omega \rightarrow E$ e $Y : \Omega \rightarrow F$ (a valori in spazi non necessariamente uguali). Si può allora considerare *la coppia* $Z := (X, Y)$, che è una variabile aleatoria definita su Ω a valori nello *spazio prodotto* $E \times F$.

Viceversa, se Z è una variabile aleatoria a valori in uno spazio prodotto $E \times F$, definita su uno spazio di probabilità discreto (Ω, \mathcal{P}) , per ogni $\omega \in \Omega$ si ha $Z(\omega) \in E \times F$ e dunque si può scrivere in modo univoco $Z(\omega) = (X(\omega), Y(\omega))$, con $X(\omega) \in E$ e $Y(\omega) \in F$. Sono quindi ben definite le variabili aleatorie $X : \Omega \rightarrow E$ e $Y : \Omega \rightarrow F$, dette *componenti* di Z .

Alle variabili aleatorie $X : \Omega \rightarrow E$ e $Y : \Omega \rightarrow F$ sono associate le rispettive distribuzioni μ_X e μ_Y , che sono probabilità discrete sugli insiemi E e F . Anche alla variabile aleatoria $Z = (X, Y) : \Omega \rightarrow E \times F$ è associata la distribuzione μ_Z , che indicheremo con $\mu_{X,Y}$, che è una probabilità discreta su $E \times F$. È naturale chiedersi quale relazione ci sia tra la distribuzione $\mu_{X,Y}$, detta *distribuzione congiunta* delle variabili aleatorie X e Y , e le distribuzioni μ_X e μ_Y , dette *distribuzioni marginali*.

Come è tipico, invece che con le distribuzioni lavoreremo con le corrispondenti densità discrete. Mostriamo che le densità discrete p_X e p_Y delle variabili aleatorie X e Y , dette *densità (discrete) marginali*, possono essere facilmente ricavate dalla densità discreta $p_{X,Y}$ della variabile aleatoria (X, Y) , che è detta *densità (discreta) congiunta*. Ricordando la relazione (3.6), per $x \in E$ e $y \in F$ si ha

$$\begin{aligned} p_X(x) &= P(X = x), & p_Y(y) &= P(Y = y), \\ p_{X,Y}(x, y) &= P((X, Y) = (x, y)) = P(X = x, Y = y), \end{aligned}$$

dove nell'ultima probabilità *la virgola indica l'intersezione di eventi*, ossia usiamo la notazione $\{X = x, Y = y\} := \{X = x\} \cap \{Y = y\}$ (lo faremo di frequente).

Proposizione 3.24 (Densità discrete congiunte e marginali). *Siano $X : \Omega \rightarrow E$ e $Y : \Omega \rightarrow F$ variabili aleatorie, definite sullo stesso spazio di probabilità discreto (Ω, \mathcal{P}) . Le densità marginali p_X, p_Y e la densità congiunta $p_{X,Y}$ soddisfano le seguenti relazioni:*

$$\begin{aligned} \forall x \in E : \quad p_X(x) &= \sum_{y \in F} p_{X,Y}(x, y); \\ \forall y \in F : \quad p_Y(y) &= \sum_{x \in E} p_{X,Y}(x, y). \end{aligned} \tag{3.15}$$

Dimostrazione. Si osservi che lo spazio Ω può essere espresso come unione al più numerabile di eventi disgiunti come segue:

$$\Omega = \bigcup_{y \in Y(\Omega)} \{Y = y\}. \quad (3.16)$$

Infatti, per definizione di funzione, per ogni $\omega \in \Omega$ esiste ed è unico l'elemento $y \in Y(\Omega)$ tale che $Y(\omega) = y$, il che è equivalente a $\omega \in \{Y = y\}$. Dal momento che $\{X = x\} = \{X = x\} \cap \Omega$, si ottiene

$$\{X = x\} = \bigcup_{y \in Y(\Omega)} \{X = x, Y = y\}, \quad \forall x \in E,$$

dove l'unione è disgiunta e al più numerabile. Pertanto, per la σ -additività,

$$\begin{aligned} p_X(x) &= P(X = x) = \sum_{y \in Y(\Omega)} P(X = x, Y = y) = \sum_{y \in Y(\Omega)} P((X, Y) = (x, y)) \\ &= \sum_{y \in Y(\Omega)} p_{X,Y}(x, y) = \sum_{y \in F} p_{X,Y}(x, y), \end{aligned}$$

dove, per l'ultima uguaglianza, abbiamo usato il fatto che se $y \notin Y(\Omega)$, allora l'evento $\{(X, Y) = (x, y)\} \subseteq \{Y = y\} = \emptyset$ è vuoto e ha dunque probabilità zero. \square

Il risultato precedente si estende senza difficoltà concettuali, solo con notazioni più pesanti, al caso di più di due variabili aleatorie. Più precisamente, siano X_1, \dots, X_n variabili aleatorie definite sullo stesso spazio di probabilità discreto (Ω, P) a valori rispettivamente negli insiemi E_1, \dots, E_n . La n -upla (X_1, \dots, X_n) è allora una variabile aleatoria definita su (Ω, P) a valori nello spazio prodotto $E_1 \times \dots \times E_n$. La densità congiunta p_{X_1, \dots, X_n} – ossia la densità discreta della variabile aleatoria (X_1, \dots, X_n) – è legata alle densità marginali p_{X_i} – ossia le densità discrete delle singole variabili aleatorie X_i – dalla seguente relazione:

$$p_{X_i}(x_i) = \sum_{j \neq i} \sum_{x_j \in E_j} p_{X_1, \dots, X_n}(x_1, \dots, x_n), \quad \forall x_i \in E_i,$$

in completa analogia con la relazione (3.15).

Più in generale, per ogni sottoinsieme di indici $\{i_1, \dots, i_m\} \subseteq \{1, 2, \dots, n\}$, in cui conveniamo che $i_1 < i_2 < \dots < i_m$, è possibile considerare la variabile aleatoria $(X_{i_1}, X_{i_2}, \dots, X_{i_m})$ definita su (Ω, P) , a valori nello spazio prodotto $E_{i_1} \times \dots \times E_{i_m}$. Vale allora la seguente relazione:

$$p_{X_{i_1}, X_{i_2}, \dots, X_{i_m}}(x_{i_1}, x_{i_2}, \dots, x_{i_m}) = \sum_{j \notin I} \sum_{x_j \in E_j} p_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n),$$

ossia la densità congiunta di un sottoinsieme di componenti può essere ricavata dalla densità congiunta di tutte le componenti, sommando sulle variabili relative alle “altre” componenti.

Esempio 3.25. Lanciamo due dadi regolari a sei facce e indichiamo con X il numero del primo dado e con Y la somma dei numeri dei due dadi. Determiniamo la densità congiunta $p_{X,Y}$ e le densità marginali p_X e p_Y .

Come capita spesso, non è necessario scrivere in dettaglio uno spazio di probabilità discreto (Ω, P) su cui definire X e Y , ma è sufficiente ragionare sugli eventi di interesse. Per calcolare la densità congiunta $p_{X,Y}(x, y)$, notiamo che $\{X = x, Y = y\}$

è l'evento “il numero del primo dado è x e la somma dei due numeri è y ”, o equivalentemente “il numero del primo dado è x e il numero del secondo dado è $y - x$ ”. È pertanto necessario che tanto x quanto $y - x$ siano in $\{1, \dots, 6\}$ e, in tal caso, si ha $P(X = x, Y = y) = \frac{1}{6} \cdot \frac{1}{6} = \frac{1}{36}$ perché gli esiti dei due dadi sono indipendenti. Di conseguenza

$$\begin{aligned} p_{X,Y}(x,y) &= \frac{1}{36} \mathbb{1}_{\{1,\dots,6\}}(x) \mathbb{1}_{\{1,\dots,6\}}(y-x) \\ &= \begin{cases} \frac{1}{36} & \text{se } x \in \{1, \dots, 6\} \text{ e } y \in \{x+1, \dots, x+6\} \\ 0 & \text{altrimenti} \end{cases}. \end{aligned}$$

Le densità marginali p_X e p_Y possono essere ricavate considerando singolarmente X e Y , oppure ottenute dalla densità congiunta $p_{X,Y}$ attraverso la formula (3.15). Procediamo nel secondo modo. Notiamo innanzitutto che $X(\Omega) = \{1, \dots, 6\}$ e $Y(\Omega) = \{2, 3, \dots, 12\}$. Per ogni $x \in X(\Omega) = \{1, \dots, 6\}$ esistono esattamente 6 valori di $y \in \mathbb{N}$ tali che $y - x \in \{1, \dots, 6\}$, precisamente $\{x+1, \dots, x+6\}$, pertanto

$$p_X(x) = \sum_{y \in \mathbb{N}} p_{X,Y}(x,y) = \frac{1}{36} \mathbb{1}_{\{1,\dots,6\}}(x) \sum_{y \in \mathbb{N}} \mathbb{1}_{\{1,\dots,6\}}(y-x) = \frac{1}{6} \mathbb{1}_{\{1,\dots,6\}}(x).$$

Questo è naturalmente lo stesso risultato dell'Esempio 3.12, dato che X rappresenta il numero ottenuto lanciando un dado regolare a sei facce.

Per quanto riguarda p_Y , dobbiamo determinare, per ogni $y \in Y(\Omega) = \{2, \dots, 12\}$ fissato, quanti valori di $x \in \mathbb{N}$ ci sono per cui $x \in \{1, \dots, 6\}$ e allo stesso tempo $y - x \in \{1, \dots, 6\}$, ossia $x \in \{1, \dots, 6\} \cap \{y - 6, \dots, y - 1\}$. Per enumerazione diretta, ci si rende facilmente conto che

$$\{1, \dots, 6\} \cap \{y - 6, \dots, y - 1\} = \begin{cases} \{1, \dots, y - 1\} & \text{se } y \in \{2, \dots, 7\} \\ \{y - 6, \dots, 6\} & \text{se } y \in \{8, \dots, 12\} \end{cases},$$

pertanto

$$\begin{aligned} p_Y(y) &= \sum_{x \in \mathbb{N}} p_{X,Y}(x,y) = \frac{|\{1, \dots, 6\} \cap \{y - 6, \dots, y - 1\}|}{36} \\ &= \begin{cases} \frac{y-1}{36} & \text{se } y \in \{2, \dots, 7\} \\ \frac{13-y}{36} & \text{se } y \in \{8, \dots, 12\} \end{cases}. \quad \square \end{aligned}$$

La Proposizione 3.24 mostra che le distribuzioni marginali di due variabili aleatorie possono essere determinate a partire dalla distribuzione congiunta. Ci si può chiedere se valga il viceversa, ossia se la distribuzione congiunta sia univocamente determinata dalle distribuzioni marginali. La risposta è *negativa*, come mostrano i due seguenti esempi, in cui costruiamo due coppie di variabili aleatorie con le stesse distribuzioni marginali ma con diversa distribuzione congiunta.

Esempio 3.26 (Estrazioni con reimmissione). Un'urna contiene n palline, numerate da 1 a n . Estraiamo due palline *con* reimmissione e indichiamo con X_1 e X_2 le palline estratte. Dato che ogni pallina ha la stessa probabilità $\frac{1}{n}$ di essere estratta, sia alla

prima che alla seconda estrazione, le densità marginali di X_1 e X_2 sono date da

$$p_{X_1}(x) = p_{X_2}(x) = \frac{1}{n} \mathbb{1}_{\{1, \dots, n\}}(x). \quad (3.17)$$

Si ha inoltre $P(X_2 = x_2 | X_1 = x_1) = \frac{1}{n}$ per ogni $x_1, x_2 \in \{1, \dots, n\}$. Infatti, sapere di aver pescato la pallina x_1 alla prima estrazione non influenza la seconda estrazione, perché la pallina viene reimmessa nell'urna. Pertanto la densità congiunta vale

$$p_{X_1, X_2}(x_1, x_2) = P(X_1 = x_1) P(X_2 = x_2 | X_1 = x_1) = \frac{1}{n^2} \mathbb{1}_{\{1, \dots, n\}}(x_1) \mathbb{1}_{\{1, \dots, n\}}(x_2),$$

da cui, volendo, si ritrovano le densità marginali in (3.17).

(Si noti che la distribuzione di X_1 e X_2 non è altro che la probabilità uniforme sull'insieme $\{1, \dots, n\}$, definita nell'Esempio 1.13. Variabili aleatorie con tale distribuzione sono dette *uniformi discrete*: ci ritorneremo nel Paragrafo 3.5.1.) \square

Esempio 3.27 (Estrazioni senza reimmisione). Un'urna contiene n palline, numerate da 1 a n . Estraiamo due palline *senza* reimmisione e indichiamo con X_1 e X_2 i numeri delle palline estratte. Chiaramente la densità di X_1 è la stessa (3.17) dell'esempio precedente.

Per la densità congiunta, notiamo che $P(X_2 = x_2 | X_1 = x_1) = 0$ se $x_1 = x_2$, visto che la prima pallina estratta *non* viene reimmessa nell'urna. Se invece $x_1 \neq x_2$, dopo la prima estrazione restano $n - 1$ palline nell'urna e dunque $P(X_2 = x_2 | X_1 = x_1) = \frac{1}{n-1}$. Dato che $p_{X_1, X_2}(x_1, x_2) = P(X_1 = x_1) P(X_2 = x_2 | X_1 = x_1)$, si ottiene

$$p_{X_1, X_2}(x_1, x_2) = \begin{cases} 0 & \text{se } x_1 = x_2 \\ \frac{1}{n-1} & \text{se } x_1 \neq x_2, \text{ con } x_1, x_2 \in \{1, \dots, n\} \end{cases}.$$

Applicando la relazione (3.15), si ricava facilmente le densità marginale di X_2 :

$$p_{X_2}(x) = \frac{1}{n} \mathbb{1}_{\{1, \dots, n\}}(x) = p_{X_1}(x).$$

Quindi anche X_2 ha distribuzione uniforme in $\{1, \dots, n\}$, un fatto non evidente a priori. Sottolineiamo che le distribuzioni marginali di X_1 e X_2 sono le stesse dell'esempio precedente, mentre la distribuzione congiunta è diversa. \square

Intuitivamente, la distribuzione congiunta descrive il comportamento collettivo delle variabili aleatorie in questione e *contiene dunque più informazione* delle distribuzioni marginali, che ne descrivono solo i comportamenti individuali.

3.2.2 Indipendenza di variabili aleatorie

Nel Paragrafo 1.3.3 abbiamo definito l'indipendenza di eventi di uno spazio di probabilità discreto (Ω, P) . Estendiamo ora la nozione di indipendenza alle variabili aleatorie, cominciando dal caso di una famiglia finita.

Definizione 3.28 (Indipendenza di n variabili aleatorie). Siano X_1, X_2, \dots, X_n variabili aleatorie, definite nello stesso spazio di probabilità, a valori rispettivamente negli insiemi E_1, E_2, \dots, E_n . Esse si dicono *indipendenti* se per ogni scelta dei sottoinsiemi $A_1 \subseteq E_1, A_2 \subseteq E_2, \dots, A_n \subseteq E_n$ si ha

$$P(X_1 \in A_1, X_2 \in A_2, \dots, X_n \in A_n) = \prod_{i=1}^n P(X_i \in A_i). \quad (3.18)$$

Nel caso speciale di due sole variabili aleatorie $X : \Omega \rightarrow E$ e $Y : \Omega \rightarrow F$, la relazione (3.18) che caratterizza l'indipendenza di X e Y si scrive come

$$P(X \in A, Y \in B) = P(X \in A) P(Y \in B), \quad \forall A \subseteq E, \forall B \subseteq F. \quad (3.19)$$

La nozione di indipendenza di due variabili aleatorie formalizza l'idea intuitiva che “*avere informazioni sul valore di una variabile non modifica le previsioni che si possono fare sul valore dell'altra*”. Infatti, la relazione (3.19) mostra che gli eventi $\{X \in A\}$ e $\{Y \in B\}$ sono indipendenti, per ogni scelta dei sottoinsiemi A e B negli spazi di arrivo di X e Y . Di conseguenza, se $P(X \in A) \neq 0$, possiamo scrivere

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

ossia l'informazione che X ha assunto valori in A non modifica la probabilità che Y assuma valori in B , per qualunque scelta di A e B .

Analogamente, l'indipendenza delle variabili aleatorie X_1, X_2, \dots, X_n significa intuitivamente che “*avere informazioni sui valori di alcune variabili della famiglia non modifica le previsioni che si possono fare sui valori delle altre variabili*”.

Osservazione 3.29. Ricordando la Definizione 1.62 di indipendenza di una famiglia di eventi, può sembrare strano che la relazione (3.18) non venga imposta anche per ogni sottofamiglia di indici $\{i_1, i_2, \dots, i_k\} \subset \{1, 2, \dots, n\}$. La ragione è che ciò non è necessario: infatti, scegliendo $A_j = E_j$ per $j \notin \{i_1, i_2, \dots, i_k\}$, si ha che $\{X_j \in A_j\} = \{X_j \in E_j\} = \Omega$ per tali valori di j , pertanto si ottiene da (3.18)

$$\begin{aligned} P(X_{i_1} \in A_{i_1}, X_{i_2} \in A_{i_2}, \dots, X_{i_k} \in A_{i_k}) &= P(X_1 \in A_1, X_2 \in A_2, \dots, X_n \in A_n) \\ &= \prod_{i=1}^n P(X_i \in A_i) = \prod_{j=1}^k P(X_{i_j} \in A_{i_j}). \end{aligned}$$

In particolare, se X_1, X_2, \dots, X_n sono variabili aleatorie indipendenti, anche $X_{i_1}, X_{i_2}, \dots, X_{i_k}$ lo sono, per ogni scelta di $\{i_1, i_2, \dots, i_k\} \subseteq \{1, 2, \dots, n\}$. \square

La nozione di indipendenza di una famiglia arbitraria (non necessariamente finita) di variabili aleatorie è data nella seguente definizione che, grazie all'osservazione appena fatta, è consistente con la Definizione 3.28.

Definizione 3.30 (Indipendenza di variabili aleatorie). Siano $(X_i)_{i \in I}$ variabili aleatorie definite nello stesso spazio di probabilità, dove l'insieme di indici I è arbitrario. Esse si dicono indipendenti se ogni sottofamiglia finita $(X_j)_{j \in J}$, con $J \subseteq I$ e $|J| < \infty$, è formata da variabili aleatorie indipendenti.

Confrontando questa definizione con la Definizione 1.62 di indipendenza tra eventi, si ottiene facilmente (esercizio) la seguente proposizione.

Proposizione 3.31. Siano $(X_i)_{i \in I}$ variabili aleatorie definite nello stesso spazio di probabilità, a valori rispettivamente negli insiemi $(E_i)_{i \in I}$. Le seguenti affermazioni sono equivalenti:

- (i) le variabili aleatorie $(X_i)_{i \in I}$ sono indipendenti;
- (ii) per ogni scelta di $A_i \subseteq E_i$, gli eventi $(\{X_i \in A_i\})_{i \in I}$ sono indipendenti.

Mostriamo ora come l'indipendenza di variabili aleatorie si possa caratterizzare in termini della loro densità congiunta. Per semplicità di notazione, enunciamo e dimostriamo la proposizione nel caso di due variabili aleatorie, ma il risultato si estende facilmente al caso di n variabili aleatorie, come discutiamo nel seguito.

Proposizione 3.32 (Indipendenza e densità discreta). Siano $X : \Omega \rightarrow E$ e $Y : \Omega \rightarrow F$ variabili aleatorie definite sullo stesso spazio di probabilità, di cui indichiamo le densità marginali con p_X, p_Y e la densità congiunta con $p_{X,Y}$. Esse sono indipendenti se e solo se vale la seguente relazione:

$$p_{X,Y}(x,y) = p_X(x) p_Y(y), \quad \forall x \in E, y \in F. \quad (3.20)$$

Dimostrazione. Siano X e Y indipendenti. Per ogni $x \in E$ e $y \in F$, applicando la relazione (3.19) con $A = \{x\}$ e $B = \{y\}$ si ottiene

$$p_{X,Y}(x,y) = P(X=x, Y=y) = P(X=x) P(Y=y) = p_X(x) p_Y(y).$$

Viceversa, assumiamo che valga la relazione (3.20). Si ha allora

$$\begin{aligned} P(X \in A, Y \in B) &= P((X,Y) \in A \times B) = \sum_{x \in A, y \in B} p_{X,Y}(x,y) \\ &= \sum_{x \in A, y \in B} p_X(x) p_Y(y) = \left(\sum_{x \in A} p_X(x) \right) \left(\sum_{y \in B} p_Y(y) \right) \\ &= P(X \in A) P(Y \in B), \end{aligned}$$

dove nella seconda uguaglianza abbiamo usato il legame tra distribuzione e densità discreta per la variabile aleatoria (X,Y) (si ricordino le relazioni (3.5) e (3.8)) mentre nella quarta uguaglianza abbiamo usato la somma a blocchi (0.10). Abbiamo ottenuto la relazione (3.19), che caratterizza l'indipendenza di X e Y . \square

La Proposizione 3.32 si estende al caso di n variabili aleatorie X_1, X_2, \dots, X_n , definite nello stesso spazio di probabilità (Ω, \mathcal{P}) , a valori rispettivamente negli insiemi E_1, E_2, \dots, E_n . Indicando con p_{X_1, \dots, X_n} la loro densità congiunta e con p_{X_i} le densità marginali, esse sono indipendenti se e solo se

$$p_{X_1, \dots, X_n}(x_1, \dots, x_n) = \prod_{i=1}^n p_{X_i}(x_i), \quad \forall (x_1, \dots, x_n) \in E_1 \times \dots \times E_n. \quad (3.21)$$

Osservazione 3.33. La Proposizione 3.32 mostra che la densità congiunta può essere ricavata dalle densità marginali, se le variabili aleatorie sono indipendenti. *Sottolineiamo che questo non è possibile in generale:* esistono infatti coppie di variabili aleatorie con le stesse densità marginali ma con densità congiunte diverse. Ad esempio, estraendo due palline da un'urna che ne contiene n (numerate progressivamente da 1 a n) e indicando con X_1 e X_2 i numeri delle palline estratte:

- se l'estrazione avviene con reimmissione, X_1 e X_2 sono variabili aleatorie indipendenti: infatti, come abbiamo visto nell'Esempio 3.26, si ha

$$p_{X_1, X_2}(x_1, x_2) = p_{X_1}(x_1) p_{X_2}(x_2), \quad \forall x_1, x_2 \in \{1, \dots, n\};$$

- tale relazione invece non è verificata nell'Esempio 3.27, pertanto X_1 e X_2 non sono indipendenti se l'estrazione avviene senza reimmissione.

Tutto ciò si generalizza al caso in cui dall'urna vengano estratte più di due palline, come mostreremo negli Esempi 3.98 e 3.99, che sono accessibili già da ora. \square

Osservazione 3.34. Siano X, Y due variabili aleatorie, definite sullo stesso spazio di probabilità a valori negli insiemi E e F rispettivamente, la cui densità congiunta si fattorizza nella forma

$$p_{X,Y}(x, y) = \alpha(x) \beta(y), \quad \forall x \in E, y \in F,$$

dove $\alpha: E \rightarrow \mathbb{R}^+$ e $\beta: F \rightarrow \mathbb{R}^+$ sono arbitrarie funzioni positive. Mostriamo che X e Y sono indipendenti. Si ha

$$p_X(x) = \sum_{y \in F} p_{X,Y}(x, y) = B \alpha(x),$$

grazie alla Proposizione 3.24, dove $B := \sum_{y \in F} \beta(y) \in [0, +\infty]$. Analogamente

$$p_Y(y) = A \beta(y),$$

con $A = \sum_{x \in E} \alpha(x) \in [0, +\infty]$. Inoltre

$$1 = \sum_{x \in E, y \in F} p_{X,Y}(x, y) = \left(\sum_{x \in E} \alpha(x) \right) \left(\sum_{y \in F} \beta(y) \right) = AB,$$

quindi $A \in (0, \infty)$, $B \in (0, \infty)$ e $B = 1/A$. Di conseguenza

$$p_{X,Y}(x, y) = \alpha(x) \beta(y) = p_X(x) p_Y(y), \quad \forall x \in E, y \in F,$$

e dunque X e Y sono indipendenti per la Proposizione 3.32.

Questo argomento è facilmente generalizzabile a una famiglia finita: se X_1, \dots, X_n sono n variabili aleatorie, a valori rispettivamente in E_1, \dots, E_n , tali che

$$p_{X_1, \dots, X_n}(x_1, \dots, x_n) = \alpha_1(x_1) \cdots \alpha_n(x_n), \quad \forall x_1 \in E_1, \dots, x_n \in E_n,$$

per un'opportuna scelta delle funzioni positive $\alpha_i : E_i \rightarrow \mathbb{R}^+$, con $i = 1, \dots, n$, allora le variabili aleatorie X_1, \dots, X_n sono indipendenti. \square

3.2.3 Rivisitazione delle prove ripetute e indipendenti

Consideriamo n prove ripetute e indipendenti con probabilità di successo p , ossia n eventi indipendenti C_1, \dots, C_n con la stessa probabilità $p = P(C_i)$, in uno spazio di probabilità discreto (Ω, P) (si ricordi il Paragrafo 1.3.4). Indichiamo con X_1, \dots, X_n le corrispondenti variabili aleatorie indicatrici:

$$X_i(\omega) := \mathbb{1}_{C_i}(\omega) = \begin{cases} 1 & \text{se } \omega \in C_i \\ 0 & \text{se } \omega \notin C_i \end{cases}, \quad (3.22)$$

che intuitivamente descrivono gli esiti delle singole prove: $X_i = 1$ se l' i -esima prova ha avuto successo e $X_i = 0$ altrimenti. Si noti che gli eventi $C_i = \{X_i = 1\}$ possono essere ricostruiti a partire da tali variabili aleatorie e, di conseguenza, ogni quantità di interesse può essere espressa in funzione di X_1, \dots, X_n .[†]

L'indipendenza degli eventi C_1, \dots, C_n è equivalente all'indipendenza delle variabili aleatorie X_1, \dots, X_n (si veda l'Esercizio 3.5). Inoltre, per l'Esempio 3.18, il fatto che gli eventi abbiano probabilità $p = P(C_i)$ è equivalente al fatto che le variabili aleatorie X_i abbiano densità discreta

$$p_{X_i}(t) = q(t) := \begin{cases} p & \text{se } t = 1 \\ 1 - p & \text{se } t = 0 \\ 0 & \text{se } t \notin \{0, 1\} \end{cases}. \quad (3.23)$$

Quindi, per descrivere n prove ripetute e indipendenti con probabilità di successo p , d'ora in avanti ci riferiremo equivalentemente a una delle seguenti costruzioni:

- n eventi C_1, \dots, C_n indipendenti e con la stessa probabilità $p = P(C_i)$;
- n variabili aleatorie X_1, \dots, X_n indipendenti e con la stessa distribuzione, con densità discreta $q(\cdot)$ data in (3.23).

La distribuzione delle X_i , detta *Bernoulli di parametro p* , gioca un ruolo importante in probabilità e sarà analizzata nel Paragrafo 3.5.2.

Osservazione 3.35. Come per le prove ripetute e indipendenti, molti modelli probabilistici studiati nel Capitolo 1 possono essere riformulati efficacemente, e spesso più intuitivamente, usando variabili aleatorie. In particolare, i modelli di estrazione di palline da un'urna (con e senza reimmissione) e di inserimento di oggetti in cassetti,

[†] Ad esempio, le variabili aleatorie S e T degli Esempi 3.4 e 3.19, che descrivono rispettivamente il numero di prove che hanno avuto successo e la prima prova che ha avuto successo, sono date da $S = X_1 + \dots + X_n$ e $T = \min\{i \in \{1, \dots, n\} : X_i = 1\}$.

introdotti negli Esempi 1.38, 1.39 e 1.41, verranno riconsiderati negli Esempi 3.98 e 3.99, che sono accessibili già da ora. \square

Osservazione 3.36. Ricordando la relazione (3.21), che caratterizza l'indipendenza di variabili aleatorie, e le formule (1.60), (1.61) e (1.62), otteniamo la densità congiunta di X_1, \dots, X_n :

$$p_{X_1, \dots, X_n}(x_1, \dots, x_n) = \prod_{i=1}^n p_{X_i}(x_i) = \prod_{i=1}^n q(x_i) = P_{n,p}(\{(x_1, \dots, x_n)\}). \quad (3.24)$$

Ciò significa che la distribuzione congiunta del vettore aleatorio (X_1, \dots, X_n) , sullo spazio di arrivo $\{0, 1\}^n = \Omega_n$, non è altro che la probabilità “canonica” $P_{n,p}$ costruita nel Paragrafo 1.3.4. \square

3.2.4 Proprietà dell'indipendenza

Un'osservazione elementare ma importante è che l'indipendenza di n variabili aleatorie X_1, \dots, X_n è una proprietà della loro distribuzione congiunta, o equivalentemente della loro densità discreta congiunta.

Proposizione 3.37. *Siano X_1, \dots, X_n variabili aleatorie discrete indipendenti. Se le variabili aleatorie discrete X'_1, \dots, X'_n hanno la stessa distribuzione congiunta (equivalentemente, la stessa densità congiunta) di X_1, \dots, X_n , anche esse sono variabili aleatorie indipendenti.*

Dimostrazione. Per ipotesi, le variabili aleatorie X_1, \dots, X_n e X'_1, \dots, X'_n hanno la stessa densità discreta congiunta e dunque le stesse densità discrete marginali (perché?). Di conseguenza la relazione (3.21), che caratterizza l'indipendenza e vale per ipotesi per X_1, \dots, X_n , vale anche per X'_1, \dots, X'_n . \square

Osservazione 3.38. Una dimostrazione alternativa più “elegante”, che non usa esplicitamente la densità discreta, consiste nel mostrare che la relazione (3.18), nella Definizione 3.28 di indipendenza di variabili aleatorie, può essere espressa in funzione della *distribuzione congiunta* delle variabili aleatorie X_1, \dots, X_n (esercizio). \square

Mostriamo ora una proprietà, piuttosto intuitiva, secondo cui, se si divide una famiglia di variabili aleatorie indipendenti in sottofamiglie disgiunte (“blocchi”), si ottengono variabili aleatorie indipendenti. Per semplicità di notazione, enunciamo il risultato nel caso di due blocchi.

Proposizione 3.39 (Indipendenza a blocchi). *Siano X_1, X_2, \dots, X_n variabili aleatorie indipendenti e siano $I = \{i_1, \dots, i_h\}$, $J = \{j_1, \dots, j_k\}$ sottoinsiemi non vuoti e disgiunti di $\{1, 2, \dots, n\}$. Allora le variabili aleatorie X_I e X_J , definite da*

$$X_I := (X_{i_1}, \dots, X_{i_h}), \quad X_J := (X_{j_1}, \dots, X_{j_k}),$$

sono indipendenti.

Dimostrazione. Osserviamo che le variabili aleatorie $(X_\ell)_{\ell \in I \cup J}$ sono indipendenti, così come lo sono le variabili aleatorie $(X_i)_{i \in I}$ e le variabili aleatorie $(X_j)_{j \in J}$. Di conseguenza, scrivendo per brevità $x_I := (x_{i_1}, \dots, x_{i_h})$ e $x_J := (x_{j_1}, \dots, x_{j_k})$, per ogni x_I e x_J si ha, grazie alla Proposizione 3.24,

$$\begin{aligned} P_{X_I, X_J}(x_I, x_J) &= P_{X_{i_1}, \dots, X_{i_h}, X_{j_1}, \dots, X_{j_k}}(x_{i_1}, \dots, x_{i_h}, x_{j_1}, \dots, x_{j_k}) \\ &= \prod_{r=1}^h \prod_{s=1}^k P_{X_{i_r}}(x_{i_r}) P_{X_{j_s}}(x_{j_s}) = \left(\prod_{r=1}^h P_{X_{i_r}}(x_{i_r}) \right) \left(\prod_{s=1}^k P_{X_{j_s}}(x_{j_s}) \right) \\ &= P_{X_I}(x_I) P_{X_J}(x_J), \end{aligned}$$

da cui segue l'indipendenza di X_I e X_J , grazie alla Proposizione 3.32. \square

La prossima proposizione stabilisce che *funzioni di variabili aleatorie indipendenti sono indipendenti*. Anche in questo caso, il risultato si estende facilmente a più di due variabili aleatorie. Qui e nel seguito, se $X : \Omega \rightarrow E$ e $f : E \rightarrow H$ sono due applicazioni, la loro composizione $f \circ X : \Omega \rightarrow H$ verrà indicata con $f(X)$.

Proposizione 3.40 (Indipendenza per trasformazioni). *Siano X e Y variabili aleatorie indipendenti, a valori rispettivamente negli insiemi E e F . Siano inoltre H, K due insiemi e $f : E \rightarrow H$, $g : F \rightarrow K$ funzioni arbitrarie. Allora anche le variabili aleatorie $f(X)$ e $g(Y)$ sono indipendenti.*

Dimostrazione. Basta osservare che, se $A \subseteq H$, $B \subseteq K$,

$$\begin{aligned} P(f(X) \in A, g(Y) \in B) &= P(X \in f^{-1}(A), Y \in g^{-1}(B)) \\ &= P(X \in f^{-1}(A)) P(Y \in g^{-1}(B)) = P(f(X) \in A) P(g(Y) \in B), \end{aligned}$$

dunque $f(X)$ e $g(Y)$ sono indipendenti. \square

Vediamo un esempio di applicazione congiunta delle Proposizioni 3.39 e 3.40.

Corollario 3.41. *Siano $X_1, \dots, X_n, X_{n+1}, \dots, X_{n+m}$ variabili aleatorie reali indipendenti, con $n, m \in \mathbb{N}$. Allora $X_1 + \dots + X_n$ e $X_{n+1} + \dots + X_{n+m}$ sono variabili aleatorie indipendenti.*

Dimostrazione. Applicando la Proposizione 3.39 con $I = \{1, 2, \dots, n\}$, $J = \{n+1, \dots, n+m\}$, si ha che $X_I = (X_1, \dots, X_n)$ e $X_J = (X_{n+1}, \dots, X_{n+m})$ sono variabili aleatorie indipendenti. Indicando con $f_k : \mathbb{R}^k \rightarrow \mathbb{R}$ la funzione somma, definita da $f_k(x_1, \dots, x_k) := x_1 + \dots + x_k$, possiamo allora scrivere $X_1 + \dots + X_n = f_n(X_I)$ e $X_{n+1} + \dots + X_{n+m} = f_m(X_J)$. La conclusione segue dalla Proposizione 3.40. \square

Concludiamo con alcune considerazioni, che sono una conseguenza immediata delle definizioni di indipendenza e di distribuzione per variabili aleatorie:

- due variabili aleatorie X, Y *indipendenti* devono essere necessariamente definite sullo *stesso* spazio di probabilità (Ω, P) , ma possono assumere valori in insiemi diversi E, F ;
- due variabili aleatorie X, Y *con la stessa distribuzione* devono invece assumere necessariamente valori nello stesso insieme E , ma possono essere definite su spazi di probabilità diversi (Ω, P) e (Ω', P') .

Di conseguenza, nell'enunciato della Proposizione 3.37, l'ipotesi che le variabili aleatorie X_1, \dots, X_n siano indipendenti sottintende la richiesta che esse siano definite sullo stesso spazio di probabilità, mentre l'ipotesi che le variabili aleatorie X'_1, \dots, X'_n abbiano la stessa distribuzione congiunta di X_1, \dots, X_n sottintende che X_i abbia lo stesso spazio di arrivo di X'_i . Queste condizioni non sono state esplicitamente indicate per alleggerire la notazione, ma è importante esserne coscienti: parlare di variabili aleatorie indipendenti definite su spazi di probabilità diversi non ha alcun senso! Discorsi analoghi valgono per gli enunciati delle Proposizioni 3.39 e 3.40.

3.2.5 Costruzione di variabili aleatorie indipendenti

Consideriamo il problema di costruire *una famiglia di variabili aleatorie indipendenti con distribuzioni marginali assegnate*. Più precisamente, siano assegnati degli insiemi arbitrari E_1, \dots, E_n e delle probabilità discrete μ_1 su E_1, \dots, μ_n su E_n (equivalentemente, le densità discrete associate p_1 su E_1, \dots, p_n su E_n), con $n \in \mathbb{N}$ fissato. Una domanda molto naturale dal punto di vista teorico, e molto importante per le applicazioni, è la seguente: esiste uno spazio di probabilità discreto (Ω, P) su cui siano definite variabili aleatorie *indipendenti* X_1, \dots, X_n , tali che X_i abbia distribuzione $\mu_{X_i} = \mu_i$ (equivalentemente, $p_{X_i} = p_i$) per ogni $i = 1, \dots, n$?

La risposta è *affermativa* ed è un semplice corollario della *costruzione canonica* descritta nel Paragrafo 3.1.6. Infatti, grazie alla Proposizione 3.32, il problema è equivalente alla costruzione di una variabile aleatoria $X := (X_1, \dots, X_n)$ a valori in $E := E_1 \times \dots \times E_n$ con densità discreta p data da[†]

$$p(x) := p_1(x_1) \cdots p_n(x_n), \quad \forall x = (x_1, \dots, x_n) \in E. \quad (3.25)$$

Come abbiamo visto nel Paragrafo 3.1.6, è sempre possibile costruire una tale variabile aleatoria su uno spazio di probabilità discreto (Ω, P) . In particolare, se gli insiemi E_1, \dots, E_n sono finiti o numerabili, si può scegliere come spazio campionario Ω lo spazio di arrivo “congiunto” $E = E_1 \times \dots \times E_n$ e come probabilità P su Ω quella associata alla densità discreta (3.25).

Osservazione 3.42. È naturale chiedersi se si possa passare dal caso finito al caso infinito, diciamo numerabile: assegnati arbitrariamente una successione di insiemi $(E_i)_{i \in \mathbb{N}}$ e, per ogni $i \in \mathbb{N}$, una densità discreta p_i su E_i , esiste uno spazio di probabi-

[†] Si verifica facilmente, usando la somma a blocchi (0.11), che la formula (3.25) definisce effettivamente una densità discreta p su E ; inoltre, se $X = (X_1, \dots, X_n)$ ha densità discreta $p_X = p$ data da (3.25), segue facilmente dalla Proposizione 3.24 che $p_{X_i} = p_1, \dots, p_{X_n} = p_n$.

lità discreto (Ω, \mathcal{P}) su cui siano definite variabili aleatorie *indipendenti* $(X_i)_{i \in \mathbb{N}}$, tali che la densità discreta di X_i sia $p_{X_i} = p_i$, per ogni $i \in \mathbb{N}$?

Questa volta la risposta è, in generale, negativa: la costruzione richiede la nozione generale di spazio di probabilità, che discuteremo nel Capitolo 5. Per esempio, immaginiamo di voler costruire una successione infinita di prove ripetute e indipendenti con probabilità di successo $p \in (0, 1)$, ossia una successione $(X_i)_{i \in \mathbb{N}}$ di variabili aleatorie discrete indipendenti, con spazi di arrivo $E_i = \{0, 1\}$ e distribuzioni marginali $p_i(\cdot) = q(\cdot)$ date da (3.23). Mostriamo che *non esiste alcuno spazio di probabilità discreto (Ω, \mathcal{P}) su cui una tale successione $(X_i)_{i \in \mathbb{N}}$ può essere definita*. Procediamo per assurdo: supponiamo che tale spazio esista e fissiamo $\bar{\omega} \in \Omega$. Definendo $\bar{x}_i := X_i(\bar{\omega})$ per ogni $i \in \mathbb{N}$, per costruzione si ha

$$\{\bar{\omega}\} \subseteq \{X_1 = \bar{x}_1, \dots, X_n = \bar{x}_n\}, \quad \forall n \in \mathbb{N}.$$

Dato che le variabili aleatorie $(X_i)_{1 \leq i \leq n}$ sono indipendenti, ricordando la relazione (3.23) segue che per ogni $n \in \mathbb{N}$

$$\begin{aligned} 0 \leq P(\{\bar{\omega}\}) &\leq P(X_1 = \bar{x}_1, \dots, X_n = \bar{x}_n) \\ &= P(X_1 = \bar{x}_1) \cdots P(X_n = \bar{x}_n) = q(\bar{x}_1) \cdots q(\bar{x}_n) \leq \alpha^n, \end{aligned}$$

dove $\alpha := \max\{p, 1 - p\} < 1$. Dato che $\lim_{n \rightarrow \infty} \alpha^n = 0$, segue che $P(\{\bar{\omega}\}) = 0$. Ma l'elemento $\bar{\omega} \in \Omega$ è arbitrario, dunque $P(\{\omega\}) = 0$ per ogni $\omega \in \Omega$, il che è chiaramente impossibile, dovendo essere $\sum_{\omega \in \Omega} P(\{\omega\}) = 1$.

Una facile estensione di questo argomento mostra che *su uno spazio di probabilità discreto non è possibile definire una successione $(X_i)_{i \in \mathbb{N}}$ di variabili aleatorie indipendenti e con la stessa distribuzione*, eccetto nel caso “banale” in cui tale distribuzione sia concentrata in un punto, ossia le variabili X_i siano q.c. costanti. \square

3.2.6 Dallo spazio di probabilità alle variabili aleatorie

Come si sarà notato, in questo capitolo c'è stato uno spostamento progressivo di enfasi *dallo spazio di probabilità alle variabili aleatorie*, che saranno sempre più l'oggetto centrale della nostra analisi. Le ragioni sono molteplici e si chiariranno nel seguito, ma vale la pena di fare alcune considerazioni.

Se in molti casi è naturale formulare un problema probabilistico in termini di variabili aleatorie, in altri casi l'uso di variabili aleatorie è una questione di “gusto”. Ad esempio, per descrivere l'esito del lancio di un dado regolare a sei facce:

- (i) si può considerare lo spazio di probabilità (E, μ) , in cui $E := \{1, 2, 3, 4, 5, 6\}$ e μ è la probabilità uniforme su E , senza fare riferimento a variabili aleatorie;
- (ii) in alternativa, si può considerare una variabile aleatoria X , *definita su uno spazio di probabilità (Ω, \mathcal{P}) generico*, a valori in E e con distribuzione $\mu_X = \mu$.

Preferiremo il secondo approccio, all'apparenza più astratto, perché permette più flessibilità. Ad esempio, se successivamente si vuole incorporare nell'esperimento anche il lancio di una moneta equilibrata, è sufficiente supporre che su (Ω, \mathcal{P}) sia definita

una seconda variabile Y , indipendente da X e con distribuzione uniforme nell'insieme $\{T, C\}$. A tal fine, potrebbe essere necessario ampliare lo spazio di probabilità originario (Ω, \mathcal{P}) , come nell'Esempio 3.12, ma è *possibile fare questa estensione implicitamente*, continuando a chiamare X l'esito del dado, perché non abbiamo imposto vincoli particolari su (Ω, \mathcal{P}) .

Nel primo approccio, invece, per incorporare il lancio della moneta, il modello deve essere modificato *esplicitamente*, considerando lo spazio di probabilità (E', μ') dove $E' = E \times \{T, C\}$ e μ' è la probabilità uniforme su E' . Per di più, se in questo modello ampliato si vuole descrivere l'esito del lancio del dado, si è “costretti” a introdurre una variabile aleatoria, la proiezione $X : E' \rightarrow E$ (come nell'Esempio 3.12): ma questo è un caso particolare del secondo approccio, in cui $(\Omega, \mathcal{P}) = (E', \mu')$! Tanto vale, allora, usare variabili aleatorie sin dal principio, lasciando implicito lo spazio di probabilità (Ω, \mathcal{P}) su cui sono definite.

Un'altra ragione fondamentale per l'uso di variabili aleatorie *reali* è la loro ricca struttura matematica: esse possono essere sommate, moltiplicate, ecc. e se ne può definire il valor medio, come vedremo a breve, che permette di ottenere *stime esplicite sulle probabilità di eventi* (cf. il Teorema 3.77). Ne vedremo un'applicazione importante nel Capitolo 7, con la legge dei grandi numeri.

Nel seguito formuleremo spesso problemi nel modo seguente: “Siano X_1, \dots, X_n variabili aleatorie indipendenti, con le seguenti distribuzioni marginali...”. Per quanto visto nel Paragrafo 3.2.5, problemi di questo tipo sono ben posti, nel senso che esiste sempre uno spazio di probabilità (Ω, \mathcal{P}) – ottenuto esplicitamente grazie alla costruzione canonica del Paragrafo 3.1.6 – su cui sono definite variabili aleatorie con le proprietà richieste. Sottolineiamo tuttavia che *qualunque altro spazio di probabilità andrebbe bene*: infatti, la probabilità di qualunque evento che dipende dalle variabili aleatorie X_1, \dots, X_n può essere calcolata conoscendone la distribuzione congiunta e questa, a sua volta, è univocamente determinata dalle distribuzioni marginali fissate, grazie alla Proposizione 3.32.

In conclusione, se un problema può essere espresso in funzione di una famiglia di variabili aleatorie, quello che conta è la loro distribuzione congiunta: dettagli aggiuntivi sullo spazio di probabilità su cui esse sono definite non sono rilevanti. Per questa ragione, lo spazio di probabilità resterà sempre più “sullo sfondo” e spesso non sarà nemmeno menzionato (ma è concettualmente importante sapere che c'è!).

Esercizi

Esercizio 3.4. Siano A e B due eventi in uno spazio di probabilità discreto (Ω, \mathcal{P}) . Si determini la densità congiunta del vettore aleatorio $(\mathbb{1}_A, \mathbb{1}_B)$.

Esercizio 3.5. Siano C_1, C_2, \dots, C_n eventi in uno spazio di probabilità discreto (Ω, \mathcal{P}) . Posto $X_i := \mathbb{1}_{C_i}$, si mostri l'equivalenza delle seguenti affermazioni

- (i) le variabili aleatorie X_1, X_2, \dots, X_n sono indipendenti;
- (ii) gli eventi C_1, C_2, \dots, C_n sono indipendenti.

Si mostri quindi che vale lo stesso per una famiglia $(C_i)_{i \in I}$ arbitraria di eventi.

[Sugg. Si ricordi la Proposizione 1.63.]

Esercizio 3.6. Date n prove ripetute e indipendenti con probabilità di successo p , consideriamo le variabili aleatorie $S :=$ “numero di successi nelle n prove” e $T :=$ “prova in cui si ha il primo successo”. Si determini la densità congiunta $p_{S,T}$.

(Osserviamo che le densità marginali p_S e p_T sono già state determinate nell'Esempio 3.19: si veda la relazione (3.14)).

3.3 Valor medio e disuguaglianze

La nozione di *media aritmetica*

$$\mu = \frac{x_1 + \dots + x_n}{n}$$

di un insieme finito di numeri reali $\{x_1, x_2, \dots, x_n\}$ è ben nota e molto naturale. Una delle sue possibili interpretazioni è quella che si ottiene considerando un sistema di n punti materiali posizionati su una retta, ognuno nel punto corrispondente alla coordinata x_i . Se i punti materiali hanno tutti la stessa massa, il punto di coordinata pari alla media aritmetica è il baricentro del sistema.

Nel caso in cui i punti non abbiano tutti la stessa massa, il baricentro si ottiene attraverso una media “pesata”: più precisamente, se m_i è la massa del punto materiale in x_i , indicando con $M := m_1 + \dots + m_n$ la massa totale, il baricentro del sistema ha coordinata

$$\mu = \frac{x_1 m_1 + \dots + x_n m_n}{M} = x_1 p_1 + \dots + x_n p_n,$$

dove i “pesi” p_i sono dati dalle masse relative, ossia

$$p_i := \frac{m_i}{M} = \frac{m_i}{m_1 + \dots + m_n}.$$

Si noti p_1, \dots, p_n possono essere visti come i valori di una densità discreta su \mathbb{R} , poiché $p_i \geq 0$ e $\sum_{i=1}^n p_i = 1$.

3.3.1 Definizione

Il *valor medio* $E(X)$ di una variabile aleatoria discreta *reale* X corrisponde precisamente alla nozione di baricentro sopra descritta, una volta interpretate le x_i come i valori assunti da X e i pesi $p_i = P(X = x_i) = p_X(x_i)$ come le probabilità che tali valori vengano assunti.

Dal momento che una variabile aleatoria discreta può assumere una quantità infinita (numerabile) di valori, occorre un po' di cautela. Ricordiamo che la nozione di somma di una famiglia $(a_x)_{x \in \mathbb{R}}$ di numeri reali è stata definita nel capitolo introduttivo “Nozioni Preliminari”.

Definizione 3.43 (Valor medio). Sia X una variabile aleatoria discreta *reale*, con densità discreta p_X . Si dice che X ammette valor medio se la famiglia di numeri reali $(x p_X(x))_{x \in \mathbb{R}}$ ammette somma; in questo caso, si definisce valor medio la somma di tale famiglia.

Più esplicitamente, la variabile aleatoria reale X ammette valor medio se almeno una delle due somme seguenti è finita:

$$\sum_{x \in (0, \infty)} x p_X(x) < +\infty, \quad \sum_{x \in (-\infty, 0)} x p_X(x) > -\infty. \quad (3.26)$$

Se X ammette valor medio, si definisce *valor medio* (o *media*, *valore atteso*, *speranza matematica*, *aspettativa*) di X la somma delle due somme in (3.26):

$$E(X) := \sum_{x \in \mathbb{R}} x p_X(x) = \sum_{x \in X(\Omega)} x p_X(x) \in [-\infty, +\infty]. \quad (3.27)$$

Osservazione 3.44. La seconda uguaglianza in (3.27) è dovuta al fatto che $p_X(x) = 0$ se $x \notin X(\Omega)$. Analogamente, le somme in (3.26) possono essere ristrette rispettivamente a $X(\Omega) \cap (0, \infty)$ e $X(\Omega) \cap (-\infty, 0)$.

Ricordiamo che $X(\Omega)$ è sempre un insieme finito o numerabile. In particolare, se X assume un insieme finito di valori, allora X ammette valor medio finito, perché entrambe le somme in (3.26) hanno un numero finito di addendi non nulli. \square

Esempio 3.45 (Variabili aleatorie costanti). Sia X una variabile aleatoria reale quasi certamente costante, ossia esiste $c \in \mathbb{R}$ tale che $P(X = c) = 1$. Per l'Esempio 3.17, si ha $p_X(c) = 1$ e $p_X(x) = 0$ se $x \neq c$, dunque X ammette valor medio dato da

$$E(X) = \sum_{x \in \mathbb{R}} x \mathbb{1}_{\{c\}}(x) = c.$$

come segue dalla relazione (3.27). \square

Esempio 3.46 (Variabili aleatorie indicatrici). Sia $X := \mathbb{1}_A$ la variabile aleatoria indicatrice di un evento $A \subseteq \Omega$. Ricordando l'Esempio 3.18, la densità discreta di X vale $p_X(1) = P(A)$ e $p_X(0) = 1 - P(A)$. Pertanto X ammette valor medio, dato da

$$E(X) = 0 \cdot (1 - P(A)) + 1 \cdot P(A) = P(A),$$

avendo applicato la relazione (3.27). \square

Un'osservazione fondamentale è che *ogni variabile aleatoria positiva X ammette valor medio* (che può eventualmente valere $+\infty$): infatti, se $X \geq 0$, si ha $p_X(x) = 0$ per $x < 0$ e dunque la seconda somma in (3.26) è nulla. Di conseguenza, $E(X)$ è sempre ben definito (dalla formula (3.27)) se la variabile aleatoria X è positiva.

Ora, per ogni variabile aleatoria reale X , definita su uno spazio di probabilità (Ω, P) , possiamo sempre considerare le variabili aleatorie positive

$$X^+ := \max\{X, 0\}, \quad X^- := \max\{-X, 0\}, \quad |X| = X^+ + X^-,$$

ossia $X^+(\omega) := \max\{X(\omega), 0\}$ per ogni $\omega \in \Omega$, ecc. I valori medi $E(X^+)$, $E(X^-)$ e $E(|X|)$ sono sempre ben definiti, e applicando la formula (3.27) si ricava[†]

$$E(X^+) = \sum_{x \in (0, \infty)} x p_X(x), \quad E(X^-) = - \sum_{x \in (-\infty, 0)} x p_X(x), \quad (3.28)$$

$$E(|X|) = \sum_{x \in \mathbb{R}} |x| p_X(x) = E(X^+) + E(X^-). \quad (3.29)$$

L'utilità di queste espressioni è che esse permettono di stabilire se una variabile aleatoria X ammette valore medio, e di determinarlo, analizzando i valori medi $E(X^+)$, $E(X^-)$ e $E(|X|)$, che sono sempre ben definiti. Possiamo infatti riformulare la definizione di valor medio come segue.

Proposizione 3.47. *Una variabile aleatoria reale X ammette valor medio se e solo se almeno uno tra i valori medi $E(X^+)$ e $E(X^-)$, dati da (3.28), è finito. In tal caso, si ha*

$$E(X) = E(X^+) - E(X^-).$$

Inoltre, X ammette valor medio finito se e solo se $E(|X|) < \infty$ (equivalentemente, se e solo se entrambi i valori medi $E(X^+)$ e $E(X^-)$ sono finiti).

Prima di enunciare le proprietà fondamentali del valor medio, discutiamone l'interpretazione. Abbiamo visto che $E(X)$ rappresenta una sorta di “baricentro” della distribuzione dei valori di X , pesati con le rispettive probabilità. Sottolineiamo tuttavia che spesso $E(X)$ non è il “valore più probabile” di X e nemmeno un valore che ci si aspetta tipicamente di osservare. Infatti, può benissimo accadere che $E(X)$ non sia neppure uno dei valori assunti da X , come mostrano gli esempi seguenti.

Esempio 3.48. Lanciamo un dado regolare a sei facce e indichiamo con X il numero ottenuto. Notiamo che $X(\Omega) = \{1, 2, 3, 4, 5, 6\}$, dunque $E(X)$ è ben definito e finito. Dato che $p_X(x) = \frac{1}{6}$ per $x \in X(\Omega)$, segue che

$$E(X) = \sum_{x \in X(\Omega)} x p_X(x) = 1 \frac{1}{6} + 2 \frac{1}{6} + \dots + 6 \frac{1}{6} = \frac{1+2+3+4+5+6}{6} = 3.5,$$

che mostra come in questo caso $E(X) \notin X(\Omega)$. □

Esempio 3.49 (Size bias). Riprendiamo l'Esempio 3.20, in cui 120 studenti sono suddivisi in 3 gruppi di 36, 40 e 44 studenti rispettivamente. Abbiamo mostrato che, se si sceglie un gruppo a caso, il numero X di studenti nel gruppo scelto ha densità discreta data da $p_X(36) = p_X(40) = p_X(44) = \frac{1}{3}$, pertanto

$$E(X) = \sum_{x \in X(\Omega)} x p_X(x) = 36 p_X(36) + 40 p_X(40) + 44 p_X(44) = 40.$$

[†] Se $x < 0$, si ha $p_{X^+}(x) = 0$; se $x = 0$, il termine $x p_{X^+}(x) = 0$ non contribuisce in (3.27); se $x > 0$, si ha $p_{X^+}(x) = P(X^+ = x) = P(X = x) = p_X(x)$; in questo modo si ottiene la formula (3.28) per $E(X^+)$. Un discorso simile vale per X^- e $|X|$, notando che $p_{|X|}(x) = p_X(x) + p_X(-x)$ per $x > 0$.

Se invece si sceglie uno studente a caso, il numero Y di studenti nel suo gruppo (lui compreso) ha densità discreta $p_Y(36) = \frac{3}{10}$, $p_Y(40) = \frac{1}{3}$, $p_Y(44) = \frac{11}{30}$, pertanto

$$E(Y) = \sum_{x \in Y(\Omega)} x p_Y(x) = 36 \frac{3}{10} + 40 \frac{1}{3} + 44 \frac{11}{30} = \frac{604}{15} \simeq 40.3.$$

Si noti che $E(Y) > E(X)$. La ragione intuitiva è che, scegliendo uno studente a caso, è più probabile sceglierne uno da un gruppo numeroso. Effetti di questo tipo sono indicati nella letteratura anglofona con il termine *size bias*. \square

Oltre all'idea di “baricentro” della distribuzione di una variabile aleatoria X , un'interpretazione fondamentale, più concreta e più probabilistica, del valor medio è fornita dalla *legge dei grandi numeri*, un teorema che studieremo nel Capitolo 7. Anticipiamone il contenuto per sommi capi.

Consideriamo n variabili aleatorie X_1, \dots, X_n indipendenti e con la stessa distribuzione di X . Se X rappresenta una quantità che dipende dall'esito di un esperimento aleatorio – per esempio, il numero ottenuto lanciando un dado regolare a sei facce – le variabili aleatorie X_1, \dots, X_n rappresentano l'analoga quantità osservata in n ripetizioni indipendenti dello stesso esperimento aleatorio – per esempio, i numeri ottenuti lanciando n volte un dado regolare a sei facce. Osserviamo che le n ripetizioni dell'esperimento aleatorio in questione costituiscono un nuovo (macro-)esperimento aleatorio, di cui X_1, \dots, X_n sono funzione.

Consideriamo ora la media aritmetica $\bar{X}_n := \frac{1}{n}(X_1 + \dots + X_n)$, detta *media campionaria*. Si noti che \bar{X}_n è anch'essa una variabile aleatoria (nell'esempio dei dadi, se si ripetono gli n lanci, si ottiene in generale un valore diverso di \bar{X}_n). È naturale chiedersi quali valori possa assumere \bar{X}_n e con quali probabilità. La legge dei grandi numeri afferma che, se X ammette valor medio finito, la variabile aleatoria \bar{X}_n assume con grande probabilità valori vicini a $E(X)$, quando n è sufficientemente grande. I dettagli precisi verranno forniti nel Capitolo 7.

In definitiva, il valor medio $E(X)$, quando esiste finito, può essere interpretato come il “valore tipico” (non di X , ma) di \bar{X}_n , ossia della media aritmetica dei risultati di n “copie indipendenti di X ”, se n è sufficientemente grande. Un'illustrazione concreta è fornita dall'esempio seguente.

Esempio 3.50 (Strategia del raddoppio). Riprendiamo l'Esempio 3.21, in cui un giocatore adotta una strategia “al raddoppio” puntando sull'uscita del “rosso” alla roulette, partendo con un capitale di 1023 euro. Abbiamo mostrato che il guadagno (con segno) X del giocatore, ossia la differenza tra il capitale finale e quello iniziale, può assumere solo i valori $X(\Omega) = \{-1023, 1\}$, con probabilità rispettive

$$p_X(-1023) = q^{10}, \quad p_X(1) = 1 - q^{10}, \quad \text{dove} \quad q := \frac{19}{37}.$$

(Ricordiamo che q è la probabilità che in un giro di roulette non esca il rosso.) Il valor medio è pertanto dato da

$$E(X) = -1023 p_X(-1023) + 1 p_X(1) = 1 - 1024 q^{10} = 1 - (2q)^{10}.$$

Dato che $2q = 2\frac{19}{37} > 1$, si ha $E(X) < 0$ (il valore numerico è $E(X) \simeq -0.3$).

Il fatto che il valor medio $E(X)$ sia strettamente negativo ha conseguenze molto rilevanti. Supponiamo infatti che il giocatore si rechi al casinò per n giorni, ogni volta con un capitale di 1023 euro, adottando ogni giorno la strategia descritta. Se indichiamo con X_i il guadagno ottenuto l' i -esimo giorno, è naturale supporre che X_1, \dots, X_n siano variabili aleatorie indipendenti con la stessa distribuzione di X .

Per la legge dei grandi numeri, la *media aritmetica* $\bar{X}_n = \frac{1}{n}(X_1 + \dots + X_n)$ dei guadagni ottenuti negli n giorni è una variabile aleatoria che, se n è elevato, assume con grande probabilità valori vicini a $E(X) \simeq -0.3$; dunque, la *somma* $X_1 + \dots + X_n$ dei guadagni ottenuti negli n giorni assume con grande probabilità valori dell'ordine di $E(X) \cdot n \simeq -0.3 \cdot n$. Di conseguenza, ripetendo la strategia un numero elevato n di volte, si è pressoché certi di finire in grande perdita! \square

3.3.2 Proprietà del valor medio

Cominciamo con un'osservazione elementare ma importante.

Osservazione 3.51. Il fatto che una variabile aleatoria reale X ammetta valor medio e, in tal caso, il valore di $E(X)$, sono *proprietà della distribuzione di X* . Infatti nella Definizione 3.43 compare solo la densità discreta p_X . \square

Mostriamo ora che la conoscenza della densità discreta p_X di una variabile aleatoria X permette di determinare il valor medio di *ogni funzione di X* .

Proposizione 3.52 (Calcolo del valor medio). *Sia X una variabile aleatoria a valori in un insieme generico E e sia $g : E \rightarrow \mathbb{R}$ una funzione reale. La variabile aleatoria $g(X)$ ammette valor medio se e solo se la famiglia di numeri reali $(g(x) p_X(x))_{x \in E}$ ammette somma; in questo caso, $E(g(X))$ è dato da*

$$E(g(X)) = \sum_{x \in E} g(x) p_X(x) = \sum_{x \in X(\Omega)} g(x) p_X(x). \quad (3.30)$$

Sottolineiamo che la formula (3.30) può essere sempre applicata alle variabili aleatorie $g(X)^+$, $g(X)^-$ e $|g(X)|$, che sono positive e dunque ammettono valor medio. Ciò permette, grazie alla Proposizione 3.47, di determinare se $g(X)$ ammette valor medio e se esso è finito. Per esempio, $g(X)$ ammette valor medio finito se e solo se

$$E(|g(X)|) = \sum_{x \in E} |g(x)| p_X(x) = \sum_{x \in X(\Omega)} |g(x)| p_X(x) < \infty. \quad (3.31)$$

Dimostrazione (della Proposizione 3.52). La densità discreta $p_{g(X)}$ della variabile aleatoria $g(X)$ è data da

$$\begin{aligned} p_{g(X)}(t) &= P(g(X) = t) = P(X \in g^{-1}(\{t\})) = \sum_{x \in g^{-1}(\{t\})} p_X(x) \\ &= \sum_{x \in E: g(x)=t} p_X(x), \quad \forall t \in \mathbb{R}. \end{aligned} \quad (3.32)$$

Assumiamo per il momento che la funzione g sia a valori in \mathbb{R}^+ . In questo caso $g(X)$ ammette valor medio, perché è una variabile aleatoria positiva, e la famiglia di numeri reali $(g(x) p_X(x))_{x \in E}$ ammette somma, perché a termini positivi. Applicando la formula (3.27), otteniamo

$$\begin{aligned} E(g(X)) &= \sum_{t \in \mathbb{R}} t p_{g(X)}(t) = \sum_{t \in \mathbb{R}} t \sum_{x \in E: g(x)=t} p_X(x) = \sum_{t \in \mathbb{R}} \sum_{x \in E: g(x)=t} t p_X(x) \\ &= \sum_{t \in \mathbb{R}} \sum_{x \in E: g(x)=t} g(x) p_X(x) = \sum_{x \in E} g(x) p_X(x), \end{aligned}$$

avendo usato nell'ultima uguaglianza la somma a blocchi (0.10), perché

$$E = \bigcup_{t \in \mathbb{R}} \{x \in E : g(x) = t\},$$

e l'unione è disgiunta. La formula (3.31) è dunque dimostrata se $g \geq 0$.

Rimuoviamo ora l'ipotesi che g sia positiva e poniamo $a_x := g(x) p_X(x)$ per semplicità. La famiglia di numeri reali $(a_x)_{x \in E}$ ammette somma se e solo se almeno una delle due somme $\sum_{x \in E} a_x^+$ e $\sum_{x \in E} a_x^-$ è finita, nel qual caso $\sum_{x \in E} a_x := \sum_{x \in E} a_x^+ - \sum_{x \in E} a_x^-$. Dato che $a_x^\pm = g(x)^\pm p_X(x)$, la relazione (3.31) applicata alle funzioni positive g^+ e g^- mostra che

$$\sum_{x \in E} a_x^+ = E(g(X)^+), \quad \sum_{x \in E} a_x^- = E(g(X)^-),$$

Pertanto la famiglia $(a_x)_{x \in E}$ ammette somma se e solo se almeno uno tra i valori medi $E(g(X)^+)$ e $E(g(X)^-)$ è finito, e in tal caso $\sum_{x \in E} a_x = E(g(X)^+) - E(g(X)^-)$. Ricordando la Proposizione 3.47, la dimostrazione è conclusa. \square

Osservazione 3.53. La formula (3.30) è di uso *molto* frequente nei calcoli (per esempio, essa è stata già implicitamente applicata in (3.28) e (3.29)). Essa può essere vista come una generalizzazione della formula (3.27) che definisce il valor medio, a cui si riduce se $E = \mathbb{R}$ e $g(x) = x$ è l'identità.

Notiamo che il valor medio della variabile aleatoria $Y = g(X)$ può essere anche ottenuto applicando la Definizione 3.43, determinando innanzitutto la densità discreta $p_Y = p_{g(X)}$, come in (3.32), e applicando quindi la formula (3.27) a Y . Tuttavia, in molti casi l'applicazione diretta della formula (3.30) risulta più pratica: si veda l'Esempio 3.64 più sotto (accessibile già da ora). \square

Diamo ora una formula alternativa per il valor medio, di grande rilevanza teorica, in cui si effettua la somma sullo spazio di partenza Ω anziché sullo spazio di arrivo \mathbb{R} . La dimostrazione è del tutto analoga a quella della Proposizione 3.52.

Proposizione 3.54. *Sia X una variabile aleatoria reale, definita su uno spazio di probabilità discreto (Ω, P) . X ammette valor medio se e solo se la famiglia di numeri reali $(X(\omega) P(\{\omega\}))_{\omega \in \Omega}$ ammette somma; in questo caso si ha:*

$$E(X) = \sum_{\omega \in \Omega} X(\omega) P(\{\omega\}). \quad (3.33)$$

Dimostrazione. Se X è positiva, il suo valor medio $E(X)$ è ben definito ed è dato da

$$\begin{aligned} E(X) &= \sum_{x \in \mathbb{R}} x p_X(x) = \sum_{x \in \mathbb{R}} x P(X=x) = \sum_{x \in \mathbb{R}} x \sum_{\omega \in \{X=x\}} P(\{\omega\}) \\ &= \sum_{x \in \mathbb{R}} \sum_{\omega \in \{X=x\}} x P(\{\omega\}) = \sum_{x \in \mathbb{R}} \sum_{\omega \in \{X=x\}} X(\omega) P(\{\omega\}), \end{aligned}$$

dove nell'ultima uguaglianza abbiamo usato il fatto che $X(\omega) = x$ per ogni $\omega \in \{X=x\}$. Analogamente a (3.16), possiamo scrivere $\Omega = \bigcup_{x \in \mathbb{R}} \{X=x\}$ e l'unione è disgiunta, pertanto per la somma a blocchi (0.10) otteniamo la relazione (3.33).

Rimuovendo l'ipotesi che X sia positiva, la relazione (3.33) vale per X^+ e X^- . Da ciò segue che la famiglia $(X(\omega) P(\{\omega\}))_{\omega \in \Omega}$ ammette somma se e solo se almeno uno tra i valori medi $E(X^+)$ e $E(X^-)$ è finito, e in tal caso $\sum_{\omega \in \Omega} X(\omega) P(\{\omega\}) = E(X^+) - E(X^-)$. Questo completa la dimostrazione, grazie alla Proposizione 3.47. \square

Enunciamo quindi alcune proprietà del valor medio di grandissima importanza, sia sul piano teorico sia su quello applicativo. Le dimostrazioni sono una semplice conseguenza della Proposizione 3.54 e delle proprietà delle somme infinite, descritte nel capitolo introduttivo “Nozioni Preliminari”, e sono lasciate al lettore.

Proposizione 3.55 (Proprietà del valor medio). *Siano X, Y variabili aleatorie discrete reali, definite nello stesso spazio di probabilità discreto (Ω, P) .*

- (i) **(Monotonia)** *Se X, Y ammettono entrambe valor medio e $X(\omega) \leq Y(\omega)$ per ogni $\omega \in \Omega$, allora $E(X) \leq E(Y)$.*
- (ii) *Se X ammette valor medio, allora*

$$|E(X)| \leq E(|X|). \quad (3.34)$$

- (iii) **(Linearità)** *Se X e Y ammettono valor medio finito (risp. se X e Y sono positive) e $a, b \in \mathbb{R}$ (risp. $a, b \in \mathbb{R}^+$), allora la variabile aleatoria $aX + bY$ definita da*

$$(aX + bY)(\omega) := aX(\omega) + bY(\omega), \quad \forall \omega \in \Omega,$$

ammette valor medio finito (risp. è positiva) e

$$E(aX + bY) = aE(X) + bE(Y).$$

La proprietà di linearità del valor medio, in particolare, è di uso molto frequente e può avere conseguenze meno ovvie di quanto si potrebbe supporre.

Esempio 3.56 (Valor medio, compleanni e permutazioni). Siano dati gli eventi (non necessariamente indipendenti) C_1, \dots, C_m , in uno spazio di probabilità discreto (Ω, P) . Consideriamo la variabile aleatoria reale

$$X := \mathbb{1}_{C_1} + \dots + \mathbb{1}_{C_m},$$

che ha un significato molto intuitivo: X conta il numero di eventi C_1, \dots, C_m che si verificano. La distribuzione di X dipende dalle relazioni che ci sono tra gli eventi C_1, \dots, C_m e, in generale, può essere complicata. Tuttavia, *il valor medio di X è sempre dato da* (si ricordi l'Esempio 3.46)

$$E(X) = E(\mathbb{1}_{C_1}) + \dots + E(\mathbb{1}_{C_m}) = P(C_1) + \dots + P(C_m).$$

In particolare, $E(X) = mp$ se tutti gli eventi hanno la stessa probabilità $p = P(C_i)$.

Riconsideriamo allora l'Esempio 1.33 (paradosso dei compleanni). Selezioniamo casualmente n persone nate in un anno non bisestile, numeriamole da 1 a n e introduciamo, per ogni coppia di persone $\{i, j\} \subseteq \{1, \dots, n\}$ (con $i \neq j$), l'evento

$$C_{\{i,j\}} := \text{"le persone } i \text{ e } j \text{ sono nate lo stesso giorno"}.$$

Abbiamo allora $m = \binom{n}{2} = \frac{1}{2}n(n-1)$ eventi $C_{\{i,j\}}$ (non indipendenti!), tutti con la stessa probabilità $p = P(C_{\{i,j\}}) = \frac{1}{365}$. In questo caso, la variabile aleatoria

$$X := \sum_{\{i,j\} \subseteq \{1, \dots, n\}} \mathbb{1}_{C_{\{i,j\}}}$$

conta il numero di coppie che compiono gli anni lo stesso giorno e ha valor medio

$$E(X) = mp = \frac{1}{2}n(n-1) \cdot \frac{1}{365} = \frac{(n-1)n}{730}. \quad (3.35)$$

Sappiamo che già per $n = 23$, dunque per un gruppo relativamente piccolo, la probabilità che almeno due persone compiano gli anni lo stesso giorno, ossia $P(X \geq 1)$, è strettamente maggiore di $\frac{1}{2}$. Questo risultato è “confermato” dalla relazione (3.35): infatti, per $n = 23$ si ha $E(X) \simeq 0.7$, un valore vicino a 1. Ciò fornisce una possibile “spiegazione” al paradosso dei compleanni: sebbene il numero di persone $n = 23$ sia relativamente basso, la quantità rilevante è il numero di *coppie* $\frac{1}{2}n(n-1) = 253$, che è comparabile col numero 365 di giorni dell'anno.

Un altro esempio interessante è fornito dalle permutazioni aleatorie, in particolare dal Problema 2.4 analizzato nel Paragrafo 2.1. Indicando con $X(\sigma)$ il numero di punti fissi di una permutazione $\sigma \in S_n$, possiamo scrivere

$$X = \sum_{i=1}^n \mathbb{1}_{C_i}, \quad \text{dove} \quad C_i = \{\sigma \in S_n : \sigma(i) = i\}.$$

Munendo S_n della probabilità uniforme P_n , si calcola $p = P_n(C_i) = (n-1)!/n! = \frac{1}{n}$; di conseguenza, X ha valor medio $E_n(X) = np = 1$, che non dipende da $n \in \mathbb{N}$. Abbiamo dunque mostrato il fatto, non intuitivo, che *una permutazione di n elementi scelta uniformemente ha “mediamente” un punto fisso, qualunque sia $n \in \mathbb{N}$* . \square

È possibile sfruttare la proprietà di monotonia del valor medio, espressa nella Proposizione 3.55(i), per *mostrare* che una variabile aleatoria reale X ammette valor me-

dio finito. Il caso tipico è il seguente: siano Z, X due variabili aleatorie reali, definite sullo stesso spazio di probabilità (Ω, \mathcal{P}) , tali che

$$|X(\omega)| \leq Z(\omega), \quad \text{per ogni } \omega \in \Omega.$$

Allora, se Z ammette valor medio finito, anche X ammette valor medio finito. Infatti, già sappiamo che $|X|$ ammette valor medio, perché positiva, e applicando la proprietà di monotonia si ottiene $E(|X|) \leq E(Z) < \infty$; dunque X ammette valor medio finito, per la Proposizione 3.47 (e inoltre $|E(X)| \leq E(|X|) \leq E(Z)$).

Come caso particolare, dato che le variabili aleatorie costanti ammettono valor medio finito (si ricordi l'Esempio 3.45), otteniamo il seguente utile criterio.

Proposizione 3.57. *Sia X una variabile aleatoria reale limitata:*

$$\exists C \in \mathbb{R}^+ : \quad |X(\omega)| \leq C, \quad \text{per ogni } \omega \in \Omega.$$

Allora X ammette valor medio finito: $|E(X)| \leq C < \infty$.

Un'altra utile conseguenza dalla Proposizione 3.54 è il seguente risultato.

Proposizione 3.58. *Sia X una variabile aleatoria reale positiva:*

$$X(\omega) \geq 0, \quad \text{per ogni } \omega \in \Omega.$$

Se $E(X) = 0$, allora X è quasi certamente uguale a zero, ossia $P(X = 0) = 1$.

Dimostrazione. La famiglia di numeri reali $(X(\omega) P(\{\omega\}))_{\omega \in \Omega}$ ha tutti i termini positivi e dunque ammette somma. Applicando la formula (3.33), si ha

$$E(X) = \sum_{\omega \in \Omega} X(\omega) P(\{\omega\}), \quad (3.36)$$

e per ipotesi $E(X) = 0$. Ma una somma a termini positivi vale zero se e solo se ogni addendo è nullo: $X(\omega) P(\{\omega\}) = 0$ per ogni $\omega \in \Omega$, ossia $P(\{\omega\}) = 0$ per tutti gli ω tali che $X(\omega) > 0$. Ciò implica che $P(X > 0) = 0$ (perché?) e dunque

$$P(X = 0) = P(\{X \geq 0\} \setminus \{X > 0\}) = P(X \geq 0) - P(X > 0) = 1 - 0 = 1,$$

come volevasi dimostrare. \square

Concludiamo infine con un'osservazione importante.

Osservazione 3.59. Siano X e X' variabili aleatorie reali, definite sullo stesso spazio di probabilità (Ω, \mathcal{P}) , quasi certamente uguali, ossia tali che $P(X = X') = 1$. Allora X ammette valor medio se e solo se X' ammette valor medio, nel qual caso $E(X) = E(X')$. Questo segue immediatamente dal fatto che X e X' hanno la stessa distribuzione, come mostrato nella Proposizione 3.14, e dall'Osservazione 3.51. \square

Grazie all'osservazione precedente, è possibile generalizzare molte proprietà del valor medio. Per esempio, le ipotesi di limitatezza della Proposizione 3.57 e di positività della Proposizione 3.58 possono essere indebolite, richiedendo soltanto che $P(|X| \leq C) = 1$ (ossia che X sia “quasi certamente limitata”) e $P(X \geq 0) = 1$ (ossia che X sia “quasi certamente positiva”), rispettivamente. Un discorso analogo vale per la proprietà di monotonia del valor medio (Proposizione 3.55): la condizione $X(\omega) \leq Y(\omega)$ per ogni $\omega \in \Omega$ può essere sostituita da $P(X \leq Y) = 1$.

3.3.3 Momenti, varianza e covarianza

Sia (Ω, P) uno spazio di probabilità discreto fissato. L'insieme delle variabili aleatorie reali definite su Ω che ammettono valor medio finito riveste grande importanza e viene indicato con il simbolo $L^1(\Omega, P)$:

$$\begin{aligned} L^1(\Omega, P) &:= \{X : \Omega \rightarrow \mathbb{R} \text{ tali che } X \text{ ammette valor medio finito}\} \\ &= \{X : \Omega \rightarrow \mathbb{R} \text{ tali che } E(|X|) < +\infty\}. \end{aligned} \quad (3.37)$$

D'ora in avanti, scriveremo $X \in L^1(\Omega, P)$, o anche soltanto $X \in L^1$, per indicare che X ammette valore medio finito. Si osservi che $X \in L^1$ se e soltanto se $|X| \in L^1$.

Più in generale, per ogni $p \in (0, \infty)$ fissato, associamo a ogni variabile aleatoria reale X un numero $\|X\|_p \in [0, +\infty]$, detto *norma p di X*:

$$\|X\|_p := [E(|X|^p)]^{1/p} \in [0, +\infty]. \quad (3.38)$$

(Si noti che $|X|^p$ è una variabile aleatoria positiva, dunque il suo valore atteso è sempre ben definito, eventualmente $+\infty$.) Si indica quindi con $L^p = L^p(\Omega, P)$ l'insieme delle variabili aleatorie X di norma p finita:

$$\begin{aligned} L^p(\Omega, P) &:= \{X : \Omega \rightarrow \mathbb{R} \text{ tali che } \|X\|_p < +\infty\} \\ &= \{X : \Omega \rightarrow \mathbb{R} \text{ tali che } E(|X|^p) < +\infty\}. \end{aligned}$$

Si noti che $X \in L^p$ se e solo se $|X|^p \in L^1$.

Ricordiamo che, dato uno spazio vettoriale V sul campo \mathbb{R} , si dice *seminorma* su V ogni funzione $\|\cdot\| : V \rightarrow [0, +\infty)$ che soddisfa le seguenti proprietà:

- omogeneità: $\|\lambda x\| = |\lambda| \|x\|$ per ogni $x \in V$, $\lambda \in \mathbb{R}$;
- disuguaglianza triangolare: $\|x + y\| \leq \|x\| + \|y\|$ per ogni $x, y \in V$.

Segue facilmente che $\|0\| = 0$ per ogni seminorma. Se vale inoltre che $\|x\| \neq 0$ per ogni $x \neq 0$, la seminorma è detta *norma*. Come discuteremo nell'Osservazione 3.62 più avanti, la “norma p” $\|\cdot\|_p$ è in generale solo una seminorma su $L^p(\Omega, P)$, e solo per $p \geq 1$.

Mostriamo alcune proprietà basilari degli spazi $L^p(\Omega, P)$.

Proposizione 3.60. *Sia (Ω, \mathbb{P}) uno spazio di probabilità discreto fissato.*

- Se $0 < p \leq q < \infty$, allora $L^q(\Omega, \mathbb{P}) \subseteq L^p(\Omega, \mathbb{P})$.
- Se $X, Y \in L^2(\Omega, \mathbb{P})$, allora $XY \in L^1(\Omega, \mathbb{P})$.

Dimostrazione. Per il primo punto, notiamo che vale la disuguaglianza

$$|x|^p \leq 1 + |x|^q, \quad \forall x \in \mathbb{R} \quad (3.39)$$

(è sufficiente osservare che $|x|^p \leq 1$ se $|x| \leq 1$, mentre $|x|^p \leq |x|^q$ se $|x| > 1$), pertanto $|X(\omega)|^p \leq 1 + |X(\omega)|^q$ per ogni $\omega \in \Omega$. Per la monotonia e la linearità del valor medio, vista nella Proposizione 3.55, applicate a variabili aleatorie positive, si ha

$$E(|X|^p) \leq 1 + E(|X|^q),$$

quindi se $E(|X|^q) < \infty$ anche $E(|X|^p) < \infty$, ossia $X \in L^q(\Omega, \mathbb{P}) \Rightarrow X \in L^p(\Omega, \mathbb{P})$.

Per il secondo punto, sviluppando il quadrato $(x \pm y)^2 \geq 0$, si ha la disuguaglianza

$$|xy| \leq \frac{1}{2}(x^2 + y^2),$$

valida per ogni $x, y \in \mathbb{R}$, da cui segue, per monotonia del valor medio, che

$$E(|XY|) \leq \frac{1}{2}(E(X^2) + E(Y^2)).$$

Di conseguenza, se $X, Y \in L^2(\Omega, \mathbb{P})$, si ha $XY \in L^1(\Omega, \mathbb{P})$. □

Proposizione 3.61. *Sia (Ω, \mathbb{P}) uno spazio di probabilità discreto fissato.*

- Per ogni $p \in (0, \infty)$, l'insieme $L^p(\Omega, \mathbb{P})$ è uno spazio vettoriale su \mathbb{R} , in cui lo “zero” è dato dalla funzione costantemente uguale a 0.
- Per ogni $p \in [1, \infty)$, il valor medio $E(\cdot) : L^p(\Omega, \mathbb{P}) \rightarrow \mathbb{R}$ è un operatore lineare.

Dimostrazione. Per mostrare che $L^p(\Omega, \mathbb{P})$ è uno spazio vettoriale, conviene usare la formulazione alternativa di valor medio fornita dalla Proposizione 3.54. Il fatto che per ogni $X \in L^p(\Omega, \mathbb{P})$ e $\lambda \in \mathbb{R}$ si ha $\lambda X \in L^p(\Omega, \mathbb{P})$ è immediato, e lo lasciamo verificare al lettore. Resta da dimostrare che se $X, Y \in L^p(\Omega, \mathbb{P})$ allora $X + Y \in L^p(\Omega, \mathbb{P})$. Osserviamo che, se x, y sono numeri reali, valgono le semplici disuguaglianze

$$|x + y| \leq |x| + |y| \leq \max\{2|x|, 2|y|\}$$

(per la seconda, basta considerare i due casi $|x| \geq |y|$ oppure $|x| < |y|$). Dato che $t \mapsto$

t^p è una funzione crescente per $t \geq 0$ (se $p > 0$), segue che

$$|x + y|^p \leq \max\{2^p|x|^p, 2^p|y|^p\} \leq 2^p|x|^p + 2^p|y|^p$$

(il massimo di due numeri positivi è minore della loro somma). Scegliendo $x = X(\omega)$ e $y = Y(\omega)$, per $\omega \in \Omega$ arbitrario, e ricordando la Proposizione 3.55, si ottiene

$$E(|X + Y|^p) \leq E(2^p|X|^p + 2^p|Y|^p) = 2^p E(|X|^p) + 2^p E(|Y|^p),$$

pertanto se $X, Y \in L^p(\Omega, \mathcal{P})$ allora $X + Y \in L^p(\Omega, \mathcal{P})$.

Per quanto riguarda la seconda parte dell'enunciato, il fatto che il valor medio $E(\cdot)$ è un operatore lineare su $L^1(\Omega, \mathcal{P})$ è una conseguenza immediata della Proposizione 3.55. Sappiamo inoltre che, per $p \geq 1$, $L^p(\Omega, \mathcal{P})$ è uno spazio vettoriale e $L^p(\Omega, \mathcal{P}) \subseteq L^1(\Omega, \mathcal{P})$. Ma allora $L^p(\Omega, \mathcal{P})$ è un sottospazio vettoriale di $L^1(\Omega, \mathcal{P})$, per cui $E(\cdot)$ è un operatore lineare anche su $L^p(\Omega, \mathcal{P})$ per $p \geq 1$. \square

Osservazione 3.62. Se facciamo l'ipotesi che lo spazio di probabilità discreto (Ω, \mathcal{P}) sia tale che

$$P(\{\omega\}) > 0 \text{ per ogni } \omega \in \Omega, \quad (3.40)$$

è possibile mostrare che per ogni $p \in [1, \infty)$ la funzione $\|\cdot\|_p$ definita in (3.38) è effettivamente una *norma* sullo spazio vettoriale $L^p(\Omega, \mathcal{P})$ (che rende tale spazio *completo*, cioè uno *spazio di Banach*). Inoltre, per il Corollario 3.79 che vedremo più avanti,

$$\|X\|_q \leq \|X\|_p, \quad \forall 1 \leq p \leq q < \infty,$$

dunque l'inclusione $L^q(\Omega, \mathcal{P}) \rightarrow L^p(\Omega, \mathcal{P})$ è continua rispetto alle corrispondenti norme. Se (3.40) non è verificata, la funzione $\|\cdot\|_p$ è solo una seminorma (si può avere $\|X\|_p = 0$ per $X \neq 0$).

Se invece $p \in (0, 1)$, la funzione $\|\cdot\|_p : L^p(\Omega, \mathcal{P}) \rightarrow \mathbb{R}^+$ non è una norma e nemmeno una seminorma, tranne in casi banali, perché non vale la disuguaglianza triangolare.[†] Per questa ragione, lo spazio vettoriale $L^p(\Omega, \mathcal{P})$ è studiato soprattutto per $p \geq 1$.

Osserviamo infine che, se Ω è un insieme infinito (numerabile) e vale la relazione (3.40), per ogni $p \in (0, \infty)$ lo spazio vettoriale $L^p(\Omega, \mathcal{P})$ ha dimensione infinita. \square

Definizione 3.63 (Momenti). Dato $k \in \mathbb{N} = \{1, 2, \dots\}$, si dice che una variabile aleatoria reale X *ammette momento di ordine k finito* se $E(|X|^k) < \infty$, ossia se $X \in L^k$. In questo caso, la quantità $E(X^k)$ si dice *momento di ordine k di X* .

[†] Per esempio, basta considerare uno spazio campionario $\Omega = \{0, 1\}$ che contiene due elementi, indicati con 0 e 1, con la probabilità uniforme, ossia $P(\{0\}) = P(\{1\}) = \frac{1}{2}$. Per ogni variabile aleatoria $X : \Omega \rightarrow \mathbb{R}$ si ha $\|X\|_p = E(|X|^p)^{1/p} = \frac{1}{2^{1/p}}(|X(0)|^p + |X(1)|^p)^{1/p}$. Definendo $X(0) := 1, X(1) := 0$ e $Y(0) := 0, Y(1) := 1$, si ha allora $\|X\|_p = \|Y\|_p = \frac{1}{2^{1/p}}$ mentre $\|X + Y\|_p = 1$, da cui è chiaro che la disuguaglianza triangolare $\|X + Y\|_p \leq \|X\|_p + \|Y\|_p$ non vale se $p < 1$.

Grazie alla Proposizione 3.52, è possibile stabilire se una variabile aleatoria reale X ammette momento d'ordine k finito, e nel caso calcolarlo, conoscendo la densità discreta p_X : infatti $X \in L^k$ se e solo se

$$E(|X|^k) = \sum_{x \in \mathbb{R}} |x|^k p_X(x) = \sum_{x \in X(\Omega)} |x|^k p_X(x),$$

e in questo caso

$$E(X^k) = \sum_{x \in \mathbb{R}} x^k p_X(x) = \sum_{x \in X(\Omega)} x^k p_X(x).$$

I momenti più importanti sono quello di ordine 1, che non è altro che il valor medio $E(X)$, e quello di ordine 2, ossia $E(X^2)$. Mostriamo che esistono variabili aleatorie $X \in L^1$ che non appartengono a L^2 .

Esempio 3.64. Sia X una variabile aleatoria a valori in \mathbb{N} con densità discreta

$$p_X(k) = \frac{C}{k^3}, \quad \forall k \in \mathbb{N},$$

dove la costante $C := (\sum_{k \in \mathbb{N}} \frac{1}{k^3})^{-1}$, strettamente positiva e finita (si ricordi (0.2)), è scelta in modo che p_X sia una densità discreta su \mathbb{N} . Usando la relazione (0.2),

$$E(|X|) = E(X) = \sum_{k \in \mathbb{N}} k p_X(k) = C \sum_{k \in \mathbb{N}} \frac{1}{k^2} < \infty,$$

quindi X ammette valor medio finito: $X \in L^1$. Consideriamo ora X^2 , che è una variabile aleatoria a valori in (un sottoinsieme di) \mathbb{N} . Per calcolare $E(X^2)$, applichiamo la formula (3.30) con $g(x) := x^2$: applicando ancora la relazione (0.2), si ottiene

$$E(X^2) = \sum_{x \in \mathbb{R}} g(x) p_X(x) = \sum_{k \in \mathbb{N}} k^2 p_X(k) = C \sum_{k \in \mathbb{N}} \frac{1}{k} = +\infty.$$

Dunque X^2 non ammette valor medio finito: $X \notin L^2$.

Osserviamo infine che, definendo $Y := X$, abbiamo un esempio di due variabili aleatorie X e Y che ammettono entrambe valor medio finito, ma il cui prodotto *non* ammette valor medio finito: $X \in L^1, Y \in L^1$ ma $XY \notin L^1$. \square

Diamo ora alcune definizioni molto importanti. Ricordiamo che, se X e Y sono variabili aleatorie in L^2 , allora X, Y e XY sono in L^1 , per le Proposizioni 3.60.

Definizione 3.65 (Varianza e covarianza). Sia (Ω, \mathcal{P}) uno spazio di probabilità discreto e siano X e Y variabili aleatorie reali definite su Ω .

- Se X, Y e il loro prodotto XY sono in L^1 – in particolare, se X e Y sono in L^2 – si definisce *covarianza di X e Y* la quantità

$$\begin{aligned} \text{Cov}(X, Y) &:= E((X - E(X))(Y - E(Y))) \\ &= E(XY) - E(X)E(Y) \in (-\infty, +\infty). \end{aligned} \quad (3.41)$$

Se $\text{Cov}(X, Y) = 0$, le variabili aleatorie X e Y si dicono *scorrelate*.

- Se X è in L^2 , si definisce *varianza di X* la quantità

$$\text{Var}(X) := \text{Cov}(X, X) = E((X - E(X))^2) = E(X^2) - E(X)^2 \in [0, \infty).$$

Qualche precisazione: se X, Y e $XY \in L^1(\Omega, \mathcal{P})$, allora $(X - E(X))(Y - E(Y)) \in L^1(\Omega, \mathcal{P})$, perché $L^1(\Omega, \mathcal{P})$ è uno spazio vettoriale (Proposizione 3.61), quindi la definizione (3.41) è ben posta; la seconda uguaglianza in (3.41) segue dalla linearità del valore medio (Proposizione 3.55).

Diamo ora formule esplicite per il calcolo di $\text{Cov}(X, Y)$ e $\text{Var}(X)$, in analogia con la formula (3.27) per $E(X)$. Queste sono di grande utilità negli esercizi.

Proposizione 3.66. Se $\text{Cov}(X, Y)$ è ben definita, essa è data dalla formula

$$\begin{aligned} \text{Cov}(X, Y) &= \sum_{(x,y) \in \mathbb{R}^2} (x - E(X))(y - E(Y)) p_{X,Y}(x, y) \\ &= \sum_{(x,y) \in \mathbb{R}^2} xy p_{X,Y}(x, y) - E(X)E(Y). \end{aligned} \quad (3.42)$$

Se $X \in L^2$, la varianza di X è data dalla formula

$$\text{Var}(X) = \sum_{x \in \mathbb{R}} (x - E(X))^2 p_X(x) = \sum_{x \in \mathbb{R}} x^2 p_X(x) - E(X)^2. \quad (3.43)$$

Le somme precedenti possono essere ristrette a $x \in X(\Omega)$, $y \in Y(\Omega)$.

Dimostrazione. La relazione (3.43) si ottiene applicando la formula (3.30) alla variabile aleatoria X a valori in $E = \mathbb{R}$, con $g(x) := (x - E(X))^2$. Analogamente, per la relazione (3.42) basta applicare ancora la formula (3.30), ma questa volta per la variabile aleatoria (X, Y) a valori in $E = \mathbb{R}^2$ e per la funzione $g(z) = g(x, y) := (x - E(X))(y - E(Y))$. (Si noti che i valori attesi $E(X)$ e $E(Y)$ che compaiono nelle definizioni di $g(\cdot)$ sono delle semplici costanti!) \square

Se il valor medio $E(X)$ di una variabile aleatoria X rappresenta intuitivamente il “baricentro” della distribuzione dei valori assunti da X , la varianza $\text{Var}(X)$ è connessa alla “dispersione” di tali valori attorno al valor medio. Infatti $\text{Var}(X)$ è una

media degli “scarti dal valor medio” *elevati al quadrato* $(x - E(X))^2$, pesati con le probabilità $p_X(x)$, come mostra la relazione (3.43). Di conseguenza, la radice quadrata della varianza $\sqrt{\text{Var}(X)}$, detta *deviazione standard*, fornisce una misura della *larghezza della distribuzione dei valori assunti da X attorno al valor medio*. Si noti che $\sqrt{\text{Var}(X)}$ ha la stessa “unità di misura” dei valori assunti da X .

Osservazione 3.67. Se non si elevassero gli scarti al quadrato, si avrebbe banalmente

$$E[(X - E(X))] = E(X) - E[E(X)] = E(X) - E(X) = 0, \quad (3.44)$$

per la linearità del valor medio e per il fatto che il valor medio di una costante è uguale alla costante (si ricordi l'Esempio 3.45, notando che $c = E(X)$ è una costante). Questo è intuitivo: essendo il valor medio $E(X)$ il baricentro della distribuzione, gli scarti sopra e sotto $E(X)$ hanno segno opposto e mediando si cancellano.

Per evitare questa cancellazione, una misura alternativa della dispersione, altrettanto legittima a priori, potrebbe essere $E(|X - E(X)|)$, che ha tra l'altro il vantaggio di essere definita per ogni $X \in L^1$, mentre la varianza richiede l'ipotesi più forte che $X \in L^2$. Tuttavia, la varianza è la misura di dispersione più importante e più usata, sia per ragioni pratiche (il quadrato è una funzione più regolare e maneggevole del valore assoluto) sia, soprattutto, per importanti ragioni teoriche, che hanno a che fare con il *teorema limite centrale*, che studieremo nel Capitolo 7. \square

La prossima proposizione descrive alcune proprietà basilari della covarianza. Ricordiamo dall'Esempio 3.17 che una variabile aleatoria reale X si dice *quasi certamente costante* se esiste $c \in \mathbb{R}$ tale che $P(X = c) = 1$.

Proposizione 3.68 (Proprietà della covarianza). *Sia (Ω, \mathcal{P}) uno spazio di probabilità discreto. La covarianza $\text{Cov}(\cdot, \cdot) : L^2(\Omega, \mathcal{P}) \times L^2(\Omega, \mathcal{P}) \rightarrow \mathbb{R}$ è un operatore simmetrico e bilineare, cioè per ogni $X, Y, Z \in L^2(\Omega, \mathcal{P})$ e $\alpha, \beta \in \mathbb{R}$*

$$\begin{aligned} \text{Cov}(X, Y) &= \text{Cov}(Y, X), \\ \text{Cov}(\alpha X + \beta Y, Z) &= \alpha \text{Cov}(X, Z) + \beta \text{Cov}(Y, Z). \end{aligned}$$

Inoltre, se X o Y è quasi certamente costante, si ha $\text{Cov}(X, Y) = 0$.

Dimostrazione. Le proprietà di simmetria e bilinearità seguono facilmente dalla definizione $\text{Cov}(X, Y) = E(XY) - E(X)E(Y)$ (si ricordi la linearità del valor medio).

Veniamo all'ultima affermazione. Se X è quasi certamente uguale alla costante $c \in \mathbb{R}$, allora $E(X) = c$, per l'Esempio 3.45, quindi la variabile aleatoria $(X - E(X))$ è quasi certamente uguale a zero. Di conseguenza, anche la variabile aleatoria $(X - E(X))(Y - E(Y))$ è quasi certamente uguale a zero, e dunque

$$0 = E((X - E(X))(Y - E(Y))) = \text{Cov}(X, Y),$$

ancora per l'Esempio 3.45. \square

Veniamo dunque alle proprietà fondamentali della varianza.

Proposizione 3.69 (Proprietà della varianza). Sia (Ω, \mathcal{P}) uno spazio di probabilità discreto.

(i) La varianza $\text{Var}(\cdot)$ è un operatore da $L^2(\Omega, \mathcal{P})$ in $[0, \infty)$ non lineare:

$$\text{Var}(aX + b) = a^2 \text{Var}(X), \quad \text{per ogni } X \in L^2(\Omega, \mathcal{P}) \text{ e } a, b \in \mathbb{R}.$$

(ii) Per ogni $n \in \mathbb{N}$ e $X_1, \dots, X_n \in L^2$ vale la relazione

$$\text{Var}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \text{Var}(X_i) + \sum_{1 \leq i, j \leq n: i \neq j} \text{Cov}(X_i, X_j). \quad (3.45)$$

(iii) $\text{Var}(X) = 0$ se e solo se X è quasi certamente costante, per ogni $X \in L^2$.

(iv) $\text{Var}(X) \leq E((X - c)^2)$, per ogni $X \in L^2$ e per ogni $c \in \mathbb{R}$. L'uguaglianza $\text{Var}(X) = E((X - c)^2)$ vale se e solo se $c = E(X)$.

Dimostrazione. Per la Proposizione 3.68, la covarianza è un operatore bilineare che si annulla se uno dei suoi argomenti è costante. Dato che $\text{Var}(Z) = \text{Cov}(Z, Z)$, i punti (i) e (ii) seguono facilmente.

Per il punto (iii), definiamo per comodità $Y := (X - E(X))^2$, osservando che $\text{Var}(X) = E(Y)$. Se X è quasi certamente costante, ossia $P(X = c) = 1$ con $c \in \mathbb{R}$, allora $E(X) = c$, per l'Esempio 3.45; notando che si ha l'uguaglianza di eventi

$$\{X = E(X)\} = \{Y = 0\}, \quad (3.46)$$

segue che $P(Y = 0) = 1$, quindi $E(Y) = 0$ e dunque $\text{Var}(X) = E(Y) = 0$. Viceversa, supponiamo che $\text{Var}(X) = E(Y) = 0$. Essendo Y positiva con valor medio nullo, per la Proposizione 3.58 si ha $P(Y = 0) = 1$, dunque $P(X = E(X)) = 1$ grazie a (3.46). Questo mostra che X è quasi certamente uguale alla costante $E(X)$.

Per il punto (iv), è sufficiente osservare che

$$\begin{aligned} E[(X - c)^2] &= E[(X - E(X) + E(X) - c)^2] \\ &= E[(X - E(X))^2] + (E(X) - c)^2 + 2(E(X) - c)E[X - E(X)] \\ &= \text{Var}(X) + (E(X) - c)^2, \end{aligned} \quad (3.47)$$

dove abbiamo usato il fatto che $E[X - E(X)] = 0$, come già visto in (3.44). \square

Osservazione 3.70. Il punto (iii) della Proposizione 3.69 mostra che le variabili aleatorie con varianza zero sono costanti a meno di insiemi di probabilità zero. Se si assumesse l'ipotesi (3.40), si potrebbe affermare, in modo più elegante, che $\text{Var}(X) = 0$ se e solo se X è costante. Tuttavia, l'ipotesi (3.40), sostanzialmente non restrittiva per spazi di probabilità discreti, è invece non ragionevole in spazi di probabilità generali, come vedremo. Scegliamo pertanto di non fare tale assunzione, in modo da avere enunciati che resteranno validi nell'ambito della teoria generale. \square

Osservazione 3.71. Il punto (iv) della Proposizione 3.69 fornisce una *caratterizzazione variazionale* del valor medio, affermando che esso è la costante che realizza la distanza minima da X nello spazio (semi-)normato $(L^2(\Omega, \mathcal{P}), \|\cdot\|_2)$. Si noti infatti che $E[(X - c)^2] = \|X - c\|_2^2$. \square

3.3.4 Valor medio e indipendenza

Mostriamo ora un legame fondamentale tra le nozioni di valor medio e indipendenza per variabili aleatorie.

Proposizione 3.72 (Valor medio e indipendenza). *Se X, Y sono variabili aleatorie reali indipendenti e in L^1 , allora anche il prodotto $XY \in L^1$ e si ha*

$$E(XY) = E(X)E(Y). \quad (3.48)$$

Nel caso in cui X e Y siano variabili aleatorie indipendenti e positive, la relazione (3.48) è valida anche senza l'ipotesi che $X, Y \in L^1$.

Dimostrazione. Dato che X e Y sono indipendenti, si ha che $p_{X,Y}(x, y) = p_X(x) p_Y(y)$ per ogni $(x, y) \in \mathbb{R}^2$, grazie alla Proposizione 3.32. Di conseguenza, applicando la formula (3.30) alla variabile aleatoria (X, Y) (a valori in $E = \mathbb{R}^2$) per la funzione positiva $g(z) = g(x, y) := |xy|$ e ricordando la somma a blocchi (0.10), si ottiene

$$\begin{aligned} E(|XY|) &= \sum_{x,y \in \mathbb{R}} |x||y| p_{X,Y}(x, y) = \left(\sum_{x \in \mathbb{R}} |x| p_X(x) \right) \left(\sum_{y \in \mathbb{R}} |y| p_Y(y) \right) \\ &= E(|X|) E(|Y|). \end{aligned}$$

Questo mostra che la relazione (3.48) è sempre verificata se X e Y sono positive. Inoltre, se $E(|X|) < \infty$ e $E(|Y|) < \infty$ anche $E(|XY|) < \infty$. Ripercorrendo gli stessi passaggi togliendo i valori assoluti, si mostra che $E(XY) = E(X)E(Y)$. \square

Osservazione 3.73. Sottolineiamo che in generale, se $X, Y \in L^1$, non è detto che $XY \in L^1$; in altre parole, se due variabili aleatorie ammettono valor medio finito, non è detto che anche il loro prodotto ammetta valor medio finito (se X e Y non sono indipendenti), come mostra l'Esempio 3.64. \square

Ricordando la definizione di covarianza e la formula (3.45), otteniamo immediatamente i seguenti corollari della Proposizione 3.72.

Corollario 3.74 (Indipendenza e scorrelazione). *Due variabili aleatorie $X, Y \in L^1$ indipendenti sono scorrelate, cioè $\text{Cov}(X, Y) = 0$.*

Corollario 3.75 (Varianza della somma di variabili aleatorie indipendenti).

Se $X_1, \dots, X_n \in L^2$ sono variabili aleatorie indipendenti,

$$\text{Var}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \text{Var}(X_i).$$

Dunque due variabili aleatorie reali indipendenti e in L^1 sono scorrelate. Il viceversa non è necessariamente vero, come mostra l'esempio che segue.

Esempio 3.76. Sia Z una variabile aleatoria reale con $Z(\Omega) = \{0, \pi/2, \pi\}$ e

$$p_Z(0) = p_Z(\pi/2) = p_Z(\pi) = \frac{1}{3}.$$

Le variabili aleatorie $X := \cos(Z)$ e $Y := \sin(Z)$ sono limitate, pertanto ammettono valor medio finito. Inoltre, si verifica subito che $XY \equiv 0$ e $E(X) = 0$, da cui segue immediatamente che

$$\text{Cov}(X, Y) = 0,$$

ossia X e Y sono scorrelate. Tuttavia

$$P(X = 1, Y = 1) = 0 \neq \frac{1}{9} = P(X = 1)P(Y = 1),$$

dunque X e Y non sono indipendenti. □

3.3.5 Disuguaglianze

Come in molti altri settori della matematica, le disuguaglianze giocano un ruolo fondamentale nel calcolo delle probabilità. La disuguaglianza di Chebyshev, che vedremo ora, rafforza il significato della varianza, o meglio della deviazione standard, come indice della dispersione (o “larghezza”) dei valori di una variabile aleatoria X attorno al suo valor medio: infatti la probabilità di una deviazione dal valor medio maggiore di ε si può stimare dall'alto in funzione di $(\varepsilon/\sqrt{\text{Var}(X)})$.

Teorema 3.77. Sia X una variabile aleatoria reale, definita su uno spazio di probabilità discreto (Ω, P) . Valgono allora le seguenti disuguaglianze.

(i) **(Disuguaglianza di Markov)** Se X è a valori positivi, per ogni $\varepsilon > 0$

$$P(X \geq \varepsilon) \leq \frac{E(X)}{\varepsilon}.$$

(ii) **(Disuguaglianza di Chebyshev)** Se $X \in L^2(\Omega, P)$, per ogni $\varepsilon > 0$

$$P(|X - E(X)| \geq \varepsilon) \leq \frac{\text{Var}(X)}{\varepsilon^2}.$$

Dimostrazione. (i) Per ipotesi $X(\omega) \geq 0$ per ogni $\omega \in \Omega$. Di conseguenza si ha la disuguaglianza

$$X \geq \varepsilon \mathbb{1}_{\{X \geq \varepsilon\}}.$$

Sottolineiamo che questa è una disuguaglianza tra variabili aleatorie, ossia tra funzioni definite su Ω : in altre parole, si ha $X(\omega) \geq \varepsilon \mathbb{1}_{\{X \geq \varepsilon\}}(\omega)$ per ogni $\omega \in \Omega$ (perché?). Per monotonia e linearità del valor medio si ottiene

$$E(X) \geq E(\varepsilon \mathbb{1}_{\{X \geq \varepsilon\}}) = \varepsilon E(\mathbb{1}_{\{X \geq \varepsilon\}}) = \varepsilon P(X \geq \varepsilon),$$

avendo usato il fatto che il valor medio dell'indicatrice di un evento è uguale alla probabilità dell'evento, come mostrato nell'Esempio 3.46. La tesi segue.

(ii) Si noti che vale l'uguaglianza di eventi

$$\{|X - E(X)| \geq \varepsilon\} = \{|X - E(X)|^2 \geq \varepsilon^2\},$$

ossia $\omega \in \Omega$ appartiene al membro sinistro di questa relazione se e solo se esso appartiene al membro destro (perché?). Di conseguenza, applicando la disuguaglianza di Markov alla variabile aleatoria positiva $(X - E(X))^2$,

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

ed essendo $E[(X - E(X))^2] = \text{Var}(X)$ la dimostrazione è conclusa. \square

La seguente disuguaglianza è utile quando si voglia confrontare il valor medio di una variabile aleatoria con quelli di sue opportune funzioni.

Teorema 3.78 (Disuguaglianza di Jensen). *Sia X una variabile aleatoria reale e $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ una funzione convessa. Se X e $\varphi(X)$ ammettono entrambe valor medio finito, vale la disuguaglianza*

$$\varphi(E(X)) \leq E(\varphi(X)).$$

Dimostrazione. Per un risultato classico di analisi (che il lettore potrà verificare), il fatto che φ sia convessa è equivalente ad affermare che, per ogni $x_0 \in \mathbb{R}$, esiste una retta che passa per $(x_0, \varphi(x_0))$ e che sta sempre sotto il grafico di φ . In altri termini, per ogni $x_0 \in \mathbb{R}$ esiste $\lambda(x_0) \in \mathbb{R}$ tale che per ogni $x \in \mathbb{R}$

$$\varphi(x) \geq \varphi(x_0) + \lambda(x_0)(x - x_0). \quad (3.49)$$

Posto $x = X(\omega)$ e $x_0 = E(X)$, otteniamo che per ogni $\omega \in \Omega$

$$\varphi(X(\omega)) \geq \varphi(E(X)) + \lambda(E(X))(X(\omega) - E(X)). \quad (3.50)$$

Il membro destro di questa equazione si può riscrivere come $aX(\omega) + b$, per opportune costanti $a, b \in \mathbb{R}$, pertanto ammette valor medio finito, perché per ipotesi X ammette valor medio finito. Prendendo il valor medio dei due membri, la conclusione segue per le proprietà di monotonia e linearità del valor medio. \square

Un esempio di applicazione della disuguaglianza di Jensen è il seguente confronto tra norme p -esime di una variabile aleatoria.

Corollario 3.79. *Sia X una variabile aleatoria reale, definita su uno spazio di probabilità discreto (Ω, \mathcal{P}) . Allora, per ogni $0 < p \leq q < \infty$ si ha*

$$\|X\|_p = E(|X|^p)^{1/p} \leq E(|X|^q)^{1/q} = \|X\|_q. \quad (3.51)$$

Dimostrazione. Se $E(|X|^q) = \infty$ non c'è niente da dimostrare. Se $E(|X|^p) = \infty$, per la disuguaglianza (3.39) e la monotonia del valor medio si ha $E(|X|^q) = \infty$, dunque la relazione (3.51) è verificata. Infine, se $E(|X|^q) < \infty$ e $E(|X|^p) < \infty$, la relazione (3.51) si ottiene applicando la disuguaglianza di Jensen alla variabile aleatoria positiva $|X|^p$ e alla funzione convessa $\varphi(x) = |x|^{q/p}$. \square

La prossima disuguaglianza è utile per stimare la media di un prodotto di variabili aleatorie. Ne vedremo un'applicazione importante nel prossimo paragrafo.

Teorema 3.80 (Disuguaglianza di Cauchy-Schwarz). *Siano X e Y variabili aleatorie reali in L^2 . Allora il prodotto $XY \in L^1$ e vale la disuguaglianza*

$$|E(XY)| \leq \sqrt{E(X^2)E(Y^2)}. \quad (3.52)$$

Inoltre, si ha l'uguaglianza in (3.52) se e solo se esiste $c \in \mathbb{R}$ tale che Y è quasi certamente uguale a cX (ossia $P(Y = cX) = 1$).

Dimostrazione. Abbiamo già mostrato nella Proposizione 3.60 che $XY \in L^1(\Omega, \mathcal{P})$ per ogni $X, Y \in L^2(\Omega, \mathcal{P})$. Per ottenere la disuguaglianza (3.52), non è restrittivo supporre $P(X = 0) < 1$ e $P(Y = 0) < 1$, o equivalentemente (si ricordi la Proposizione 3.58) $E(X^2) > 0$ e $E(Y^2) > 0$, perché in caso contrario la relazione è banalmente verificata ($0 = 0$). Inoltre possiamo assumere $E(XY) \geq 0$, perché in caso contrario è sufficiente rimpiazzare X con $-X$, notando che l'intero enunciato non viene modificato da tale sostituzione. Poniamo allora

$$X_* := \frac{X}{\sqrt{E(X^2)}}, \quad Y_* := \frac{Y}{\sqrt{E(Y^2)}}.$$

Dato che $(X_* - Y_*)^2 \geq 0$, per monotonia del valor medio si ha

$$0 \leq E[(X_* - Y_*)^2] = E(X_*^2) + E(Y_*^2) - 2E(X_*Y_*) = 2 - 2 \frac{E(XY)}{\sqrt{E(X^2)E(Y^2)}},$$

da cui (3.52) segue immediatamente.

Supponiamo ora che (3.52) valga come uguaglianza. Per quanto appena visto $E[(X_* - Y_*)^2] = 0$, quindi per la Proposizione 3.58 si ha $P((X_* - Y_*) = 0) = 1$, ossia

$$P(X_* = Y_*) = 1 \iff P\left(Y = \frac{\sqrt{E(Y^2)}}{\sqrt{E(X^2)}} X\right) = 1,$$

dunque $P(Y = cX) = 1$ con $c = \sqrt{E(Y^2)}/\sqrt{E(X^2)}$.

Viceversa, se $P(Y = cX) = 1$, Y è quasi certamente uguale a cX e dunque XY è quasi certamente uguale a cX^2 . Pertanto $E(XY) = cE(X^2)$ e $E(Y^2) = c^2E(X^2)$, per l'Osservazione 3.59, da cui segue che la (3.52) vale come uguaglianza. \square

Mostriamo infine una versione speciale di una disuguaglianza importante molto importante in meccanica statistica. Ne vedremo un'applicazione nel Paragrafo 4.6.

Proposizione 3.81. *Sia Y una variabile aleatoria reale e siano $g, h: \mathbb{R} \rightarrow \mathbb{R}$ funzioni tali che $g(Y) \in L^1$, $h(Y) \in L^1$ e $g(Y)h(Y) \in L^1$. Se g e h sono entrambe crescenti, o entrambe decrescenti, si ha*

$$\text{Cov}(g(Y), h(Y)) \geq 0, \quad \text{cioè} \quad E(g(Y)h(Y)) \geq E(g(Y))E(h(Y)), \quad (3.53)$$

mentre se le funzioni g e h sono una crescente e l'altra decrescente, valgono le disuguaglianze inverse, con \leq invece di \geq .

Dimostrazione. Si noti che la relazione (3.53) dipende soltanto dalla *distribuzione* della variabile aleatoria Y (perché?). Tuttavia, risulta comodo dimostrarla usando esplicitamente variabili aleatorie.

Notiamo innanzitutto che, se g e h sono entrambe crescenti, si ha

$$(g(y_1) - g(y_2))(h(y_1) - h(y_2)) \geq 0, \quad \forall y_1, y_2 \in \mathbb{R}.$$

Infatti, se $y_1 \geq y_2$ si ha $g(y_1) \geq g(y_2)$ e $h(y_1) \geq h(y_2)$, mentre se $y_1 \leq y_2$ si ha $g(y_1) \leq g(y_2)$ e $h(y_1) \leq h(y_2)$, per la monotonia di g e h . Supponiamo ora che Y_1 e Y_2 siano variabili aleatorie indipendenti, definite ovviamente sullo stesso spazio di probabilità (Ω, \mathcal{P}) , ciascuna con la stessa distribuzione di Y . Dalla relazione precedente deduciamo che $[g(Y_1(\omega)) - g(Y_2(\omega))][h(Y_1(\omega)) - h(Y_2(\omega))] \geq 0$ per ogni $\omega \in \Omega$, pertanto per monotonia del valor medio

$$E[(g(Y_1) - g(Y_2))(h(Y_1) - h(Y_2))] \geq 0.$$

Per linearità del valor medio, il membro sinistro di questa relazione vale

$$\begin{aligned} & E[g(Y_1)h(Y_1)] + E[g(Y_2)h(Y_2)] - E[g(Y_1)h(Y_2)] - E[g(Y_2)h(Y_1)] \\ &= 2E[g(Y)h(Y)] - 2E[g(Y)]E[h(Y)] = 2\text{Cov}(g(Y), h(Y)), \end{aligned}$$

dove nella prima uguaglianza abbiamo usato il fatto che Y_1 e Y_2 hanno la stessa distribuzione di Y e sono indipendenti. La dimostrazione è completa se g e h sono

crescenti, e gli altri casi si mostrano in modo analogo (oppure si nota che moltiplicando una funzione decrescente per -1 si ottiene una funzione crescente). \square

3.3.6 Coefficiente di correlazione

Cominciamo con una semplice ma importante conseguenza della disuguaglianza di Cauchy-Schwarz.

Proposizione 3.82. *Per ogni $X, Y \in L^2(\Omega, \mathcal{P})$ è ben definita $\text{Cov}(X, Y)$ e si ha*

$$|\text{Cov}(X, Y)| \leq \sqrt{\text{Var}(X) \text{Var}(Y)}. \quad (3.54)$$

In questa relazione vale l'uguaglianza se e solo se esistono costanti $a, b \in \mathbb{R}$ tali che Y è quasi certamente uguale a $(aX + b)$, ossia

$$P(Y = aX + b) = 1. \quad (3.55)$$

Se per di più $\text{Cov}(X, Y) \neq 0$ (ossia $\text{Var}(X) \neq 0$, $\text{Var}(Y) \neq 0$), la costante a in (3.55) è non nulla e ha lo stesso segno di $\text{Cov}(X, Y)$.

Dimostrazione. Abbiamo già osservato che $\text{Cov}(X, Y)$ è ben definita per ogni $X, Y \in L^2(\Omega, \mathcal{P})$, grazie alla Proposizione 3.60. Applicando il Teorema 3.80 alle variabili aleatorie $\bar{X} = X - E(X)$ e $\bar{Y} = Y - E(Y)$, si ottiene immediatamente la disuguaglianza (3.54) come conseguenza di (3.52). Inoltre, se (3.54) vale come uguaglianza, allora (3.52) vale come uguaglianza per \bar{X} e \bar{Y} , e quindi esiste $c \in \mathbb{R}$ per cui $P(\bar{Y} = c\bar{X}) = 1$, ossia

$$P(Y = cX + E(Y) - cE(X)) = 1,$$

cioè (3.55) è verificata con $a = c$ e $b = E(Y) - cE(X)$. Infine, se vale (3.55), si ha $\text{Cov}(X, Y) = a \text{Var}(X)$ per l'Osservazione 3.59. Di conseguenza, se $\text{Cov}(X, Y) \neq 0$ e $\text{Var}(X) > 0$, segue che $a \neq 0$ ha lo stesso segno di $\text{Cov}(X, Y)$. \square

Definizione 3.83 (Coefficiente di correlazione). Se $X, Y \in L^2$ sono variabili aleatorie con $\text{Var}(X) > 0$ e $\text{Var}(Y) > 0$ (ossia, per la Proposizione 3.69(iii), se X e Y non sono quasi certamente costanti), la quantità

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

è detta *coefficiente di correlazione tra X e Y* .

Si noti che se le variabili X e Y sono in L^2 e scorrelate, ossia $\text{Cov}(X, Y) = 0$, allora il coefficiente di correlazione è nullo: $\rho(X, Y) = 0$. Le seguenti proprietà del coefficiente di correlazione discendono immediatamente dalla Proposizione 3.82.

Proposizione 3.84. Per ogni $X, Y \in L^2$ tali che $\text{Var}(X) > 0$, $\text{Var}(Y) > 0$ si ha

$$|\rho(X, Y)| \leq 1.$$

Inoltre $\rho(X, Y) = 1$ (risp. $\rho(X, Y) = -1$) se e solo se esistono $a > 0$ (risp. $a < 0$) e $b \in \mathbb{R}$ tali che

$$P(Y = aX + b) = 1.$$

Il coefficiente di correlazione è un indice del grado di correlazione *lineare* tra le variabili X e Y , ossia di quanto bene Y possa essere approssimata da funzioni lineari-affini di X . Questo è chiaro quando $|\rho(X, Y)| = 1$, vista la Proposizione 3.84. Per capire il caso generale, formuliamo il seguente problema. Siano $X, Y \in L^2(\Omega, \mathcal{P})$, con $\text{Var}(X) > 0$, $\text{Var}(Y) > 0$, due variabili aleatorie fissate, e cerchiamo di determinare la funzione lineare-affine di X che meglio approssima Y . Più precisamente, cerchiamo le costanti $a, b \in \mathbb{R}$ tali che la distanza $\|Y - (aX + b)\|_2$, o equivalentemente il suo quadrato, sia la minima possibile (si ricordi (3.38)). La minimizzazione della funzione

$$\begin{aligned} \varphi(a, b) &:= \|Y - (aX + b)\|_2^2 = E[(Y - aX - b)^2] \\ &= E(X^2)a^2 + b^2 + 2E(X)ab - 2E(XY)a - 2E(Y)b + E(Y^2), \end{aligned}$$

si può ottenere facilmente con il calcolo differenziale in due variabili. Alternativamente, possiamo usare l'espressione in (3.47):

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

da cui segue che una coppia (a^*, b^*) che minimizza $\varphi(a, b)$ deve essere tale che $b^* = E(Y) + a^*E(X)$, mentre a^* è il valore di a che minimizza

$$\text{Var}(Y - aX) = a^2 \text{Var}(X) - 2a \text{Cov}(X, Y) + \text{Var}(Y).$$

Essendo quest'ultima espressione un polinomio di secondo grado in a , si trova

$$a^* = \frac{\text{Cov}(Y, X)}{\text{Var}(X)}, \quad b^* = E(Y) - \frac{E(X) \text{Cov}(X, Y)}{\text{Var}(X)}. \quad (3.56)$$

Il valore del minimo assoluto della funzione $\varphi(a, b) = \|Y - (aX + b)\|_2^2$ è pertanto

$$\varphi(a^*, b^*) = \text{Var}(Y)(1 - \rho^2(X, Y)).$$

Questo mostra che Y è tanto meglio approssimabile da funzioni lineari-affini di X quanto più $\rho^2(X, Y)$ è vicino a 1. Viceversa, se le variabili Y e X sono scorrelate, ossia $\rho(X, Y) = 0$, allora $a^* = 0$, cioè la migliore approssimazione di Y con funzioni affini di X non dipende da X , ossia è una costante. La retta $y = a^*x + b^*$, con a^* e b^* dati da (3.56) è detta *retta di regressione lineare* della variabile aleatoria Y rispetto a X .

Esercizi

Esercizio 3.7. Sia X una variabile aleatoria discreta reale a valori in \mathbb{N} , con densità discreta data da $p_X(x) = \frac{c\alpha}{x^{1+\alpha}} \mathbb{1}_{\mathbb{N}}(x)$, dove $\alpha \in (0, \infty)$ è un parametro fissato, e

$$\frac{1}{c\alpha} = \sum_{y \in \mathbb{N}} \frac{1}{y^{1+\alpha}}.$$

Si determini per quali valori di $p \in (0, \infty)$ si ha $X \in L^p(\Omega, \mathcal{P})$.

Esercizio 3.8. Siano X, Y variabili aleatorie indipendenti, a valori rispettivamente negli insiemi E e F , e siano $g: E \rightarrow \mathbb{R}$ e $h: F \rightarrow \mathbb{R}$ funzioni arbitrarie. Si mostri che se le variabili aleatorie $g(X)$ e $h(Y)$ ammettono valor medio finito, allora anche il loro prodotto $g(X)h(Y)$ ammette valor medio finito e $E(g(X)h(Y)) = E(g(X))E(h(Y))$.

Esercizio 3.9. Per $n \geq 1$, sia X_n una variabile aleatoria che assume, con la stessa probabilità, i valori $1, 2, \dots, n-1, n$. Se $f: [0, 1] \rightarrow \mathbb{R}$ è una funzione continua, sia

$$m_n = E \left[\left(\frac{X_n}{n} \right) \right].$$

Si identifichi il limite

$$\lim_{n \rightarrow +\infty} m_n.$$

Esercizio 3.10. Sia Y una variabile aleatoria a valori in $[0, +\infty)$, e $f: (0, +\infty) \rightarrow (0, +\infty)$ una funzione crescente. Si mostri che per ogni $\varepsilon > 0$

$$P(Y \geq \varepsilon) \leq \frac{E[f(Y)]}{f(\varepsilon)}.$$

Esercizio 3.11. Si dimostri la seguente versione “forte” della disuguaglianza di Jensen. Sia X una variabile aleatoria reale e $\varphi: \mathbb{R} \rightarrow \mathbb{R}$ una funzione *strettamente* convessa, nel senso che la relazione (3.49) vale come disuguaglianza stretta per ogni $x \neq x_0$. Se X e $\varphi(X)$ ammettono entrambe valor medio finito, e X *non* è quasi certamente costante, vale la disuguaglianza *stretta*

$$\varphi(E(X)) > E(\varphi(X)).$$

Esercizio 3.12. Due variabili aleatorie X e Y sono definite nel modo seguente, sulla base dei risultati di tre lanci indipendenti di una moneta equilibrata. Se al primo lancio esce testa, poniamo $X = Y = 0$. Se esce croce sia al primo che al secondo lancio, poniamo $X = 1$, mentre se viene croce al primo lancio e testa al secondo poniamo $X = -1$. Similmente, se viene croce al primo e al terzo lancio poniamo $Y = 1$, mentre se viene croce al primo lancio e testa al terzo, poniamo $Y = -1$. Si determini la distribuzione congiunta di X e Y , mostrando che sono variabili aleatorie scorrelate ma non indipendenti.

3.4 Lavorare con le distribuzioni

Per rendere “operative” le nozioni fin qui introdotte sulle variabili aleatorie, è utile studiare alcune quantità e operazioni legate alle distribuzioni.

3.4.1 Somma di variabili aleatorie

Ricordiamo, per la Definizione 1.9, che una *densità discreta* p su \mathbb{R} è una funzione $p : \mathbb{R} \rightarrow [0, 1]$ tale che $\sum_{x \in \mathbb{R}} p(x) = 1$.

Definizione 3.85 (Convoluzione di densità discrete). Siano p e q due densità discrete su \mathbb{R} . Si definisce *convoluzione* di p e q la funzione $p * q : \mathbb{R} \rightarrow \mathbb{R}$ definita da

$$(p * q)(z) := \sum_{x \in \mathbb{R}} p(z - x) q(x).$$

Proposizione 3.86. *La convoluzione di due densità discrete su \mathbb{R} è una densità discreta su \mathbb{R} . Inoltre $p * q = q * p$.*

Dimostrazione. Chiaramente $(p * q)(z) \geq 0$ per ogni $z \in \mathbb{R}$, essendo definita come somma ad addendi positivi. Inoltre, usando la proprietà (0.11) delle somme infinite,

$$\sum_{z \in \mathbb{R}} (p * q)(z) = \sum_{z \in \mathbb{R}} \left(\sum_{x \in \mathbb{R}} p(z - x) q(x) \right) = \sum_{x \in \mathbb{R}} \left[\sum_{z \in \mathbb{R}} p(z - x) \right] q(x) = \sum_{x \in \mathbb{R}} q(x) = 1,$$

dove abbiamo usato il fatto che se $(a_x)_{x \in \mathbb{R}}$ ammette somma, allora

$$\sum_{x \in \mathbb{R}} a_x = \sum_{x \in \mathbb{R}} a_{z-x}, \quad \forall z \in \mathbb{R},$$

la cui verifica è lasciata al lettore. In altre parole, per una famiglia di numeri reali indicizzati da $x \in \mathbb{R}$, il “cambio di variabile” che consiste nel rimpiazzare x con $z - x$ non modifica il valore della somma. La stessa proprietà mostra la simmetria della convoluzione $p * q = q * p$, dato che le due somme

$$(p * q)(z) := \sum_{x \in \mathbb{R}} p(z - x) q(x), \quad (q * p)(z) := \sum_{x \in \mathbb{R}} p(x) q(z - x),$$

si ottengono l’una dall’altra proprio con questo cambio di variabile. \square

Il significato probabilistico della convoluzione è chiarito dal seguente risultato, in cui mostriamo come calcolare la densità discreta della somma di due variabili aleatorie reali a partire dalla loro densità congiunta.

Proposizione 3.87 (Somma di variabili aleatorie discrete). Siano X e Y due variabili aleatorie discrete reali, definite sullo stesso spazio di probabilità, e sia $p_{X,Y}$ la loro densità discreta congiunta. Allora la densità discreta della loro somma $X + Y$ è data da

$$p_{X+Y}(z) = \sum_{x \in \mathbb{R}} p_{X,Y}(x, z-x) = \sum_{y \in \mathbb{R}} p_{X,Y}(z-y, y), \quad \forall z \in \mathbb{R}. \quad (3.57)$$

Se, inoltre, X e Y sono indipendenti, allora

$$p_{X+Y} = p_X * p_Y. \quad (3.58)$$

Dimostrazione. Usando la σ -additività della probabilità, si ha

$$\begin{aligned} P(X+Y=z) &= P\left(\bigcup_{x \in X(\Omega)} \{X=x, Y=z-x\}\right) \\ &= \sum_{x \in X(\Omega)} P(X=x, Y=z-x) = \sum_{x \in \mathbb{R}} p_{X,Y}(x, z-x), \end{aligned}$$

dove l'ultima uguaglianza è giustificata dal fatto che $P(X=x, Y=z-x) = 0$ se $x \notin X(\Omega)$. L'ultima uguaglianza in (3.57) si ottiene con il cambio di variabile $x \mapsto z-x$, già usato nella dimostrazione della Proposizione 3.86. Infine, se X e Y sono indipendenti, la relazione (3.58) segue facilmente da (3.57) e dalla Proposizione 3.32, per la quale $p_{X,Y}(x, z-x) = p_X(x) p_Y(z-x)$. \square

3.4.2 Funzione di ripartizione

Definizione 3.88 (Funzione di ripartizione). Sia X una variabile aleatoria discreta reale. Si dice *funzione di ripartizione* di X la funzione $F_X : \mathbb{R} \rightarrow [0, 1]$ definita da

$$F_X(x) := P(X \leq x), \quad \forall x \in \mathbb{R}.$$

Dimostriamo alcune proprietà generali della funzione di ripartizione. Ricordiamo che per noi una funzione $f : \mathbb{R} \rightarrow \mathbb{R}$ è crescente se $f(x) \leq f(y)$ per ogni $x \leq y$.

Proposizione 3.89 (Proprietà della funzione di ripartizione). Sia F_X la funzione di ripartizione di una variabile aleatoria reale discreta X . Allora valgono le seguenti proprietà:

- (i) F_X è crescente;
- (ii) F_X è continua da destra, ossia $F_X(x) = \lim_{y \downarrow x} F_X(y)$ per ogni $x \in \mathbb{R}$;
- (iii) $\lim_{x \rightarrow -\infty} F_X(x) = 0$;
- (iv) $\lim_{x \rightarrow +\infty} F_X(x) = 1$.

Dimostrazione. (i) Se $x \leq y$ allora $\{X \leq x\} \subseteq \{X \leq y\}$, pertanto

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

(ii) Sia $x \in \mathbb{R}$. Basta dimostrare che se $(x_n)_{n \in \mathbb{N}}$ è una successione decrescente, tale che $x_n \downarrow x$, allora

$$\lim_{n \rightarrow +\infty} F_X(x_n) = F(x). \quad (3.59)$$

Si osservi che

$$\{X \leq x\} = \bigcap_{n \in \mathbb{N}} \{X \leq x_n\}.$$

Infatti, se $X(\omega) \leq x$, si deve avere $X(\omega) \leq x_n$ per ogni $n \in \mathbb{N}$, essendo $x \leq x_n$; viceversa, se $X(\omega) \leq x_n$ per ogni $n \in \mathbb{N}$, allora $X(\omega) \leq \lim_{n \rightarrow \infty} x_n = x$. Notiamo che la successione di eventi $\{X \leq x_n\}$ è decrescente: $\{X \leq x_{n+1}\} \subseteq \{X \leq x_n\}$ per ogni $n \in \mathbb{N}$, perché $x_{n+1} \leq x_n$. Per la Proposizione 1.19 si ha dunque

$$F_X(x) = P(X \leq x) = \lim_{n \rightarrow +\infty} P(X \leq x_n) = \lim_{n \rightarrow +\infty} F_X(x_n).$$

(iii) Sia $(x_n)_{n \in \mathbb{N}}$ una successione tale che $\lim_{n \rightarrow \infty} x_n = -\infty$. Consideriamo la famiglia decrescente di eventi $\{X \leq x_n\}$, osservando che

$$\bigcap_{n \in \mathbb{N}} \{X \leq x_n\} = \emptyset,$$

perché non esiste alcun numero reale $x = X(\omega)$ tale che $x \leq x_n$ per ogni $n \in \mathbb{N}$. Applicando la Proposizione 1.19 si ottiene che

$$0 = P(\emptyset) = \lim_{n \rightarrow \infty} P(X \leq x_n) = \lim_{n \rightarrow \infty} F_X(x_n).$$

(iv) In analogia con il punto precedente, sia $(x_n)_{n \in \mathbb{N}}$ una successione tale che $\lim_{n \rightarrow \infty} x_n = +\infty$. La famiglia di eventi $\{X \leq x_n\}$ è crescente e

$$\bigcup_{n \in \mathbb{N}} \{X \leq x_n\} = \Omega,$$

pertanto $\lim_{n \rightarrow \infty} F_X(x_n) = \lim_{n \rightarrow \infty} P(X \leq x_n) = 1$ per la Proposizione 1.19. \square

Notiamo che, per le relazioni (3.5) e (3.8), possiamo scrivere

$$F_X(x) = \mu_X((-\infty, x]) = \sum_{z \leq x} p_X(z), \quad (3.60)$$

da cui è chiaro che la funzione di ripartizione F_X dipende solo dalla distribuzione μ_X (o equivalentemente dalla densità discreta p_X) della variabile X . Viceversa, mostriamo ora che la densità discreta p_X è esprimibile in termini della funzione di ripartizione (si veda la Figura 3.1). In particolare, *la funzione di ripartizione determina completamente la distribuzione di una variabile aleatoria reale discreta.*

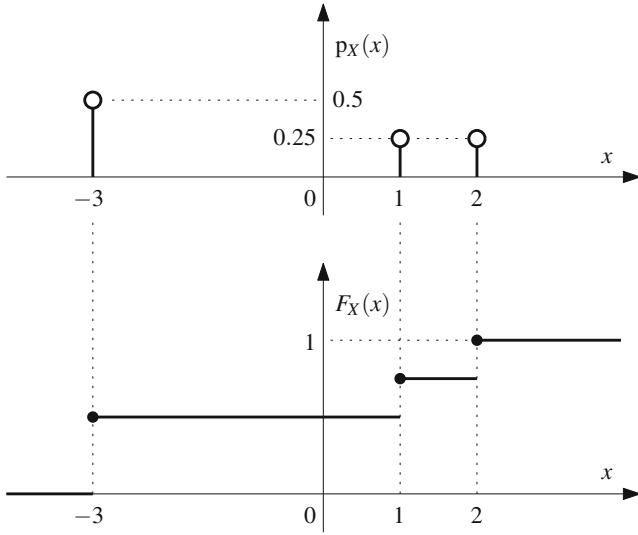


Figura 3.1 Densità discreta (sopra) e funzione di ripartizione (sotto) di una variabile aleatoria reale X , che assume i valori $\{-3, 1, 2\}$ con probabilità $P(X = -3) = \frac{1}{2}$ e $P(X = 1) = P(X = 2) = \frac{1}{4}$

Proposizione 3.90. Se X è una variabile aleatoria reale discreta, si ha

$$p_X(x) = F_X(x) - F_X(x^-), \quad \forall x \in \mathbb{R},$$

dove $F_X(x^-)$ indica il limite da sinistra della funzione F_X nel punto x :

$$F_X(x^-) := \lim_{y \uparrow x} F_X(y),$$

che esiste per monotonia di F_X .

Dimostrazione. Se $y < x$, si ha $\{X \in (y, x]\} = \{X \leq x\} \setminus \{X \leq y\}$, pertanto

$$P(X \in (y, x]) = P(X \leq x) - P(X \leq y) = F_X(x) - F_X(y).$$

Se $(y_n)_{n \in \mathbb{N}}$ è una successione di punti strettamente minori di x , tali che $y_n \uparrow x$, gli eventi $\{X \in (y_n, x]\}$ formano una successione decrescente tale che

$$\bigcap_{n \in \mathbb{N}} \{X \in (y_n, x]\} = \{X = x\}.$$

Infatti, se $X(\omega) \in (y_n, x]$ per ogni $n \in \mathbb{N}$, si deve necessariamente avere $X(\omega) = x$ (e

viceversa). Pertanto, per la Proposizione 1.19,

$$\begin{aligned} p_X(x) &= P(X = x) = \lim_{n \rightarrow +\infty} P(X \in (y_n, x]) = \lim_{n \rightarrow +\infty} [F_X(x) - F_X(y_n)] \\ &= F_X(x) - \lim_{n \rightarrow +\infty} F_X(y_n) = F_X(x) - F_X(x^-), \end{aligned}$$

e la dimostrazione è conclusa. \square

Osservazione 3.91. Segue dalla Proposizione 3.90 che F_X è discontinua in x se e solo se $p_X(x) > 0$, e il valore di $p_X(x)$ rappresenta l'entità del "salto" di F_X in x . Nel caso in cui $X(\Omega)$ non abbia punti di accumulazione, l'identità (3.60) mostra che F_X è una funzione costante a tratti. \square

Concludiamo con un risultato utile, che permette di esprimere il valore atteso $E(X)$ di una variabile aleatoria positiva X a valori interi in termini della funzione di ripartizione $F_X(n)$, o meglio di $1 - F_X(n) = P(X > n)$.

Proposizione 3.92. *Sia X una variabile aleatoria a valori in $\mathbb{N}_0 = \{0, 1, \dots\}$. Allora*

$$E(X) = \sum_{n=0}^{\infty} P(X > n). \quad (3.61)$$

Dimostrazione. Per le relazioni (3.5) e (3.8), possiamo scrivere

$$P(X > n) = \sum_{k > n} p_X(k) = \sum_{k \in \mathbb{N}_0} \mathbb{1}_{\{k > n\}} p_X(k), \quad \forall n \in \mathbb{N}_0.$$

Pertanto, usando la proprietà (0.11) delle somme infinite, si ottiene

$$\begin{aligned} \sum_{n=0}^{\infty} P(X > n) &= \sum_{n=0}^{\infty} \sum_{k=0}^{\infty} \mathbb{1}_{\{k > n\}} p_X(k) = \sum_{k=0}^{\infty} \left(\sum_{n=0}^{\infty} \mathbb{1}_{\{k > n\}} \right) p_X(k) \\ &= \sum_{k=0}^{\infty} k p_X(k) = E(X), \end{aligned}$$

avendo usato il fatto che $\sum_{n=0}^{\infty} \mathbb{1}_{\{k > n\}} = k$ (si noti che la quantità $\mathbb{1}_{\{k > n\}}$ vale 1 per $n = 0, 1, \dots, k-1$ e vale 0 per $n \geq k$). \square

3.4.3 Massimo e minimo di variabili aleatorie indipendenti

La funzione di ripartizione è spesso utile nei calcoli con le distribuzioni di variabili aleatorie, in particolare quando si tratta di determinare la distribuzione del massimo o del minimo di variabili aleatorie indipendenti.

Proposizione 3.93. *Siano X_1, X_2, \dots, X_n variabili aleatorie reali indipendenti. Definiamo:*

$$Z := \max(X_1, \dots, X_n), \quad W := \min(X_1, \dots, X_n).$$

Allora

$$F_Z(x) = \prod_{k=1}^n F_{X_k}(x), \quad F_W(x) = 1 - \prod_{k=1}^n [1 - F_{X_k}(x)]. \quad (3.62)$$

Dimostrazione. Cominciamo col dimostrare la prima relazione. Si osservi che

$$\{Z \leq x\} = \bigcap_{k=1}^n \{X_k \leq x\}.$$

Infatti $\omega \in \{Z \leq x\}$ se e solo se $\max(X_1(\omega), \dots, X_n(\omega)) \leq x$, il che è equivalente a richiedere che $X_k(\omega) \leq x$, ossia $\omega \in \{X_k \leq x\}$, per ogni $k = 1, \dots, n$. Usando l'indipendenza delle X_k , si ottiene

$$F_Z(x) = P(Z \leq x) = \prod_{k=1}^n P(X_k \leq x) = \prod_{k=1}^n F_{X_k}(x).$$

Per quanto riguarda la seconda relazione, con argomenti analoghi si mostra che

$$\{W > x\} = \bigcap_{k=1}^n \{X_k > x\}$$

perciò

$$\begin{aligned} F_W(x) &= 1 - P(W > x) = 1 - \prod_{k=1}^n P(X_k > x) = 1 - \prod_{k=1}^n [1 - P(X_k \leq x)] \\ &= 1 - \prod_{k=1}^n [1 - F_{X_k}(x)], \end{aligned}$$

come volevasi dimostrare. □

3.4.4 Funzione generatrice dei momenti

Definizione 3.94 (Funzione generatrice dei momenti). Sia X una variabile aleatoria reale. La funzione $M_X : \mathbb{R} \rightarrow \mathbb{R} \cup \{+\infty\}$ definita da

$$M_X(t) = E(e^{tX})$$

è detta *funzione generatrice dei momenti* della variabile aleatoria X .

Osservazione 3.95. Sottolineiamo che la funzione generatrice dei momenti di una variabile aleatoria può assumere il valore $+\infty$; più precisamente, $M_X(t) = +\infty$ per quei valori di $t \in \mathbb{R}$ per cui la variabile aleatoria positiva e^{tX} ha valor medio infinito. Si osservi, inoltre, che necessariamente $M_X(0) = 1$. \square

La funzione generatrice dei momenti gioca un ruolo importante nel calcolo delle probabilità, e ne vedremo applicazioni rilevanti nel Capitolo 4. Come suggerisce il suo nome, essa è spesso utile per il calcolo dei momenti di una variabile aleatoria, ove sia possibile far uso del seguente risultato.

Teorema 3.96 (Proprietà della funzione generatrice dei momenti). *Siano X una variabile aleatoria reale e M_X la sua funzione generatrice dei momenti. Supponiamo che esista $a > 0$ tale che $M_X(t) < +\infty$ per ogni $t \in (-a, a)$. Allora*

- (i) *La variabile aleatoria X ammette momenti di ogni ordine e, per ogni $t \in (-a, a)$, la funzione $M_X(t)$ è data dalla serie di Taylor*

$$M_X(t) = \sum_{n=0}^{+\infty} E(X^n) \frac{t^n}{n!}. \quad (3.63)$$

- (ii) *M_X è infinitamente derivabile in $(-a, a)$ e, indicando con $M_X^{(n)}$ la derivata n -esima di M_X , vale la relazione*

$$M_X^{(n)}(0) = E(X^n), \quad \forall n \in \mathbb{N}_0.$$

Dimostrazione. Cominciamo con alcune considerazioni preliminari. Per ogni $x \in \mathbb{R}$ vale la disuguaglianza $e^{t|x|} \leq e^{tx} + e^{-tx}$. Pertanto, per monotonia del valor medio,

$$M_{|X|}(t) = E(e^{t|X|}) \leq E(e^{tX}) + E(e^{-tX}) = M_X(t) + M_X(-t),$$

da cui segue che

$$M_{|X|}(t) < +\infty \quad \forall t \in (-a, a). \quad (3.64)$$

Ci saranno utili le seguenti disuguaglianze, valide per ogni $\rho > 0$:

$$\rho^n \leq n!e^\rho, \quad \forall n \geq 0; \quad \sum_{k=n}^{+\infty} \frac{\rho^k}{k!} \leq \frac{\rho e^\rho}{n}, \quad \forall n \geq 1. \quad (3.65)$$

La prima disuguaglianza segue dalla serie esponenziale (0.5), osservando che per ogni $n \in \mathbb{N}_0$

$$e^\rho = \sum_{k=0}^{+\infty} \frac{\rho^k}{k!} \geq \frac{\rho^n}{n!},$$

mentre la seconda deriva dalle disuguaglianze

$$n \sum_{k=n}^{+\infty} \frac{\rho^k}{k!} \leq \sum_{k=n}^{+\infty} k \frac{\rho^k}{k!} = \rho \sum_{k=n}^{+\infty} \frac{\rho^{k-1}}{(k-1)!} \leq \rho \sum_{\ell=0}^{+\infty} \frac{\rho^\ell}{\ell!} = \rho e^\rho.$$

Mostriamo quindi che

$$\mathbb{E} \left[|X|^n e^{|tX|} \right] < +\infty, \quad \forall n \geq 0, \forall t \in (-a, a). \quad (3.66)$$

Infatti, scelto $h > 0$ in modo tale che $|t| + h < a$, applicando la prima disuguaglianza in (3.65) con $\rho = h|X|$, abbiamo che

$$|X|^n e^{|tX|} \leq \frac{n!}{h^n} e^{h|X|} e^{|tX|} = \frac{n!}{h^n} e^{(|t|+h)|X|},$$

da cui segue che

$$\mathbb{E} \left[|X|^n e^{|tX|} \right] \leq \frac{n!}{h^n} \mathbb{M}_{|X|}(|t| + h) < +\infty.$$

Passiamo alla dimostrazione del punto (i). Il fatto che X ammetta momenti di ogni ordine segue subito da (3.66), con $t = 0$. Applicando la definizione di $\mathbb{M}_X(t)$, la serie esponenziale (0.5), la linearità del valor medio e la disuguaglianza triangolare, si ottiene

$$\begin{aligned} \left| \mathbb{M}_X(t) - \sum_{k=0}^n \mathbb{E}(X^k) \frac{t^k}{k!} \right| &= \left| \mathbb{E} \left[e^{tX} - \sum_{k=0}^n X^k \frac{t^k}{k!} \right] \right| \leq \mathbb{E} \left[\left| e^{tX} - \sum_{k=0}^n X^k \frac{t^k}{k!} \right| \right] \\ &= \mathbb{E} \left[\left| \sum_{k=0}^{\infty} X^k \frac{t^k}{k!} - \sum_{k=0}^n X^k \frac{t^k}{k!} \right| \right] = \mathbb{E} \left[\left| \sum_{k=n+1}^{\infty} X^k \frac{t^k}{k!} \right| \right] \leq \mathbb{E} \left[\sum_{k=n+1}^{\infty} |X|^k \frac{|t|^k}{k!} \right]. \end{aligned}$$

Usando, in quest'ultima espressione, la seconda disuguaglianza in (3.65) con $\rho = |tX|$, otteniamo

$$\left| \mathbb{M}_X(t) - \sum_{k=0}^n \mathbb{E}(X^k) \frac{t^k}{k!} \right| \leq \frac{|t|}{n+1} \mathbb{E} \left[|X| e^{|t||X|} \right]. \quad (3.67)$$

Ricordando la relazione (3.66), il membro destro in (3.67) tende a zero per $n \rightarrow \infty$, ossia abbiamo dimostrato la relazione (3.63).

Passiamo ora alla dimostrazione del punto (ii). Dimostreremo che \mathbb{M}_X è infinitamente derivabile in $(-a, a)$, e che

$$\mathbb{M}_X^{(n)}(t) = \mathbb{E}(X^n e^{tX}), \quad (3.68)$$

da cui quanto enunciato segue ponendo $t = 0$. Avendo dimostrato che \mathbb{M}_X è data dalla somma della *serie di potenze* (3.63), la relazione (3.68) segue facilmente dal teorema di derivazione per serie di potenze. Ne forniamo qui una dimostrazione alternativa, che non fa uso della teoria delle serie di potenze.

Mostriamo per induzione su $n \geq 0$ che \mathbb{M}_X ammette derivata n -esima in $(-a, a)$, e

$$\mathbb{M}_X^{(n)}(t) = \mathbb{E}(X^n e^{tX}), \quad \forall t \in (-a, a), \quad (3.69)$$

dove si noti che, per (3.66), il valor medio in (3.69) esiste. Per $n = 0$ non c'è nulla da

dimostrare, essendo, per definizione, la derivata di ordine 0 la funzione M_X stessa. Per il passo induttivo, si noti che, se $h > 0$ è sufficientemente piccolo, in modo che $t \pm h \in (-a, a)$, usando l'ipotesi induttiva (3.69)

$$\frac{M_X^{(n)}(t+h) - M_X^{(n)}(t)}{h} = E \left(X^n e^{tX} \left[\frac{e^{hX} - 1}{h} \right] \right).$$

Pertanto

$$\left| \frac{M_X^{(n)}(t+h) - M_X^{(n)}(t)}{h} - E(X^{n+1} e^{tX}) \right| \leq E \left(|X|^{n+1} \left| \frac{e^{hX} - 1}{h} - X \right| \right). \quad (3.70)$$

A questo punto usiamo il fatto che, per ogni $x \in \mathbb{R}$,

$$|e^x - 1 - x| \leq \frac{x^2}{2} e^{|x|}, \quad (3.71)$$

come segue facilmente dall'applicazione, alla funzione esponenziale, della formula di Taylor con resto di Lagrange:

$$e^x = 1 + x + \frac{x^2}{2} e^{\xi},$$

dove ξ è un opportuno numero reale contenuto nell'intervallo di estremi 0 e x . Ma allora, usando (3.71) in (3.70) si ha

$$\begin{aligned} \left| \frac{M_X^{(n)}(t+h) - M_X^{(n)}(t)}{h} - E(X^{n+1} e^{tX}) \right| &\leq \frac{h}{2} E \left(|X|^{n+2} e^{(|t|+|h|)|X|} \right) \\ &\leq \frac{h}{2} E \left(|X|^{n+2} e^{b|X|} \right), \end{aligned} \quad (3.72)$$

dove b è scelto in modo tale che $|t| + |h| < b < a$. Dato che $E(|X|^{n+2} e^{b|X|}) < +\infty$, per quanto visto sopra, otteniamo

$$\lim_{h \rightarrow 0} \left| \frac{M_X^{(n)}(t+h) - M_X^{(n)}(t)}{h} - E(X^{n+1} e^{tX}) \right| = 0.$$

Questo mostra che $M_X^{(n+1)}(t)$ esiste e vale $E(X^{n+1} e^{tX})$, quindi la dimostrazione del passo induttivo è conclusa. \square

Osservazione 3.97. Nella dimostrazione del Teorema 3.96 non abbiamo mai usato la seguente espressione per la funzione generatrice dei momenti, che è quella che più useremo nei calcoli espliciti: applicando la formula (3.30) a $g(x) = e^{tx}$, si ha

$$M_X(t) = \sum_{x \in \mathbb{R}} e^{tx} p_X(x). \quad (3.73)$$

Nel caso in cui la variabile aleatoria X assuma un numero finito di valori, cioè $|X(\Omega)| < +\infty$, la somma in (3.73) è una somma finita. In questo caso, tutte le affermazioni del Teorema 3.96 si dimostrano facilmente da (3.73). Se invece $|X(\Omega)|$ è numerabile, non c'è alcun vantaggio ad usare (3.73) per dimostrare il Teorema 3.96. La dimostrazione qui data ha il vantaggio di rimanere inalterata nel contesto più generale di variabili aleatorie definite in spazi di probabilità non discreti. \square

Esercizi

Esercizio 3.13 (Generalizzazione dell'Esempio 3.25). Siano X e Y due variabili aleatorie indipendenti, entrambe a valori in $\{1, 2, \dots, n\}$, tali che $P(X = k) = P(Y = k) = \frac{1}{n}$ per ogni $k = 1, 2, \dots, n$. Si calcoli la densità di $X + Y$.

Esercizio 3.14. Sia (X, Y) una variabile aleatoria a valori in \mathbb{R}^2 . Definiamo

$$F_{X,Y}(x, y) := P(X \leq x, Y \leq y).$$

Si mostri che per ogni $(x, y) \in \mathbb{R}^2$,

$$\begin{aligned} p_{X,Y}(x, y) \\ = \lim_{n \rightarrow +\infty} \left[F_{X,Y}(x, y) + F_{X,Y}\left(x - \frac{1}{n}, y - \frac{1}{n}\right) - F_{X,Y}\left(x - \frac{1}{n}, y\right) - F_{X,Y}\left(x, y - \frac{1}{n}\right) \right]. \end{aligned}$$

Esercizio 3.15. Sia X una variabile aleatoria discreta reale a valori in $\mathbb{Z} \setminus \{0\}$, con densità discreta data da $p_X(x) = \frac{c\alpha}{|x|^{1+\alpha}} \mathbb{1}_{\mathbb{N}}(|x|)$, dove $\alpha \in (0, \infty)$ è un parametro fissato, e

$$\frac{1}{c\alpha} = \sum_{y \in \mathbb{Z} \setminus \{0\}} \frac{1}{|y|^{1+\alpha}}.$$

Si mostri che la funzione generatrice M_X di X è data da $M_X(0) = 1$ e $M_X(t) = +\infty$ per ogni $t \neq 0$. In particolare, si noti che M_X non dipende dal valore di α .

Esercizio 3.16. Siano X_1, X_2, \dots, X_n variabili aleatorie reali indipendenti, e siano M_{X_i} le relative funzioni generatrici. Si mostri che

$$M_{X_1+X_2+\dots+X_n}(t) = \prod_{i=1}^n M_{X_i}(t),$$

per ogni $t \in \mathbb{R}$ tale che $M_{X_i}(t) < +\infty$ per ogni $i = 1, 2, \dots, n$,

3.5 Classi notevoli di variabili aleatorie discrete

Abbiamo visto come molte proprietà di una variabile aleatoria reale X dipendano esclusivamente dalla *distribuzione* μ_X di X , o equivalentemente dalla sua densità discreta p_X , e non dal particolare spazio di probabilità (Ω, \mathcal{P}) su cui X è definita (si pensi ad esempio al valor medio, alla funzione di ripartizione, alla funzione generatrice dei momenti, ...). Risulta pertanto conveniente “classificare” le variabili aleatorie secondo la loro distribuzione.

In questo paragrafo ci apprestiamo ad analizzare in dettaglio le proprietà di alcune variabili aleatorie “notevoli” – o, più precisamente, delle loro distribuzioni – che rivestono grande importanza, tanto per le loro proprietà intrinseche quanto per la frequenza e la rilevanza con cui esse ricorrono negli esempi. La maggior parte dei calcoli che effettueremo sarà basata sull’uso della formula (3.30).

Con un leggero abuso di notazione, spesso indicheremo con $X(\Omega)$ l’insieme dei valori assunti che una variabile aleatoria X assume *con probabilità strettamente positiva*, come spiegato nell’Osservazione 3.16.

3.5.1 Uniforme discreta

Sia E un arbitrario insieme finito. Una variabile aleatoria X che assume con ugual probabilità ogni valore $x \in E$ si dice *variabile aleatoria uniforme discreta a valori in E* , e scriveremo $X \sim \text{Unif}(E)$. Tale variabile aleatoria ha densità

$$p_X(x) = \frac{1}{|E|}, \quad \forall x \in E,$$

e pertanto la sua distribuzione è data da

$$\mu_X(A) = \sum_{x \in A} p_X(x) = \frac{|A|}{|E|}, \quad \forall A \subseteq E.$$

In altre parole, $X \sim \text{Unif}(E)$ se e solo se la distribuzione μ_X di X è la probabilità uniforme su E , introdotta nell’Esempio 1.13.

Un caso tipico è quello in cui $E = \{1, 2, \dots, n\}$, con $n \in \mathbb{N}$. Dato che $E \subseteq \mathbb{R}$, ha senso calcolare i momenti di X . Dato che

$$p_X(k) = p_{\text{Unif}\{1, \dots, n\}}(k) := \frac{1}{n} \mathbb{1}_{\{1, \dots, n\}}(k),$$

applicando la formula (3.30) si ottiene

$$\begin{aligned} E(X) &= \sum_{x \in \mathbb{R}} x p_X(x) = \frac{1}{n} \sum_{k=1}^n k = \frac{n+1}{2}, \\ E(X^2) &= \sum_{x \in \mathbb{R}} x^2 p_X(x) = \frac{1}{n} \sum_{k=1}^n k^2 = \frac{(n+1)(2n+1)}{3}, \end{aligned}$$

dove abbiamo usato le identità (0.7). Anche la funzione generatrice dei momenti si

calcola facilmente: applicando ancora la formula (3.30) e ricordando (0.3),

$$M_X(t) = E(e^{tX}) = \sum_{x \in \mathbb{R}} e^{tx} p_X(x) = \frac{1}{n} \sum_{k=1}^n e^{tk} = \frac{e^t}{n} \frac{e^{tn} - 1}{e^t - 1}, \quad \forall t \neq 0,$$

mentre ovviamente $M_X(t) = 1$ per $t = 0$.

I problemi di estrazione di palline da un'urna, studiati nel Paragrafo 1.2.6 usando il formalismo degli spazi di probabilità, possono essere equivalentemente (e più intuitivamente) riformulati in termini di variabili aleatorie uniformi, come mostriamo ora, generalizzando gli Esempi 3.26 e 3.27.

Esempio 3.98 (Estrazioni con reimmissione). Un'urna contiene N palline, numerate da 1 a N . Estraiamo, una dopo l'altra e con reimmissione, n palline dall'urna e indichiamo con U_1, U_2, \dots, U_n i numeri delle palline estratte.

Chiaramente $U_1 \sim \text{Unif}\{1, \dots, N\}$. Inoltre, per ogni $1 \leq \ell \leq n-1$,

$$P(U_{\ell+1} = k_{\ell+1} | U_1 = k_1, \dots, U_\ell = k_\ell) = \frac{1}{N}, \quad (3.74)$$

per ogni scelta di $k_1, \dots, k_{\ell+1} \in \{1, \dots, N\}$. Infatti, quali che siano i risultati delle prime ℓ estrazioni, la $(\ell+1)$ -esima estrazione avviene da un'urna che contiene N palline, una sola delle quali è la numero $k_{\ell+1}$. Dalla relazione (3.74), applicando la Proposizione 1.45, si ricava la densità congiunta:

$$p_{U_1, \dots, U_n}(k_1, \dots, k_n) = \frac{1}{N} \cdots \frac{1}{N} = \frac{1}{N^n}, \quad \forall (k_1, \dots, k_n) \in \{1, \dots, N\}^n. \quad (3.75)$$

Il vettore aleatorio $U := (U_1, \dots, U_n)$ ha dunque distribuzione uniforme sull'insieme $\{1, \dots, N\}^n$ delle disposizioni con ripetizione di n elementi estratti da $\{1, \dots, N\}$, coerentemente con l'analisi svolta a suo tempo nell'Esempio 1.38.

Dalla relazione (3.75) si deduce che $p_{U_i}(k_i) = \frac{1}{N}$; quindi, ricordando (3.21), le variabili aleatorie U_1, \dots, U_n sono indipendenti. In definitiva, in un'estrazione con reimmissione, *i numeri U_1, \dots, U_n delle palline estratte sono variabili aleatorie indipendenti, con distribuzione uniforme nell'insieme $\{1, \dots, N\}$.*

Un discorso del tutto analogo vale per l'inserimento casuale di n oggetti in r cassetti, che abbiamo discusso nell'Esempio 1.41. Numerando gli oggetti da 1 a n e i cassetti da 1 a r , e indicando con U_i il cassetto in cui viene inserito l'oggetto i -esimo, *le variabili aleatorie U_1, \dots, U_n sono indipendenti e hanno la stessa distribuzione, uniforme nell'insieme $\{1, \dots, r\}$.* \square

Esempio 3.99 (Estrazioni senza reimmissione). Un'urna contiene $N \in \mathbb{N}$ palline, numerate da 1 a N . Estraiamo, una dopo l'altra e senza reimmissione, tutte le palline dall'urna e indichiamo con U_i il numero dell' i -esima pallina estratta.

Ciascuna variabile aleatoria U_i assume valori in $\{1, \dots, N\}$ e, per costruzione, si ha $U_i \neq U_j$ per $i \neq j$. Quindi il vettore aleatorio $U := (U_1, \dots, U_N)$ assume valori nell'insieme $S_N \subseteq \{1, \dots, N\}^N$ delle *permutazioni* di $\{1, \dots, N\}$.

Chiaramente $U_1 \sim \text{Unif}\{1, \dots, N\}$. Osserviamo quindi che, per $1 \leq \ell \leq N-1$,

$$P(U_{\ell+1} = k_{\ell+1} | U_1 = k_1, \dots, U_\ell = k_\ell) = \frac{1}{N-\ell},$$

per ogni scelta di valori *distinti* $k_1, \dots, k_{\ell+1} \in \{1, \dots, N\}$. Infatti, estratte le prime ℓ palline, restano $N - \ell$ palline nell'urna, una sola delle quali è la numero $k_{\ell+1}$. Da ciò si deduce facilmente la distribuzione congiunta delle variabili U_1, \dots, U_N : ricordando la Proposizione 1.45, se $k_i \neq k_j$ per $i \neq j$, si ha

$$\begin{aligned} P(U_1 = k_1, \dots, U_N = k_N) &= P(U_1 = k_1) \prod_{\ell=1}^{N-1} P(U_{\ell+1} = k_{\ell+1} | U_1 = k_1, \dots, U_\ell = k_\ell) \\ &= \frac{1}{N} \prod_{\ell=1}^{N-1} \frac{1}{N-\ell} = \frac{1}{N!}. \end{aligned}$$

In definitiva, dato che $S_N = \{(k_1, \dots, k_N) \in \{1, \dots, N\}^N : k_i \neq k_j \text{ per } i \neq j\}$,

$$p_{U_1, \dots, U_N}(k_1, \dots, k_N) = \frac{1}{N!} \mathbb{1}_{S_N}(k_1, \dots, k_N). \quad (3.76)$$

Ricordando che $|S_N| = N!$, questo mostra che il vettore aleatorio $U := (U_1, \dots, U_N)$ ha distribuzione uniforme sull'insieme S_N .

Dal fatto che la densità congiunta $p_{U_1, \dots, U_N}(k_1, \dots, k_N)$ è una funzione invariante per permutazioni di k_1, \dots, k_N , segue che le variabili U_1, \dots, U_N hanno la stessa distribuzione marginale (perché?), dunque $U_i \sim U_1 \sim \text{Unif}\{1, \dots, N\}$, per ogni $1 \leq i \leq N$. In alternativa, con un calcolo diretto a partire dalla relazione (3.76),

$$p_{U_i}(k_i) = \sum_{k_1, \dots, k_{i-1}, k_{i+1}, \dots, k_N \in \{1, \dots, N\}} p_{U_1, \dots, U_N}(k_1, \dots, k_N) = \frac{1}{N!} (N-1)! = \frac{1}{N}.$$

Questo mostra che le variabili aleatorie U_1, \dots, U_N *non* sono indipendenti, poiché la loro densità congiunta (3.76) non è data dal prodotto delle densità marginali. \square

Osservazione 3.100. L'Esempio 3.99 fornisce un modo iterativo per generare una permutazione aleatoria $\sigma = (\sigma_1, \dots, \sigma_N)$ con distribuzione uniforme in S_N : si sceglie il primo elemento $\sigma_1 = U_1$ uniformemente in $\{1, \dots, N\}$, quindi si sceglie il secondo elemento $\sigma_2 = U_2$ uniformemente in $\{1, \dots, N\} \setminus \{\sigma_1\}$, ecc. \square

3.5.2 Bernoulli

Una variabile aleatoria X è detta *di Bernoulli* se $X(\Omega) = \{0, 1\}$. La distribuzione di X è dunque completamente determinata dal parametro $p := p_X(1) \in [0, 1]$, dal momento che $p_X(0) = 1 - p$ e $p_X(x) = 0$ se $x \notin \{0, 1\}$. Diremo in questo caso che X è una variabile aleatoria di Bernoulli di parametro p , scrivendo $X \sim \text{Be}(p)$.

In definitiva, $X \sim \text{Be}(p)$ se e solo se la densità discreta di X è data da

$$p_X(k) = p_{\text{Be}(p)}(k) := (1-p)\mathbb{1}_{\{0\}}(k) + p\mathbb{1}_{\{1\}}(k) = \begin{cases} 1-p & \text{se } k=0 \\ p & \text{se } k=1 \\ 0 & \text{se } k \notin \{0,1\} \end{cases}.$$

Dato che $X^2 = X$ quasi certamente, è immediato dedurre che

$$E(X) = E(X^2) = p,$$

e dunque

$$\text{Var}(X) = E(X^2) - (E(X))^2 = p(1-p).$$

Anche la funzione generatrice dei momenti si calcola facilmente:

$$M_X(t) := E[e^{tX}] = e^t p + (1-p), \quad \forall t \in \mathbb{R}.$$

A dispetto dell'apparenza semplice, le variabili di Bernoulli sono l'ingrediente fondamentale di molti modelli probabilistici, anche piuttosto complessi. Ad esempio, n prove ripetute e indipendenti con probabilità di successo p sono descritte in modo naturale da n variabili aleatorie indipendenti X_1, \dots, X_n con distribuzioni marginali $X_i \sim \text{Be}(p)$, come abbiamo mostrato nel Paragrafo 3.2.3.

Osservazione 3.101. Ogni variabile aleatoria X di Bernoulli è quasi certamente uguale a una variabile aleatoria indicatrice. Infatti, se definiamo l'evento $C := \{X = 1\}$, si ha che X è quasi certamente uguale a $\mathbb{1}_C$:

$$X(\omega) = \mathbb{1}_C(\omega) \quad \text{per ogni } \omega \in \Omega \text{ per cui } X(\omega) \in \{0,1\}. \quad (3.77)$$

Ritroviamo in questo modo i risultati visti negli Esempi 3.18 e 3.46. \square

3.5.3 Binomiale

Dati $n \in \mathbb{N}$ e $p \in [0,1]$, diremo che una variabile aleatoria X è *binomiale di parametri n e p* , e scriveremo $X \sim \text{Bin}(n, p)$, se X ha densità discreta data da

$$p_X(k) = p_{\text{Bin}(n,p)}(k) := \binom{n}{k} p^k (1-p)^{n-k} \mathbb{1}_{\{0,\dots,n\}}(k) \quad (3.78)$$

(si veda la Figura 3.2). Si noti che $X(\Omega) = \{0,1,\dots,n\}$. Nel caso $n = 1$ ritroviamo le variabili di Bernoulli: $\text{Bin}(1, p) = \text{Be}(p)$.

L'espressione nel membro destro di (3.78) è familiare: come abbiamo mostrato nel Paragrafo 1.3.4, essa coincide con la probabilità che in n prove ripetute e indipendenti con probabilità di successo p si abbiano esattamente k successi (si ricordino le formule (1.53) e (1.56)). Questa osservazione è alla base della seguente fondamentale caratterizzazione.

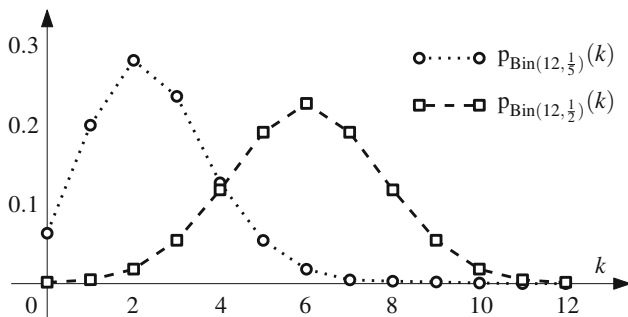


Figura 3.2 Densità discreta di una variabile aleatoria $\text{Bin}(n, p)$ con $n = 12$, per $p = \frac{1}{5}$ e $p = \frac{1}{2}$

Proposizione 3.102 (Somma di Bernoulli indipendenti). Siano X_1, \dots, X_n variabili aleatorie $\text{Be}(p)$ indipendenti. Allora $X_1 + \dots + X_n \sim \text{Bin}(n, p)$.

Dimostrazione. Sia (Ω, P) lo spazio di probabilità su cui sono definite le variabili aleatorie X_1, \dots, X_n . Ponendo $C_i := \{X_i = 1\} \subseteq \Omega$ e ricordando il Paragrafo 3.2.3, gli eventi C_1, \dots, C_n sono indipendenti e hanno la stessa probabilità $P(C_i) = p$, ossia sono n prove ripetute e indipendenti con probabilità di successo p .

Abbiamo già osservato in (3.77) che $X_i = \mathbb{1}_{C_i}$ quasi certamente. Di conseguenza, la somma $X_1 + \dots + X_n$ coincide quasi certamente con $S := \mathbb{1}_{C_1} + \dots + \mathbb{1}_{C_n}$, il numero di successi nelle n prove (si ricordi l'Esempio 3.4). Sappiamo che $S \sim \text{Bin}(n, p)$ (si veda la relazione (3.14)). Per la Proposizione 3.14, le variabili aleatorie $X_1 + \dots + X_n$ e S hanno la stessa densità discreta, quindi anche $X_1 + \dots + X_n \sim \text{Bin}(n, p)$. \square

Per calcolare media, varianza e funzione generatrice dei momenti di una variabile binomiale, è istruttivo sfruttare la Proposizione 3.102. Siano infatti X_1, \dots, X_n variabili aleatorie indipendenti con $X_i \sim \text{Be}(p)$. Allora $X := X_1 + \dots + X_n \sim \text{Bin}(n, p)$ e, per la linearità del valore atteso,

$$E(X) = E(X_1 + \dots + X_n) = np.$$

Analogamente, dato che le variabili X_i sono indipendenti, usando il Corollario 3.75

$$\text{Var}(X) = \sum_{i=1}^n \text{Var}(X_i) = np(1-p).$$

Infine, per la funzione generatrice dei momenti, ricordando l'Esercizio 3.16,

$$M_X(t) = E(e^{tX}) = (E(e^{tX_1}))^n = (M_{X_1}(t))^n = (pe^t + (1-p))^n.$$

In alternativa, sono possibili calcoli diretti: usando l'espressione (3.78) della densità discreta e la definizione (3.27) di valor medio, si ottiene

$$\begin{aligned} E(X) &= \sum_{k=0}^n k p_X(k) = \sum_{k=0}^n k \binom{n}{k} p^k (1-p)^{n-k} = np \sum_{k=1}^n \binom{n-1}{k-1} p^{k-1} (1-p)^{n-k} \\ &= np \sum_{\ell=0}^{n-1} \binom{n-1}{\ell} p^{\ell} (1-p)^{(n-1)-\ell} = np \sum_{\ell=0}^{n-1} \text{PBin}(n-1, p)(\ell) = np. \end{aligned}$$

Si possono calcolare analogamente $\text{Var}(X)$ e $M_X(t)$, applicando la formula (3.30) (si vedano gli Esercizi 3.17 e 3.18).

Una proprietà rilevante delle variabili aleatorie binomiali è la seguente.

Proposizione 3.103 (Somma di binomiali indipendenti). *Siano $X \sim \text{Bin}(n, p)$ e $Y \sim \text{Bin}(m, p)$ variabili aleatorie indipendenti. Allora $X + Y \sim \text{Bin}(n + m, p)$.*

Dimostrazione. Siano $X_1, X_2, \dots, X_n, X_{n+1}, \dots, X_{n+m}$ variabili aleatorie indipendenti, con $X_i \sim \text{Be}(p)$, definite su un opportuno spazio di probabilità discreto.[†] Le variabili aleatorie $X := X_1 + \dots + X_n$ e $Y := X_{n+1} + \dots + X_{n+m}$ sono indipendenti, per il Corollario 3.41, con $X \sim \text{Bin}(n, p)$ e $Y \sim \text{Bin}(m, p)$, per la Proposizione 3.102. D'altro canto $X + Y = X_1 + \dots + X_n + X_{n+1} + \dots + X_{n+m} \sim \text{Bin}(n + m, p)$, sempre per la Proposizione 3.102. Abbiamo dunque dimostrato il teorema per una scelta speciale delle variabili aleatorie X e Y .

Siano ora X', Y' generiche variabili aleatorie indipendenti con $X' \sim \text{Bin}(n, p)$ e $Y' \sim \text{Bin}(m, p)$. I vettori aleatori bidimensionali (X, Y) e (X', Y') hanno la stessa distribuzione congiunta, grazie alla Proposizione 3.32, perché hanno entrambi componenti indipendenti con le stesse distribuzioni marginali. Di conseguenza, per ogni funzione $f: \mathbb{R}^2 \rightarrow \mathbb{R}$, le variabili aleatorie reali $f((X, Y))$ e $f((X', Y'))$ hanno la stessa distribuzione, per la Proposizione 3.15. Scegliendo la funzione $f((x, y)) := x + y$, otteniamo che $S' := X' + Y'$ ha la stessa distribuzione di $S := X + Y$. Avendo già mostrato che $S \sim \text{Bin}(n + m, p)$, segue che anche $S' \sim \text{Bin}(n + m, p)$. \square

Le variabili aleatorie binomiali compaiono in modo naturale in molti modelli basati su uno schema di prove ripetute e indipendenti, come nei problemi di “estrazioni con reimmissione di palline da un’urna” e di “inserimento di oggetti in cassette”, descritti a suo tempo negli Esempi 1.38 e 1.41 e rivisitati nell'Esempio 3.98. Riformuliamo i calcoli fatti allora con il linguaggio delle variabili aleatorie.

Esempio 3.104. Consideriamo un’urna contenente N palline, di cui M rosse e $N - M$ verdi. Effettuiamo n estrazioni con reimmissione e indichiamo con X il numero di palline rosse estratte. Allora $X \sim \text{Bin}(n, \frac{M}{N})$, come abbiamo mostrato nella relazione (1.30). In effetti, le estrazioni costituiscono n prove ripetute e indipendenti, in cui la probabilità di successo (=estrarre una pallina rossa) vale $p = \frac{M}{N}$. \square

[†] Ad esempio, sullo spazio di probabilità canonico $(\Omega_{n+m}, \mathcal{P}_{n+m, p})$, costruito nel Paragrafo 1.3.4.

Esempio 3.105. Ci sono n oggetti distinti, ciascuno dei quali viene inserito casualmente in uno di r cassette disponibili. Sia X la variabile aleatoria che conta il numero di oggetti che finiscono in un cassetto fissato (ad esempio, il primo). Allora $X \sim \text{Bin}(n, \frac{1}{r})$, come abbiamo mostrato in (1.34). In effetti, ciascun oggetto ha, indipendentemente dagli altri, una probabilità pari a $1/r$ di finire nel primo cassetto, e gli inserimenti degli n oggetti costituiscono n prove ripetute e indipendenti. \square

Discutiamo infine un esempio dalle conclusioni a prima vista sorprendenti.

Esempio 3.106. Supponiamo di versare un bicchiere d'acqua nel mare e, un mese dopo, di ripescare un bicchiere d'acqua dal mare. Delle molecole d'acqua versate, ne ritrovo un numero aleatorio X nel bicchiere che ripesco. Quanto vale $E(X)$?

Indichiamo con $M \in \mathbb{N}$ il numero di molecole d'acqua contenute in un bicchiere e con B il numero di “bicchieri d'acqua contenuti nel mare”, ossia il rapporto tra il volume degli oceani e il volume di un bicchiere. Assumendo che le molecole versate nel mare si sparpaglino uniformemente, possiamo considerare il mare come un'urna contenente $N := M \cdot B$ molecole d'acqua, di cui M , quelle versate, sono “rosse”. Ripescare un bicchiere d'acqua dal mare corrisponde a estrarre $n := M$ palline senza reimmissione da questa urna. Trattandosi di un'urna numerosa, possiamo supporre che l'estrazione avvenga con reimmissione, in quanto le differenze sono trascurabili (si ricordi l'Osservazione 1.40). La probabilità di estrarre una pallina rossa in una singola estrazione vale $p = M/N = 1/B$. Di conseguenza la variabile causale X , che conta il numero di palline rosse estratte dall'urna, ha distribuzione approssimativamente $\text{Bin}(n, p) = \text{Bin}(M, 1/B)$ e quindi $E(X) \simeq np = M/B$.

Come vedremo nel Paragrafo 3.5.4, la distribuzione esatta di X , quando le estrazioni avvengano senza reimmissione, è detta ipergeometrica e il suo valore atteso vale *esattamente* $E(X) = np$. Non abbiamo dunque commesso errori in questa approssimazione.

Restano solo da stimare i valori di M e B . Cominciamo da M . Una molecola d'acqua è composta da due atomi di idrogeno, ciascuno dei quali ha un solo protone e nessun neutrone, e da un atomo di ossigeno, che ha 8 protoni e 8 neutroni. Pertanto il peso molecolare dell'acqua è 18, ossia una mole di acqua pesa 18 grammi, e una mole contiene per definizione un numero di Avogadro di molecole, cioè all'incirca $6 \cdot 10^{23}$. Un bicchiere d'acqua di capienza (tipica) 200 millilitri contiene 200 grammi d'acqua. In definitiva, otteniamo $M \simeq (200/18) \cdot 6 \cdot 10^{23} \simeq 6 \cdot 10^{24}$.

Determiniamo ora B . Il raggio terrestre R è lungo circa 6000 chilometri. Gli oceani coprono un'area A pari all'incirca ai due terzi della superficie terrestre, cioè $A \simeq 2/3 \cdot 4\pi R^2 \simeq 300 \cdot (1000\text{km})^2 \simeq 3 \cdot 10^8 \text{km}^2$. Assumendo una profondità media di 5 chilometri, il volume degli oceani è dunque pari a $1.5 \cdot 10^9 \text{km}^3 = 1.5 \cdot 10^{24} \text{cm}^3$. Dato che il volume di un bicchiere è di 200cm^3 (= millilitri), gli oceani contengono all'incirca $B \simeq 1.5 \cdot 10^{24}/200 = 7.5 \cdot 10^{21}$ bicchieri d'acqua.

Mettendo insieme i vari pezzi, otteniamo che il bicchiere ripescato contiene in media $E(X) = M/B \simeq 6 \cdot 10^{24}/(7.5 \cdot 10^{21}) \simeq 800$ molecole di quelle originariamente versate, un risultato che può apparire sorprendente. Come esercizio, si stimi la probabilità $P(X = 0)$ di non ripescare nessuna delle molecole originariamente versata, mostrando che è irrisoriamente piccola. \square

3.5.4 Ipergeometrica

Nell'Esempio 3.104 abbiamo visto che, in uno schema di estrazioni con reimmissione di palline da un'urna, il numero di palline estratte di un determinato colore ha distribuzione binomiale. Trattiamo ora il caso caso di estrazioni *senza* reimmissione.

Dati $n, N \in \mathbb{N}$ e $M \in \mathbb{N}_0$ con $M \leq N$ e $n \leq N$, diremo che una variabile aleatoria X è *ipergeometrica con parametri* n, N, M , e scriveremo $X \sim \text{Iper}(n, N, M)$, se X ha la stessa distribuzione del numero di palline rosse estratte in uno schema di n estrazioni senza reimmissione da un'urna che contiene N palline, di cui M rosse. Per quanto visto nell'Esempio 1.39, si ha $X \sim \text{Iper}(n, N, M)$ se e solo se la densità discreta di X è data da

$$p_X(k) = p_{\text{Iper}(n, N, M)}(k) := \frac{\binom{M}{k} \binom{N-M}{n-k}}{\binom{N}{n}} \mathbb{1}_{\{0, \dots, n\}}(k),$$

dove poniamo per comodità $\binom{a}{b} := 0$ se $a \notin \{0, \dots, b\}$. Otteniamo, in particolare, che $X(\Omega) = \{\max\{0, n - (N - M)\}, \dots, \min\{n, M\}\} \subseteq \{0, \dots, n\}$.

Come abbiamo già notato nell'Osservazione 1.40, nel caso di un'urna popolosa c'è poca differenza tra gli schemi di estrazioni con e senza reimmissione. Possiamo ora riformulare la relazione (1.33) mostrata a suo tempo in termini della convergenza della densità discreta della distribuzione ipergeometrica verso la densità discreta della distribuzione binomiale.

Proposizione 3.107 (Binomiale come limite dell'ipergeometrica). *Data una successione $(M_N)_{N \in \mathbb{N}}$ a valori in \mathbb{N}_0 tale che*

$$\lim_{N \rightarrow \infty} \frac{M_N}{N} = p \in (0, 1),$$

per ogni $n \in \mathbb{N}$ e $k \in \{0, \dots, n\}$ si ha

$$\lim_{N \rightarrow \infty} p_{\text{Iper}(n, N, M_N)}(k) = p_{\text{Bin}(n, p)}(k).$$

Tipicamente i calcoli diretti con la densità discreta della distribuzione ipergeometrica non sono molto agevoli. Un utile strumento alternativo consiste nell'esprimere una variabile ipergeometrica come funzione di variabili di Bernoulli. Consideriamo infatti uno schema di n estrazioni senza reimmissione da un'urna che contiene N palline, di cui M rosse, e indichiamo con X_1, \dots, X_n le variabili aleatorie definite da $X_i := 1$ se viene pescata una pallina rossa all' i -esima estrazione, e $X_i := 0$ altrimenti. Allora la variabile aleatoria $X := X_1 + \dots + X_n$ rappresenta il numero di palline rosse estratte, dunque $X \sim \text{Iper}(n, N, M)$.

Le X_i sono variabili di Bernoulli (non indipendenti): $X_i \sim \text{Be}(p_i)$. Il fatto interessante è che la quantità p_i , ossia la probabilità di estrarre una pallina rossa all' i -esima estrazione, vale M/N e, dunque, *non dipende da i* . Per convincersene, si ricordi l'Esempio 3.99: numerando le palline da 1 a N e indicando con U_i il numero dell' i -esima pallina estratta, si ha $U_i \sim \text{Unif}\{1, \dots, N\}$ per ogni i . Possiamo supporre che le palline

rosse siano quelle numerate da 1 a M , pertanto $p_i = P(U_i \in \{1, \dots, M\}) = M/N$. Di conseguenza

$$E(X) = \sum_{i=1}^n E(X_i) = \sum_{i=1}^n p_i = np = n \frac{M}{N}.$$

La stessa rappresentazione $X = X_1 + \dots + X_N$ permette di scrivere

$$\text{Var}(X) = \sum_{i=1}^n \text{Var}(X_i) + \sum_{i \neq j=1}^n \text{Cov}(X_i, X_j),$$

avendo usato il Corollario 3.75. Si ha $\text{Var}(X_i) = p_i(1 - p_i) = M(N - M)/N^2$, perché $X_i \sim \text{Be}(p_i)$. Inoltre, per ogni $i \neq j$ si ha

$$\text{Cov}(X_i, X_j) = E(X_i X_j) - E(X_i)E(X_j) = \frac{M(M-1)}{N(N-1)} - \frac{M^2}{N^2} = -\frac{M(N-M)}{N^2(N-1)},$$

dove il calcolo di $E(X_i X_j)$ è lasciato al lettore (si ricordi l'Esempio 3.99). Quindi

$$\text{Var}(X) = n \frac{M(N-M)}{N^2} - n(n-1) \frac{M(N-M)}{N^2(N-1)} = \frac{n(N-n)M(N-M)}{N^2(N-1)}.$$

3.5.5 Poisson

In numerose situazioni concrete, ci si trova a considerare variabili aleatorie binomiali in cui i parametri sono tali che n è molto grande e p è molto piccolo. Si consideri ad esempio il numero di accessi, in un fissato intervallo di tempo, ad un certo servizio, come lo sportello di un ufficio pubblico, un pagina web, un centralino... Sebbene vi sia un numero molto grande di utenti, diciamo n , che ha potenzialmente accesso a tale servizio, tipicamente un numero di utenti molto minore di n accede effettivamente al servizio nell'intervallo di tempo considerato. Un semplice modello matematico consiste nel supporre che ogni utente potenziale abbia una probabilità p molto piccola di accedere al servizio. Se assumiamo per semplicità che il valore di p sia uguale per tutti gli utenti, e che ogni utente si comporti in modo indipendente dagli altri, ci troviamo di fronte a uno schema di numerose prove ripetute e indipendenti: pertanto il numero X di utenti che effettivamente accede al servizio è una variabile aleatoria con distribuzione $\text{Bin}(n, p)$, ossia, ricordando (3.78),

$$p_X(k) = p_{\text{Bin}(n,p)}(k) = \binom{n}{k} p^k (1-p)^{n-k} \mathbb{1}_{\{0, \dots, n\}}(k).$$

Quando n è molto grande, calcoli espliciti con questa densità risultano pesanti, a causa della presenza dei coefficienti binomiali. È allora interessante analizzare il comportamento asintotico di $p_{\text{Bin}(n,p)}(k)$ quando $n \rightarrow +\infty$ e $p \rightarrow 0$. Per ottenere un comportamento non banale, è necessario che p “sia dell'ordine” di $1/n$. Per semplicità scegliamo $p = \frac{\lambda}{n}$, con $\lambda > 0$, ma il calcolo che segue vale nel caso più generale

in cui $p = p_n$ con $\lim_{n \rightarrow +\infty} n p_n = \lambda > 0$. Si ha

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

Si osservi che, per ogni k fissato, $n(n-1) \cdots (n-k+1)$ è un polinomio di grado k nella variabile n , il cui termine dominante per $n \rightarrow \infty$ è precisamente n^k , pertanto

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

Dato che, sempre per k fissato,

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

segue che

$$\lim_{n \rightarrow +\infty} p_X(k) = e^{-\lambda} \frac{\lambda^k}{k!}, \quad \forall k \in \mathbb{N}_0 \text{ fissato.} \quad (3.79)$$

Abbiamo dunque mostrato che, nel limite $n \rightarrow \infty$ con $p = \lambda/n$, la densità discreta $p_{\text{Bin}(n,p)}(k)$ converge verso un'espressione esplicita, che risulta essere una densità discreta su \mathbb{N}_0 , dal momento che, ricordando la serie esponenziale (0.5),

$$\sum_{n=0}^{\infty} e^{-\lambda} \frac{\lambda^k}{k!} = e^{-\lambda} \sum_{n=0}^{\infty} \frac{\lambda^k}{k!} = e^{-\lambda} e^{\lambda} = 1.$$

Questo motiva l'introduzione di una nuova classe di variabili aleatorie.

Dato $\lambda \in (0, \infty)$, diremo che una variabile aleatoria X è *Poisson di parametro λ* , e scriveremo $X \sim \text{Pois}(\lambda)$, se la densità discreta di X è data da

$$p_X(k) = p_{\text{Pois}(\lambda)}(k) = e^{-\lambda} \frac{\lambda^k}{k!} \mathbb{1}_{\mathbb{N}_0}(k). \quad (3.80)$$

per ogni $k \in \mathbb{N}_0$ (si veda la Figura 3.3). Si noti che $X(\Omega) = \mathbb{N}_0 = \{0, 1, \dots\}$. Possiamo quindi riformulare la convergenza (3.79) nel modo seguente.

Proposizione 3.108 (Poisson come limite della binomiale). *Sia $(p_n)_{n \in \mathbb{N}}$ una successione in $(0, 1)$ tale che $p_n \sim \frac{\lambda}{n}$ con $\lambda \in (0, \infty)$ (ossia $\lim_{n \rightarrow \infty} n p_n = \lambda$). Allora*

$$\lim_{n \rightarrow \infty} p_{\text{Bin}(n, p_n)}(k) = p_{\text{Pois}(\lambda)}(k), \quad \forall k \in \mathbb{N}_0.$$

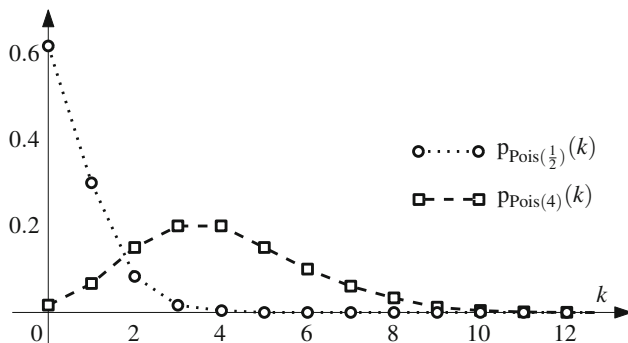


Figura 3.3 Densità discreta di una variabile aleatoria $\text{Pois}(\lambda)$, per $\lambda = \frac{1}{2}$ e $\lambda = 4$

Sulla base di questo risultato, detto talvolta *legge dei piccoli numeri*, le variabili di Poisson vengono comunemente usate per modellare, tra le altre, quantità del tipo “numero di accessi ad un servizio”, per le ragioni descritte in precedenza. Nel Paragrafo 4.1 mostreremo che, anche rimuovendo l’ipotesi che tutti gli “utenti potenziali” del servizio abbiano la stessa probabilità di accesso, la distribuzione del numero di accessi è comunque “vicina” ad una distribuzione di Poisson, purché le probabilità di accesso dei singoli utenti siano sufficientemente piccole. Questo spiega l’importanza e l’efficacia delle variabili di Poisson.

Studiamo ora le proprietà delle variabili aleatorie di Poisson. È agevole calcolare la funzione generatrice dei momenti di $X \sim \text{Pois}(\lambda)$:

$$M_X(t) = \sum_{k=0}^{\infty} e^{tk} p_X(k) = e^{-\lambda} \sum_{k=0}^{\infty} \frac{(e^t \lambda)^k}{k!} = e^{-\lambda} e^{e^t \lambda} = e^{\lambda(e^t - 1)}.$$

Ricordando il Teorema 3.96, segue che

$$M'_X(t) = \lambda e^t M_X(t), \quad M''_X(t) = \lambda e^t [M_X(t) + M'_X(t)],$$

da cui si ricava $E(X) = M'_X(0) = \lambda$, $E(X^2) = M''_X(0) = \lambda + \lambda^2$ e in particolare

$$E(X) = \lambda, \quad \text{Var}(X) = \lambda.$$

Sottolineiamo che questi valori si possono calcolare direttamente a partire dalla densità discreta, applicando la formula (3.30) (si veda l’Esercizio 3.19).

Il risultato che ora illustriamo afferma che, analogamente a quanto accade per le binomiali (Proposizione 3.103), la somma di due variabili aleatorie di Poisson *indipendenti* è ancora una variabile aleatoria di Poisson.

Proposizione 3.109 (Somma di Poisson indipendenti). *Siano $X \sim \text{Pois}(\lambda)$ e $Y \sim \text{Pois}(\mu)$ variabili aleatorie indipendenti. Allora $X + Y \sim \text{Pois}(\lambda + \mu)$.*

Dimostrazione. Ricordando la Proposizione 3.87, per ogni $n \in \mathbb{N}_0$ si ha

$$\begin{aligned} p_{X+Y}(n) &= (p_X * p_Y)(n) = \sum_{k \in \mathbb{R}} p_X(k) p_Y(n-k) = \sum_{k=0}^n e^{-\lambda} \frac{\lambda^k}{k!} e^{-\mu} \frac{\mu^{n-k}}{(n-k)!} \\ &= e^{-(\lambda+\mu)} \frac{1}{n!} \sum_{k=0}^n \binom{n}{k} \lambda^k \mu^{n-k} = e^{-(\lambda+\mu)} \frac{(\lambda + \mu)^n}{n!}, \end{aligned}$$

dove nell'ultima uguaglianza abbiamo applicato il binomio di Newton (1.29). \square

Concludiamo discutendo un esempio interessante, in cui le variabili di Poisson appaiono in modo naturale. Supponiamo di disporre casualmente n oggetti in r cassette numerati. Fissiamo $k \in \mathbb{N}_0$ e, per ogni $i \in \{1, \dots, r\}$, introduciamo l'evento

$C_i :=$ “il cassetto i -esimo contiene esattamente k oggetti”

e la variabile aleatoria

$$S := \text{“numero di cassette che contengono esattamente } k \text{ oggetti”} = \sum_{i=1}^r 1_{C_i}.$$

Gli eventi C_1, \dots, C_r non sono indipendenti, ma hanno tutti la stessa probabilità, ossia la probabilità che in un cassetto fissato finiscano esattamente k oggetti: come abbiamo ricordato nell'Esempio 3.105,

$$P(C_i) = p_{\text{Bin}(n, 1/r)}(k).$$

Per linearità del valor medio, si ha dunque $E(S) = r \cdot p_{\text{Bin}(n, 1/r)}(k)$.

Consideriamo ora il caso interessante in cui $n, r \gg 1$ e $n/r \approx \lambda \in (0, \infty)$. Per la Proposizione 3.108, in questo regime si ha

$$E(S) = r \cdot p_{\text{Bin}(n, 1/r)}(k) \approx r \cdot p_{\text{Pois}(\lambda)}(k) = r \cdot e^{-\lambda} \frac{\lambda^k}{k!}.$$

Si può inoltre mostrare che valori di S “distanti” dal valor medio $E(S)$ sono molto improbabili. Formalizzeremo precisamente questa affermazione nel Paragrafo 7.1.4 (il lettore interessato può guardare già ora l'enunciato della Proposizione 7.10). Per il momento, ci accontentiamo della seguente descrizione qualitativa, ma efficace:

Disponendo casualmente n oggetti in r cassette, con $n, r \gg 1$ e $\lambda := n/r$, il numero di cassette che contengono esattamente k oggetti è con grande probabilità dell'ordine di $\simeq r \cdot (e^{-\lambda} \frac{\lambda^k}{k!})$.

I termini “oggetti” e “cassetti” possono essere adattati a un grande numero di situazioni concrete, in cui l'accordo con la distribuzione di Poisson risulta notevole.

Esempio 3.110. Durante la seconda guerra mondiale, la parte meridionale di Londra fu colpita da $n = 535$ bombe volanti V1. Per analizzare la distribuzione geografica dei punti di impatto, tale area è stata suddivisa in $r = 576$ regioni di pari superficie, registrando quante bombe sono cadute in ciascuna regione.

Se ciascuna delle n bombe fosse assegnata “a caso” a una delle r regioni, ci aspetteremmo che, per ogni valore di k , il numero di regioni contenenti esattamente k bombe, che chiameremo “frequenza attesa”, sia all'incirca $\simeq r \cdot (e^{-\lambda} \frac{\lambda^k}{k!})$, dove $\lambda = n/r = \frac{535}{576} \simeq 0.929$. Riportiamo nella Tabella 3.1, per $k \in \{0, \dots, 6\}$, le frequenze attese accanto alle frequenze effettivamente osservate.

Tabella 3.1 Distribuzione delle bombe cadute su Londra durante la seconda guerra mondiale in 576 regioni di pari superficie (si veda l'Esempio 3.110)

Bombe ricevute	0	1	2	3	4	5	6 o più
Frequenze osservate	229	211	93	35	7	1	0
Frequenze attese	227.5	211.3	98.2	30.4	7.1	1.3	0.2

Come si vede, l'accordo è (almeno qualitativamente) impressionante, suggerendo che per lo più le bombe fossero lanciate “casualmente” e non in modo mirato. \square

3.5.6 Geometrica

Come abbiamo visto nell'Esempio 3.19 (riprendendo il Paragrafo 1.3.4), quando si eseguono n prove ripetute e indipendenti, in cui ciascuna prova ha probabilità di “successo” pari a p , la probabilità di ottenere il primo successo al tentativo k -esimo vale $p(1-p)^{k-1}$, per ogni $k \in \mathbb{N}$ con $k \leq n$ (cf. la formula (3.14)). Dato che il valore di tale probabilità non dipende da n , è naturale rimuovere la restrizione $k \leq n$ e considerare una variabile aleatoria che possa assumere ogni valore $k \in \mathbb{N}$, con probabilità $p(1-p)^{k-1}$. Si noti che, per la serie geometrica (0.4),

$$\sum_{k \in \mathbb{N}} p(1-p)^{k-1} = p \sum_{\ell=0}^{\infty} (1-p)^{\ell} = p \frac{1}{1-(1-p)} = 1,$$

quindi $k \mapsto p(1-p)^{k-1}$ definisce effettivamente una densità discreta su \mathbb{N} .

Dunque, dato $p \in (0, 1)$, diremo che una variabile aleatoria reale X è *geometrica di parametro p* , e scriveremo $X \sim \text{Geo}(p)$, se la sua densità discreta è data da

$$p_X(k) = p_{\text{Geo}(p)}(k) := p(1-p)^{k-1} \mathbb{1}_{\mathbb{N}}(k) \quad (3.81)$$

(si veda la Figura 3.4). Sottolineiamo che $X(\Omega) = \mathbb{N} = \{1, 2, \dots\}$.

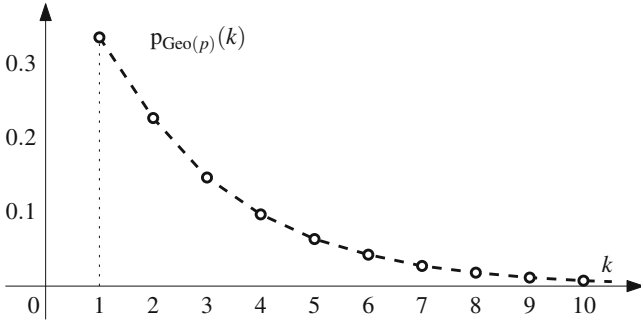


Figura 3.4 Densità discreta di una variabile aleatoria $\text{Geo}(p)$, per $p = \frac{1}{3}$

Osservazione 3.111. Un modo più “elegante” per liberarsi della restrizione $k \leq n$ consiste nel considerare una successione *infinita* di prove ripetute e indipendenti. Ricordando il Paragrafo 3.2.3, siano $(X_n)_{n \in \mathbb{N}}$ variabili aleatorie indipendenti $\text{Be}(p)$, ossia $P(X_n = 1) = 1 - P(X_n = 0) = p$, con $p \in (0, 1)$. Intuitivamente, $X_n = 1$ (risp. $X_n = 0$) significa che la n -esima prova ha avuto successo (risp. insuccesso). L’istante di primo successo è dato dalla variabile aleatoria $T : \Omega \rightarrow \mathbb{N} \cup \{+\infty\}$, definita da

$$T(\omega) := \min\{n \in \mathbb{N} : X_n(\omega) = 1\}, \quad (3.82)$$

con la convenzione $\min \emptyset := +\infty$. Si noti che vale l’uguaglianza di eventi

$$\{T = k\} = \{X_1 = 0, \dots, X_{k-1} = 0, X_k = 1\}, \quad \forall k \in \mathbb{N}, \quad (3.83)$$

dunque l’evento $\{T = k\}$ dipende solo dalle prime k variabili aleatorie X_1, \dots, X_k . Di conseguenza, la sua probabilità coincide con quella già calcolata per un numero finito di prove ripetute e indipendenti. Più direttamente, per l’indipendenza delle X_n ,

$$P(T = k) = P(X_1 = 0) \cdots P(X_{k-1} = 0) P(X_k = 1) = (1 - p)^{k-1} p, \quad (3.84)$$

per ogni $k \in \mathbb{N}$. Questo mostra che $p_T(\cdot) = p_{\text{Geo}(p)}(\cdot)$, ossia $T \sim \text{Geo}(p)$. Sottolineiamo che $P(T < +\infty) = \sum_{k \in \mathbb{N}} P(T = k) = 1$ e dunque $P(T = +\infty) = 0$.

In questa derivazione c’è un piccolo imbroglio: infatti, per l’Osservazione 3.42, la successione di variabili aleatorie $(X_n)_{n \in \mathbb{N}}$ non può essere definita su uno spazio di probabilità discreto. Tuttavia, anche in spazi di probabilità generali (che introdurremo nel Capitolo 5), la relazione (3.83) è sempre verificata, per definizione (3.82) di T , e di conseguenza il calcolo (3.84) continuerà a valere, perché esso si basa solo sul fatto che X_1, \dots, X_k siano variabili aleatorie $\text{Be}(p)$ indipendenti. \square

Osservazione 3.112. Talvolta è conveniente includere il caso $p = 1$: diremo che $X \sim \text{Geo}(1)$ se $X = 1$ quasi certamente: $P(X = 1) = 1$, o equivalentemente se la densità discreta di X è data da $p_X(k) = p_{\text{Geo}(1)}(k) := \mathbb{1}_{\{1\}}(k)$. Questo è in accordo con la formula (3.81), con la convenzione $0^0 := 1$. Nel seguito, considereremo per lo più il caso “interessante” $p \in (0, 1)$. \square

Osservazione 3.113. In diversi contesti (come vedremo nel Paragrafo 4.6) risulta utile considerare variabili aleatorie che assumono valori in $\mathbb{N}_0 = \{0, 1, 2, \dots\}$, ottenute da variabili geometriche sottraendo una unità. Indicheremo (la distribuzione di) tali variabili col simbolo $\text{Geo}_0(p)$. Più precisamente, scriveremo che $Y \sim \text{Geo}_0(p)$ se $Y + 1 \sim \text{Geo}(p)$, ossia

$$Y \sim \text{Geo}_0(p) \iff p_Y(k) = p_{\text{Geo}_0(p)}(k) = p(1-p)^k \mathbb{1}_{\mathbb{N}_0}(k). \quad (3.85)$$

Naturalmente i risultati che mostreremo per variabili $\text{Geo}(p)$ possono essere facilmente tradotti per variabili $\text{Geo}_0(p)$. Sottolineiamo che alcuni autori chiamano geometrica la distribuzione $\text{Geo}_0(p)$ invece che $\text{Geo}(p)$. \square

Calcoliamo ora media e varianza di una variabile aleatoria geometrica. A questo scopo, determiniamo innanzitutto la sua funzione generatrice dei momenti:

$$M_X(t) = \sum_{n=1}^{+\infty} e^{tn} p(1-p)^{n-1} = pe^t \sum_{\ell=0}^{+\infty} ((1-p)e^t)^\ell.$$

Quest'ultima è una serie geometrica di ragione $(1-p)e^t$: ricordando (0.4) si ha

$$M_X(t) = \begin{cases} \frac{pe^t}{1 - (1-p)e^t} = \frac{p}{e^{-t} - (1-p)} & \text{se } t < \log \frac{1}{1-p} \\ +\infty & \text{se } t \geq \log \frac{1}{1-p} \end{cases}. \quad (3.86)$$

Dato che $\log \frac{1}{1-p} > 0$, la funzione $M_X(t)$ è finita in un intorno di $t = 0$. Essendo

$$\begin{aligned} M'_X(t) &= \frac{pe^{-t}}{(e^{-t} - (1-p))^2} = \frac{p}{(e^{-t/2} - (1-p)e^{t/2})^2}, \\ M''_X(t) &= \frac{pe^{-t/2} + p(1-p)e^{t/2}}{(e^{-t/2} - (1-p)e^{t/2})^3}, \end{aligned}$$

applicando il Teorema 3.96 si trova

$$E(X) = M'(0) = \frac{1}{p}, \quad E(X^2) = M''(0) = \frac{(2-p)}{p^2}, \quad (3.87)$$

da cui segue che

$$\text{Var}(X) = E(X^2) - E^2(X) = \frac{1-p}{p^2}. \quad (3.88)$$

Osservazione 3.114. Se $Y \sim \text{Geo}_0(p)$, dato che $X := Y + 1 \sim \text{Geo}(p)$ si ottengono le formule $E(Y) = E(X) - 1 = \frac{1-p}{p}$ e $\text{Var}(Y) = \text{Var}(X) = \frac{1-p}{p^2}$. \square

Calcoliamo la probabilità che una variabile geometrica assuma valori strettamente maggiori di un determinato $n \in \mathbb{N}_0$:

$$\begin{aligned} P(X > n) &= \sum_{k=n+1}^{+\infty} p(1-p)^{k-1} = (1-p)^n \sum_{\ell=1}^{+\infty} p(1-p)^{\ell-1} \\ &= (1-p)^n \sum_{\ell=1}^{+\infty} p_{\text{Geo}(p)}(\ell) = (1-p)^n, \end{aligned} \quad (3.89)$$

dunque la funzione di ripartizione di X è data da

$$F_X(n) = P(X \leq n) = 1 - P(X > n) = 1 - (1-p)^n, \quad \forall n \in \mathbb{N}_0,$$

mentre $F_X(x) = F_X(\lfloor x \rfloor)$ per $x \in [0, \infty)$ e $F_X(x) = 0$ per $x \in (-\infty, 0)$. Grazie alla Proposizione 3.92, abbiamo anche un modo alternativo di calcolare $E(X)$:

$$E(X) = \sum_{n=0}^{\infty} P(X > n) = \sum_{n=0}^{\infty} (1-p)^n = \frac{1}{1-(1-p)} = \frac{1}{p},$$

avendo usato la serie geometrica (0.4).

Le variabili aleatorie geometriche godono di una proprietà fondamentale, detta *assenza di memoria* (per ragioni che discutiamo a breve).

Proposizione 3.115 (Assenza di memoria). *Sia $X \sim \text{Geo}(p)$. Allora*

$$P(X > n+m | X > n) = P(X > m), \quad \forall n, m \in \mathbb{N}_0. \quad (3.90)$$

Dimostrazione. Anzitutto si osservi che, poiché $\{X > n+m\} \subseteq \{X > n\}$,

$$P(X > n+m | X > n) = \frac{P(X > n+m, X > n)}{P(X > n)} = \frac{P(X > n+m)}{P(X > n)}.$$

Ricordando la formula (3.89) si ottiene

$$P(X > n+m | X > n) = \frac{(1-p)^{n+m}}{(1-p)^n} = (1-p)^m = P(X > m),$$

come voluto. □

Osservazione 3.116. Supponiamo che X rappresenti il tempo che è necessario attendere per il verificarsi di un determinato fenomeno aleatorio. Per $n \in \mathbb{N}$, indichiamo con $P^{(n)}(\cdot) := P(\cdot | X > n)$ la probabilità condizionale rispetto all'evento $\{X > n\}$, ossia “il fenomeno non si è verificato entro l'istante n ”, e definiamo la variabile aleatoria $X^{(n)} := X - n$, che rappresenta il “tempo residuo” che è ancora necessario aspettare. Possiamo allora riscrivere la relazione (3.90) come $P^{(n)}(X^{(n)} > m) = P(X > m)$, per ogni $m \in \mathbb{N}_0$. Dalla relazione $\{X = k\} = \{X > k-1\} \setminus \{X > k\}$, e da quella analoga

per $X^{(n)}$, si deduce che, se $X \sim \text{Geo}(p)$,

$$P^{(n)}(X^{(n)} = k) = P(X = k), \quad \forall k \in \mathbb{N}.$$

Questo mostra che *la variabile aleatoria $X^{(n)}$ rispetto alla probabilità condizionale $P^{(n)} = P(\cdot | X > n)$ ha la stessa densità discreta, e dunque la stessa distribuzione, della variabile aleatoria originale X* , e questo per ogni $n \in \mathbb{N}_0$. In altri termini, sapere che il fenomeno non si è verificato entro l'istante n , non modifica la distribuzione del tempo residuo $X^{(n)} = X - n$. Questa proprietà, intuitivamente chiara pensando all'istante di primo successo in una successione di prove ripetute e indipendenti, spiega il termine *assenza di memoria*. \square

In effetti, l'assenza di memoria è una proprietà che *caratterizza* le variabili aleatorie geometriche, come mostriamo ora.

Proposizione 3.117. *Sia X una variabile aleatoria a valori in \mathbb{N} , tale che la proprietà (3.90) è verificata. Allora $X \sim \text{Geo}(p)$, con $p := P(X = 1)$.*

Dimostrazione. Posto $m = 1$ in (3.90), si ottiene

$$P(X > n + 1) = P(X > n)P(X > 1),$$

che, per induzione, implica

$$P(X > n) = P(X > 1)^n, \quad \forall n \geq 1.$$

Ma allora, osservando che $\{X = n\} = \{X > n - 1\} \setminus \{X > n\}$, per ogni $n \geq 2$ si ha

$$\begin{aligned} p_X(n) &= P(X = n) = P(X > n - 1) - P(X > n) = P(X > 1)^{n-1} - P(X > 1)^n \\ &= (1 - P(X > 1))P(X > 1)^{n-1}. \end{aligned}$$

Per ipotesi $P(X \geq 1) = 1$, perché X prende valori in $\mathbb{N} = \{1, 2, \dots\}$, pertanto si ha $1 - P(X > 1) = P(X = 1)$. Ponendo $p := P(X = 1) = p_X(1) \in [0, 1]$, abbiamo dunque

$$p_X(n) = p(1 - p)^{n-1}, \quad \forall n \geq 2,$$

e la stessa relazione vale, per definizione di p , anche per $n = 1$ (purché conveniamo che $0^0 := 1$ se $p = 1$). Osserviamo che necessariamente $p \neq 0$, altrimenti si avrebbe $p_X(n) = 0$ per ogni $n \in \mathbb{N}$, il che è impossibile, dovendo essere $\sum_{n \in \mathbb{N}} p_X(n) = 1$. In definitiva, abbiamo mostrato che $p_X = p_{\text{Geo}(p)}$, dunque $X \sim \text{Geo}(p)$. \square

Concludiamo determinando la distribuzione del massimo e minimo di variabili aleatorie geometriche indipendenti.

Esempio 3.118. Siano $X_1, \dots, X_n \sim \text{Geo}(p)$ variabili aleatorie indipendenti. Determiniamo la densità discreta di

$$Z := \max(X_1, \dots, X_n), \quad W := \min(X_1, \dots, X_n).$$

Segue dalla formula (3.89) che la funzione di ripartizione di X_i è data da

$$F_{X_i}(k) = P(X_i \leq k) = 1 - P(X_i > k) = 1 - (1 - p)^k, \quad \forall k \in \mathbb{N}_0.$$

Ricordando la Proposizione 3.93, si ottengono le funzioni di ripartizione di Z e W :

$$F_Z(k) = [1 - (1 - p)^k]^n, \quad F_W(k) = 1 - (1 - p)^{nk}, \quad \forall k \in \mathbb{N}_0.$$

Dato che $F_Z(k^-) = F_Z(k - 1)$ e $F_W(k^-) = F_W(k - 1)$, per ogni $k \in \mathbb{N}$, si ha, usando la Proposizione 3.90,

$$\begin{aligned} p_Z(k) &= F_Z(k) - F_Z(k - 1) = [1 - (1 - p)^k]^n - [1 - (1 - p)^{k-1}]^n, \\ p_W(k) &= F_W(k) - F_W(k - 1) = (1 - p)^{n(k-1)} - (1 - p)^{nk}. \end{aligned}$$

Si noti che quest'ultima espressione può essere riscritta come

$$p_W(k) = (1 - (1 - p)^n)(1 - p)^{n(k-1)} = q(1 - q)^{k-1}, \quad \text{con } q := 1 - (1 - p)^n,$$

pertanto abbiamo mostrato che $W \sim \text{Geo}(q) = \text{Geo}(1 - (1 - p)^n)$. □

Esercizi

Esercizio 3.17. Sia $X \sim \text{Bin}(n, p)$. Con un calcolo diretto a partire dall'espressione (3.78) della densità discreta, si mostri che $E(X(X - 1)) = n(n - 1)p^2$. Ricordando che $E(X) = np$, si deduca quindi che $\text{Var}(X) = np(1 - p)$.

[Sugg. Si applichi la formula (3.30) e si noti che $k(k - 1)\binom{n}{k}p^k = p^2 \cdot \binom{n-2}{k-2}p^{k-2}$.]

Esercizio 3.18. Sia $X \sim \text{Bin}(n, p)$. Con un calcolo diretto a partire dall'espressione (3.78) della densità discreta, si mostri che $M_X(t) = (pe^t + (1 - p))^n$. Applicando il Teorema 3.96, si deducano quindi i valori di $E(X)$ e $E(X^2)$, e si ricavi $\text{Var}(X)$.

[Sugg. Si applichi la formula (3.30) e si ricordi il binomio di Newton (1.29).]

Esercizio 3.19. Sia $X \sim \text{Pois}(\lambda)$. Usando l'espressione (3.80) della densità discreta, si mostri che $E(X) = \lambda$ e $E(X(X - 1)) = \lambda^2$. Si deduca quindi che $\text{Var}(X) = \lambda$.

Esercizio 3.20. Siano $X \sim \text{Pois}(\lambda)$, $Y \sim \text{Pois}(\mu)$ variabili aleatorie indipendenti. Per $n \geq 0$ fissato, si determini la distribuzione della variabile aleatoria X rispetto alla probabilità condizionata $P(\cdot | X + Y = n)$.

Esercizio 3.21. Siano X, Y variabili aleatorie indipendenti con distribuzione $\text{Geo}(p)$. Per $n \geq 2$ fissato, si determini la distribuzione della variabile aleatoria X rispetto alla probabilità condizionata $P(\cdot | X + Y = n)$.

3.6 Esercizi di riepilogo

Esercizio 3.22. Si sceglie “a caso” un campione di 5 oggetti da un lotto di 100 di cui 10 sono difettosi per effettuare un controllo di qualità. Sia X il numero di oggetti difettosi contenuti nel campione. Si determini la densità discreta di X .

Esercizio 3.23. Siano X_1, X_2 variabili aleatorie indipendenti, ciascuna delle quali assume i valori $\{1, \dots, n\}$ con la stessa probabilità, dove $n \in \mathbb{N}$. Definiamo la variabile $Y := \min\{X_1, X_2\}$.

- (i) Si calcoli $P(Y = k)$ per ogni $k \in \mathbb{N}$.
- (ii) Si mostri che, per ogni $t \in (0, 1)$, si ha che

$$\lim_{n \rightarrow \infty} P(Y \leq tn) = 2t - t^2.$$

Esercizio 3.24. Un gioco a premi ha un montepremi di 512 euro. Vengono poste ad un concorrente 10 domande. Ad ogni risposta errata il montepremi viene dimezzato. Alla prima risposta esatta il concorrente vince il montepremi rimasto. Se non si dà alcuna risposta esatta non si vince nulla. Un certo concorrente risponde esattamente a ciascuna domanda con probabilità $p \in (0, 1)$, indipendentemente dalle risposte alle altre domande. Sia X la vincita in euro di questo concorrente. Si determini la densità discreta p_X di X .

Esercizio 3.25. Due mazzi di 40 carte sono ognuno costituito da 20 carte rosse e 20 nere. Si mescolano entrambi i mazzi, quindi si dispongono uno accanto all'altro. Cominciamo con lo scoprire la prima carta di entrambi i mazzi. Se entrambe sono rosse vinciamo un euro, altrimenti non vinciamo alcunché. Proseguiamo scoprendo le seconde carte dei due mazzi: se sono entrambe rosse vinciamo un altro euro, e così via. Sia X il numero di euro vinti dopo aver scoperto tutte le carte. Si calcoli la densità discreta di X .

Esercizio 3.26. Un'urna contiene $n \geq 1$ palline bianche e 2 palline rosse. Si eseguono estrazioni ripetute *senza reimmissione*. Introduciamo la variabile aleatoria

X = numero di palline bianche estratte prima di estrarre una pallina rossa,

la cui densità discreta verrà indicata con $p_X(k) = P(X = k)$.

- (i) Si mostri che, per $k = 0, 1, \dots, n$,

$$p_X(k) = \frac{2}{(n+2)(n+1)}(n-k+1).$$

- (ii) Si calcoli $E(X)$.

[Sugg. Si ricordi la formula (0.7).]

Esercizio 3.27. Da un'urna contenente r palline rosse e v palline verdi si estraggono successivamente, senza reimmissione, k palline, con $k \leq \min(r, v)$. Sia

$$X_i := \begin{cases} 1 & \text{se l}'i\text{-ma pallina estratta è rossa} \\ 0 & \text{altrimenti} \end{cases}, \quad \text{per } i = 1, 2, \dots, k,$$

e definiamo $X = X_1 + \dots + X_k$.

- (i) Si determini la distribuzione di X .
- (ii) Si determinino le distribuzioni delle X_i .
- (iii) Si calcoli $E(X)$.
- (iv) Si mostri, per induzione su k , che la densità congiunta delle X_i è data da

$$\begin{aligned} p_{X_1, \dots, X_k}(x_1, \dots, x_k) \\ = \frac{r(r-1) \cdots (r - \sum_{i=1}^k x_i + 1) v(v-1) \cdots (v - k + \sum_{i=1}^k x_i + 1)}{(r+v)(r+v-1) \cdots (r+v-k+1)} \end{aligned}$$

Esercizio 3.28. Siano W, T variabili aleatorie indipendenti, a valori in $\{0, 1\}$ e $\mathbb{N} = \{1, 2, \dots\}$ rispettivamente, con le seguenti distribuzioni marginali:

$$P(T=0) = P(T=1) = \frac{1}{2}, \quad P(W=n) = p(1-p)^{n-1} \mathbb{1}_{\mathbb{N}}(n),$$

dove $p \in (0, 1)$ è un parametro fissato. Definiamo la variabile

$$X := W \mathbb{1}_{\{T=0\}} + \frac{1}{W} \mathbb{1}_{\{T=1\}},$$

che può dunque assumere come valori i numeri naturali e i reciproci dei numeri naturali, ossia $X(\Omega) = \mathbb{N} \cup \{\frac{1}{n}\}_{n \in \mathbb{N}}$.

- (i) Si determini la densità discreta di X .
- (ii) Si mostri che la variabile aleatoria $Y := 1/X$ ha la stessa distribuzione di X .
- (iii) Si calcoli $E(X)$.

[Sugg. Si ricordi che $\sum_{n=1}^{\infty} \frac{1}{n} x^n = -\log(1-x)$ e $\sum_{n=1}^{\infty} n x^{n-1} = (1-x)^{-2}$, per $|x| < 1$.]

Esercizio 3.29. Sia $n \in \mathbb{N}$, $n \geq 3$ e indichiamo con S_n il gruppo delle permutazioni di $\{1, \dots, n\}$, munito della probabilità P uniforme. Gli elementi di S_n saranno indicati con $\sigma = (\sigma(1), \dots, \sigma(n))$. Introduciamo le variabili aleatorie reali X, Y definite da

$$X(\sigma) := \sigma(1), \quad Y(\sigma) := \sigma(2).$$

- (i) Si mostri che, per ogni $i, j \in \{1, 2, \dots, n\}$, la densità congiunta di X e Y vale

$$p_{X,Y}(i, j) = \begin{cases} \frac{1}{c_n} & \text{se } i \neq j \\ 0 & \text{se } i = j \end{cases},$$

dove c_n è un'opportuna costante che è richiesto di determinare.

(ii) Si determini la densità della variabile $D := Y - X$.

[Sugg. Basta calcolare $p_D(m)$ per $m > 0$, poiché per simmetria $p_D(-m) = p_D(m)$.]

Siano ora Z, W due variabili aleatorie reali indipendenti, definite su uno spazio di probabilità (Ω, \tilde{P}) , tali che $\tilde{P}(Z = i) = \tilde{P}(W = i) = \frac{1}{n}$, per ogni $i \in \{1, \dots, n\}$.

(iii) Si calcoli $\tilde{P}(Z \neq W)$.

(iv) Si mostri che $\tilde{P}(Z = i, W = j | Z \neq W) = p_{X,Y}(i, j)$ per ogni $i, j \in \{1, 2, \dots, n\}$.
In altri termini, il vettore aleatorio (Z, W) , rispetto alla probabilità condizionale $\tilde{P}(\cdot | Z \neq W)$, ha la stessa distribuzione congiunta del vettore iniziale (X, Y) .

Esercizio 3.30. Fissati $p \in (0, 1)$ e $n \geq 2$, siano Z_1, \dots, Z_n variabili aleatorie indipendenti a valori in $\{-1, 1\}$, con $P(Z_i = 1) = p$ per ogni $i = 1, \dots, n$. Definiamo

$$X := \prod_{i=1}^n Z_i = Z_1 \cdot Z_2 \cdots Z_n.$$

(i) Si determini $E(X)$ e si deduca la distribuzione di X .

[Sugg. Che valori può assumere X ?]

(ii) X è indipendente dal vettore aleatorio (Z_1, \dots, Z_n) ?

(iii) X è indipendente dal vettore aleatorio (Z_2, \dots, Z_n) ?

Esercizio 3.31. Da un mazzo di 50 carte numerate da 1 a 50, si estraggono a caso 3 carte. Introduciamo le seguenti variabili aleatorie:

$X :=$ numero più basso estratto;

$Z :=$ numero più alto estratto;

$Y :=$ terzo numero estratto.

(i) Si determinino le distribuzioni marginali di X, Y e Z .

(ii) Si determini la distribuzione di (X, Y) e si mostri che $Y - X$ ha la stessa distribuzione di X .

Esercizio 3.32. Si estraggono, una dopo l'altra, senza reimmissione, quattro palline da un'urna che contiene 10 palline, numerate da 1 a 10.

(i) Sia X il più alto numero estratto. Qual è la distribuzione di X ?

(ii) Sia A l'evento "tra i numeri delle quattro palline estratte non ce ne sono due di consecutivi". Qual è la probabilità di A ?

(iii) Si calcoli $P(A | X = 7)$.

(iv) Qual è la probabilità di A condizionata al fatto che tra i numeri estratti c'è il 7?

Esercizio 3.33. Sia Ω l'insieme dei sottoinsiemi non vuoti di $\{1, 2, \dots, N\}$, dove $N \in \mathbb{N}$ con $N \geq 2$. In altre parole

$$\Omega := \{\omega \subseteq \{1, 2, \dots, N\} : \omega \neq \emptyset\}.$$

Se $\omega \in \Omega$ sia $X(\omega) := \max(\omega)$ il massimo elemento di ω e $Y(\omega) := \min(\omega)$ il minimo elemento di ω . Infine, sia P la probabilità uniforme su Ω .

- (i) Si mostri che, per $n \in \{1, 2, \dots, N\}$,

$$P(X = n) = \frac{2^{n-1}}{2^N - 1}.$$

- (ii) Si calcoli la funzione generatrice dei momenti di X .
 (iii) Si determini la densità congiunta di (X, Y) .
 (iv) Si determini la densità di $X - Y$.

Esercizio 3.34. Sia $\Omega_n := \{0, 1\}^n$, e

$$\Omega := \bigcup_{n=1}^{+\infty} \Omega_n.$$

In altre parole gli elementi di Ω sono sequenze binarie di lunghezza arbitraria, ma finita. Se $\omega \in \Omega$, l'unico $n \geq 1$ per cui $\omega \in \Omega_n$ si dice *lunghezza* di ω , e si denota con $l(\omega)$. Per ogni $\omega \in \Omega$, definiamo

$$p(\omega) := 2^{-2l(\omega)}.$$

- (i) Si mostri che p è una densità discreta su Ω (nel senso della Definizione 1.9). Pertanto, secondo la Proposizione 1.11, essa identifica una probabilità P .
 (ii) Sullo spazio di probabilità (Ω, P) appena definito, si definisca per ogni $i \geq 1$ la variabile aleatoria

$$X_i(\omega) := \begin{cases} \omega_i & \text{se } i \leq l(\omega) \\ 0 & \text{altrimenti} \end{cases}$$

Si determini la distribuzione di X_i .

Esercizio 3.35. Un insetto depone un numero aleatorio $N \sim \text{Pois}(\lambda)$ di uova. Ciascun uovo deposto si schiude con probabilità $p \in (0, 1)$, indipendentemente dal numero di uova deposte e dal fatto che le altre si schiudano. Indichiamo con X il numero (aleatorio) di uova che si schiudono.

- (i) Qual è il valore di $P(X = k | N = n)$, per $n \in \mathbb{N}_0$ e $k \in \mathbb{R}$?
 (ii) Si determini la distribuzione di X .

Esercizio 3.36. Siano X e Y due variabili aleatorie a valori in \mathbb{N}_0 aventi la seguente densità congiunta:

$$p_{X,Y}(k, n) = \begin{cases} \binom{n}{k} p^k (1-p)^{n-k} e^{-\lambda} \frac{\lambda^n}{n!} & \text{se } 0 \leq k \leq n \\ 0 & \text{altrimenti,} \end{cases}$$

dove $p \in (0, 1)$ e $\lambda > 0$ sono parametri fissati. Si determinino le densità marginali di X e Y e si calcoli $\text{Cov}(X, Y)$.

Esercizio 3.37. Siano X, Z e W variabili aleatorie indipendenti con $X \sim \text{Be}(p)$ e $Z, W \sim \text{Pois}(\lambda)$. Definiamo $Y := XZ + W$.

- (i) Si spieghi la seguente uguaglianza di eventi: per ogni $m \in \{0, 1\}$ e $n \in \mathbb{N}_0$

$$\{X = m, Y = n\} = \{X = m, mZ + W = n\}.$$

- (ii) Si determinino le densità discrete di (X, Y) e di Y .

- (iii) Usando la densità p_Y ottenuta al punto precedente, si calcoli $E(Y)$ e $\text{Var}(Y)$.

- (iv) Si calcoli $E(Y)$ e $\text{Var}(Y)$ *senza* utilizzare p_Y .

Esercizio 3.38. Un gioco a premi ha il seguente funzionamento. Il montepremi iniziale è $C > 0$. Il concorrente lancia due volte una moneta. Se l'esito del primo e del secondo lancio sono uguali, il concorrente vince l'intero montepremi. In caso contrario il montepremi si dimezza. Il concorrente lancia la moneta una terza volta. Se l'esito del terzo lancio è uguale a quello del secondo, allora il concorrente vince $C/2$. In caso contrario il montepremi si dimezza ulteriormente, e così via. Sia X il montepremi vinto dal concorrente al termine del gioco.

(Si accetti il fatto che il gioco ha termine, cioè che non sia possibile avere una successione infinita di esiti in cui ogni esito è diverso dal precedente.)

- (i) Assumendo che la moneta sia equilibrata, si determini la densità discreta di X .
(ii) Sempre nell'ipotesi che la moneta sia equilibrata, si determini il valor medio di X (si ricordi la serie geometrica (0.4)).
(iii) Supponiamo ora che la moneta non sia equilibrata: la probabilità che esca *testa* è $p \in (0, 1)$. Si determinino la densità e il valor medio di X .

Esercizio 3.39. Elena lancia ripetutamente una moneta. La probabilità di ottenere testa vale $p \in (0, 1)$. Detto $X_k \in \{0, 1\}$ l'esito dell' k -esimo lancio (testa = 1, croce = 0), indichiamo con S, T rispettivamente il primo e secondo istante in cui esce testa:

$$S := \min\{k \in \mathbb{N} : X_k = 1\}, \quad T := \min\{k \in \mathbb{N}, k > S : X_k = 1\},$$

con la convenzione $\min \emptyset := +\infty$. Poniamo quindi

$$U := T - S$$

e definiamo infine per $k \in \mathbb{N}$ gli eventi

$$A_k := \{X_k = 1\}.$$

- (i) Fissiamo $n, m \in \mathbb{N}$ con $n > m$. Si esprima l'evento $\{S = m, T = n\}$ in funzione degli eventi A_1, \dots, A_n mediante opportune operazioni insiemistiche. Si deduca quindi la densità congiunta

$$p_{S,T}(m, n) = P(S = m, T = n) = (1 - p)^{n-2} p^2 \mathbb{1}_{\{1 \leq m < n\}}.$$

- (ii) Si determini la densità congiunta delle variabili aleatorie S, U . Si deduca che S e U sono indipendenti e hanno la stessa distribuzione marginale (quale?).

- (iii) Si calcolino $E(T)$ e $\text{Var}(T)$.

Esercizio 3.40. Stefano lancia ripetutamente un dado regolare a sei facce. Indichiamo con X_k il risultato dell' k -esimo lancio, per $k \in \mathbb{N}$. Matteo sta a guardare gli esiti dei lanci, aspettando il primo istante T in cui esce un numero in $\{1, 2, 3, 4, 5\}$. Quando ciò accade, si appunta su un foglio il numero uscito $Y := X_T$.

- (i) Per ogni $n \in \mathbb{N}$, si esprima l'evento $\{T = n\}$ in termini delle variabili aleatorie X_1, \dots, X_n . Si deduca quindi la densità discreta di T e la si riconosca.
- (ii) Per ogni $n \in \mathbb{N}$ e $a \in \{1, 2, 3, 4, 5\}$, si esprima l'evento $\{T = n, Y = a\}$ in termini delle variabili aleatorie X_1, \dots, X_n . Si deduca quindi la densità discreta congiunta delle variabili aleatorie T e Y .
- (iii) Si determini la distribuzione di Y . Le variabili T e Y sono indipendenti?

Esercizio 3.41. Consideriamo una famiglia di urne, indicizzate dai numeri naturali incluso lo zero: l'urna k -esima, con $k \in \mathbb{N}_0 = \{0, 1, 2, \dots\}$, contiene 1 pallina nera e k palline bianche. Katia sceglie un numero casuale X , con distribuzione $\text{Pois}(\lambda)$, e pesca una pallina a caso dall'urna numero X . Indichiamo con A l'evento "la pallina pescata da Katia è nera".

- (i) Si determini $P(A|X = k)$, per ogni $k \in \mathbb{N}_0$. Si deduca che $P(A) = \frac{1-e^{-\lambda}}{\lambda}$.
- (ii) Si determini $q(k) := P(X = k|A)$ per ogni $k \in \mathbb{N}_0$.
- (iii) Si mostri che $q(\cdot)$ coincide con la densità discreta della variabile aleatoria $Y := X - 1$ rispetto alla probabilità $P(\cdot | X \geq 1)$.
- (iv) Si calcoli $E(X|A)$, ossia il valore atteso di X rispetto alla probabilità $P(\cdot | A)$.

3.7 Note bibliografiche

Come nel Capitolo 1, anche il contenuto di questo capitolo è trattato in tutti i testi introduttivi di probabilità. Oltre ai riferimenti bibliografici già consigliati per il Capitolo 1, citiamo il testo di G. Grimmett e D. Stirzaker [28] contenente, tra l'altro, numerose applicazioni della funzione generatrice dei momenti.

La disuguaglianza (3.53) è un caso speciale di un'importante disuguaglianza in meccanica statistica, detta FKG dal nome dei fisici matematici C.M. Fortuin, P.W. Kasteleyn e J. Ginibre [26].

L'Esempio 3.110, descritto nel libro di W. Feller [23], è tratto dall'articolo [13] di R.D. Clarke.

Variabili aleatorie discrete: esempi e applicazioni

Sommario. In questo capitolo studiamo alcuni modelli probabilistici la cui analisi coinvolge variabili aleatorie. In alcuni casi, i modelli verranno definiti a partire da una successione infinita di variabili aleatorie indipendenti e con la stessa distribuzione, *non* concentrata in un punto. Come abbiamo spiegato nell'Osservazione 3.42, una tale successione non può essere costruita in uno spazio di probabilità discreto e richiede una nozione più generale di spazio di probabilità, che introdurremo nel Capitolo 5. Abbiamo deciso di usare questa formulazione, in anticipo rispetto alle nozioni teoriche, perché consente maggiore sintesi e, in ogni caso, comporta l'uso di strumenti di calcolo che restano sempre nell'ambito della probabilità discreta. Infatti, nei problemi che affronteremo, ci si potrà sempre ricondurre a eventi che dipendono solo da un tratto finito della successione.

4.1 Sulla legge dei piccoli numeri

In questo paragrafo ci proponiamo di migliorare la Proposizione 3.108, nota come *legge dei piccoli numeri*, in due direzioni:

- mostreremo che la densità di una variabile di Poisson è una buona approssimazione per la densità di una somma di variabili di Bernoulli indipendenti, anche nel caso in cui i parametri delle variabili di Bernoulli siano diversi, purché siano piccoli (in un senso preciso);
- forniremo un risultato di approssimazione non solo nella forma di limite, ma dando una stima esplicita dell'errore per ogni fissato valore del numero di variabili di Bernoulli e dei rispettivi parametri.

Ricordiamo che una *densità discreta* su un insieme E è una funzione $p : E \rightarrow [0, 1]$ tale che $\sum_{x \in E} p(x) = 1$. Definiamo una nozione di *distanza* fra densità discrete.

Ricordiamo che si dice *distanza su un insieme* V qualsiasi funzione $d(\cdot, \cdot) : V \times V \rightarrow [0, +\infty)$ che soddisfa le seguenti proprietà:

- $d(x, y) = 0$ se e solo se $x = y$;

Caravenna F., Dai Pra P.: *Probabilità. Un'introduzione attraverso modelli e applicazioni*.

Unitext – La Matematica per il 3+2 67

DOI 10.1007/978-88-470-2595-0_4, © Springer-Verlag Italia 2013

- simmetria: $d(x, y) = d(y, x)$ per ogni $x, y \in V$;
- disuguaglianza triangolare: $d(x, y) \leq d(x, z) + d(z, y)$ per ogni $x, y, z \in V$.

Se V è uno spazio vettoriale reale e $\|\cdot\|$ è una norma su V (definita a pagina 139) allora la funzione $d(x, y) := \|y - x\|$ è una distanza, come si verifica facilmente.

Definizione 4.1. Siano p e q due densità discrete sullo stesso insieme E . Chiameremo *distanza in variazione totale* fra p e q la quantità

$$d(p, q) := \frac{1}{2} \sum_{x \in E} |p(x) - q(x)|.$$

Non è difficile mostrare che $d(\cdot, \cdot)$ è effettivamente una distanza nell'insieme delle densità discrete su E . In particolare, soddisfa la disuguaglianza triangolare: $d(p, q) \leq d(p, q') + d(q', q)$ per ogni scelta delle densità discrete p, q, q' .

Illustreremo ora un metodo molto utile per fornire stime sulla distanza in variazione totale tra due densità discrete, noto come il *metodo dell'accoppiamento*. Accoppiare due densità p e q significa realizzare nello stesso spazio di probabilità due variabili aleatorie discrete X e Y le cui densità siano rispettivamente p e q .[†] Questo può essere fatto in molti modi diversi: ad esempio, si possono considerare due variabili aleatorie *indipendenti* di densità, rispettivamente, p e q . Questo modo semplice risulta, nella gran parte di casi, scarsamente efficace.

Il risultato che segue costituisce lo strumento fondamentale del metodo dell'accoppiamento.

Proposizione 4.2. Siano X e Y due variabili aleatorie discrete, definite nello stesso spazio di probabilità (Ω, P) , e a valori in E . Allora

$$d(p_X, p_Y) \leq P(X \neq Y).$$

Dimostrazione. Osserviamo che

$$\begin{aligned} p_X(x) - p_Y(x) &= P(X = x) - P(Y = x) \leq P(X = x) - P(X = x, Y = x) \\ &= P(\{X = x\} \setminus \{X = x, Y = x\}) = P(X = x, Y \neq x). \end{aligned}$$

Analogamente $p_Y(x) - p_X(x) \leq P(Y = x, X \neq x)$, da cui segue che

$$|p_X(x) - p_Y(x)| \leq P(X = x, Y \neq x) + P(Y = x, X \neq x). \quad (4.1)$$

Inoltre, considerando il fatto che $(\{X = x\})_{x \in X(\Omega)}$ è una partizione numerabile di Ω ,

[†] Notiamo che ciò è equivalente (perché?) a definire una densità congiunta di cui p e q siano le densità marginali.

abbiamo che

$$\begin{aligned} P(X \neq Y) &= P\left(\bigcup_{x \in X(\Omega)} \{X \neq Y\} \cap \{X = x\}\right) \\ &= \sum_{x \in X(\Omega)} P(X \neq Y, X = x) = \sum_{x \in X(\Omega)} P(X = x, Y \neq x). \end{aligned}$$

Per simmetria, si ha anche $P(X \neq Y) = \sum_{x \in X(\Omega)} P(Y = x, X \neq x)$, e quindi

$$2P(X \neq Y) = \sum_{x \in X(\Omega)} [P(X = x, Y \neq x) + P(Y = x, X \neq x)]. \quad (4.2)$$

La conclusione segue dal confronto di (4.1) e (4.2). \square

Vediamo ora l'applicazione del precedente risultato a due casi speciali.

Lemma 4.3. *Siano $X \sim \text{Be}(p)$ e $Y \sim \text{Pois}(p)$. Allora*

$$d(p_X, p_Y) \leq p^2.$$

Dimostrazione. Per capire l'idea della dimostrazione, supponiamo che $p \ll 1$ sia piccolo e $Y \sim \text{Pois}(p)$. Poiché $P(Y = 0) = e^{-p} \simeq 1$, è abbastanza facile vedere che, posto $X := \mathbb{1}_{\{Y \geq 1\}}$, le densità di X e Y sono “vicine” in variazione totale (le variabili aleatorie X e Y sono definite sullo stesso spazio di probabilità, e quindi è possibile applicare la Proposizione 4.2). Si noti che $X \sim \text{Be}(1 - e^{-p})$ mentre, nell'enunciato del Teorema, compare $X \sim \text{Be}(p)$. Notiamo che per p piccolo si ha $1 - e^{-p} \simeq p$. Per eliminare la discrepanza, correggiamo come segue il precedente argomento.

Consideriamo una variabile aleatoria Z a valori in $\{-1, 0, 1, 2, \dots\}$, con la seguente distribuzione:

$$P(Z = k) = \begin{cases} e^{-p} \frac{p^k}{k!} & \text{se } k \geq 1 \\ 1 - p & \text{se } k = 0 \\ e^{-p} - (1 - p) & \text{se } k = -1 \end{cases}.$$

Essendo $e^{-p} \geq 1 - p$, si tratta di una buona definizione. Definiamo poi

$$X := \mathbb{1}_{\{Z \neq 0\}}, \quad Y := Z \mathbb{1}_{\{Z \neq -1\}}.$$

Lasciamo al lettore la semplice verifica del fatto che $X \sim \text{Be}(p)$ e $Y \sim \text{Pois}(p)$. Inoltre, osservando che $\{X \neq Y\} = \{Z \notin \{0, 1\}\}$, per la Proposizione 4.2 e la disuguaglianza $e^{-p} \geq 1 - p$, abbiamo:

$$\begin{aligned} d(p_X, p_Y) &\leq P(Z \notin \{0, 1\}) = 1 - P(Z = 0) - P(Z = 1) \\ &= 1 - e^{-p} - pe^{-p} \leq 1 - (1 - p) - p(1 - p) = p^2, \end{aligned}$$

che completa la dimostrazione. \square

Lemma 4.4. *Siano $X \sim \text{Pois}(\lambda)$ e $Y \sim \text{Pois}(\mu)$. Allora*

$$d(p_X, p_Y) \leq |\lambda - \mu|.$$

Dimostrazione. Senza perdita di generalità, possiamo supporre che $\lambda > \mu$. Siano (X, B_1, B_2, \dots) variabili aleatorie indipendenti, definite su un opportuno spazio di probabilità Ω , con $X \sim \text{Pois}(\lambda)$ e $B_i \sim \text{Be}(\frac{\mu}{\lambda})$. Definiamo $Z_0 := 0$ e, per ogni $n \in \mathbb{N}$, poniamo $Z_n := B_1 + \dots + B_n$. Si ha allora $Z_n \sim \text{Bin}(n, \frac{\mu}{\lambda})$ per la Proposizione 3.102. Si noti inoltre che, per ogni $n \geq 1$ fissato, X e Z_n sono variabili aleatorie indipendenti (perché?). Sullo stesso spazio di probabilità Ω definiamo la variabile aleatoria Y ponendo

$$Y(\omega) := Z_{X(\omega)}(\omega), \quad \forall \omega \in \Omega.$$

Cominciamo con il mostrare che $Y \sim \text{Pois}(\mu)$. Per ogni $k \in \mathbb{N}_0$ vale l'uguaglianza di eventi $\{Y = k\} = \bigcup_{n \geq k} \{X = n, Z_n = k\}$, dove la restrizione $n \geq k$ è dovuta al fatto che, per costruzione, $Z_n \leq n$. Pertanto

$$\begin{aligned} P(Y = k) &= P\left(\bigcup_{n \geq k} \{X = n, Z_n = k\}\right) = \sum_{n \geq k} P(X = n, Z_n = k) \\ &= \sum_{n \geq k} P(X = n) P(Z_n = k) = \sum_{n \geq 1} e^{-\lambda} \frac{\lambda^n}{n!} \binom{n}{k} \left(\frac{\mu}{\lambda}\right)^k \left(1 - \frac{\mu}{\lambda}\right)^{n-k} \\ &= e^{-\lambda} \frac{\lambda^k}{k!} \left(\frac{\mu}{\lambda}\right)^k \sum_{n \geq k} \frac{\lambda^{n-k}}{(n-k)!} \left(1 - \frac{\mu}{\lambda}\right)^{n-k} = e^{-\lambda} \frac{\mu^k}{k!} e^{\lambda-\mu} = e^{-\mu} \frac{\mu^k}{k!}. \end{aligned}$$

Usando ora la Proposizione 4.2, notando che $\{X = Y\} = \bigcup_{n \geq 0} \{X = n, Y = n\} = \bigcup_{n \geq 0} \{X = n, Z_n = n\}$, per definizione di Y , si ha

$$\begin{aligned} d(p_X, p_Y) &\leq 1 - P(X = Y) = 1 - P\left(\bigcup_{n \geq 0} \{X = n, Z_n = n\}\right) \\ &= 1 - \sum_{n \geq 0} P(X = n) P(Z_n = n) = 1 - \sum_{n \geq 0} e^{-\lambda} \frac{\lambda^n}{n!} \left(\frac{\mu}{\lambda}\right)^n \\ &= 1 - e^{-\lambda} \sum_{n \geq 0} \frac{\mu^n}{n!} = 1 - e^{-(\lambda-\mu)} \leq \lambda - \mu, \end{aligned}$$

avendo usato il fatto che $1 - e^{-x} \leq x$ per ogni $x \in \mathbb{R}$. □

Siamo ora in condizione di enunciare il principale risultato di questo paragrafo.

Teorema 4.5. *Siano X_1, X_2, \dots, X_n variabili aleatorie indipendenti tali che $X_i \sim \text{Be}(q_i)$. Posto $S_n := X_1 + X_2 + \dots + X_n$ e $W_n \sim \text{Pois}(q_1 + q_2 + \dots + q_n)$, si ha che*

$$d(p_{S_n}, p_{W_n}) \leq \sum_{i=1}^n q_i^2.$$

Prima di dimostrare il Teorema 4.5, mostriamo che esso costituisce un'estensione della Proposizione 3.108. Se $q_i \equiv p_n$ per ogni $i = 1, \dots, n$, allora $S_n \sim \text{Bin}(n, p_n)$, per la Proposizione 3.102, e $W_n \sim \text{Pois}(np_n)$, perché $q_1 + q_2 + \dots + q_n = np_n$. Supponiamo ora che $np_n \rightarrow \lambda \in (0, \infty)$ per $n \rightarrow \infty$, e poniamo $Y \sim \text{Pois}(\lambda)$. Notando che $\sum_{i=1}^n q_i^2 = np_n^2$, segue dalla disuguaglianza triangolare e dal Lemma 4.4 che

$$d(p_{S_n}, p_Y) \leq d(p_{S_n}, p_{W_n}) + d(p_{W_n}, p_Y) \leq np_n^2 + |\lambda - np_n|.$$

Dato che $np_n^2 = (np_n)^2/n \sim \lambda/n \rightarrow 0$ per $n \rightarrow \infty$, abbiamo mostrato che

$$\lim_{n \rightarrow +\infty} d(p_{S_n}, p_Y) = 0. \quad (4.3)$$

Ovviamente la Proposizione 3.108 segue da (4.3), essendo

$$|p_{S_n}(k) - p_Y(k)| \leq 2d(p_{S_n}, p_Y), \quad \forall k \in \mathbb{N}_0.$$

Dimostrazione (del Teorema 4.5). Siano Z_1, \dots, Z_n variabili aleatorie indipendenti, con le seguenti distribuzioni marginali:

$$P(Z_i = k) = \begin{cases} e^{-p_i} \frac{p_i^k}{k!} & \text{se } k \geq 1 \\ 1 - p_i & \text{se } k = 0 \\ e^{-p_i} - (1 - p_i) & \text{se } k = -1 \end{cases}.$$

Posto $X_i := \mathbb{1}_{\{Z_i \neq 0\}}$ e $Y_i := Z_i \mathbb{1}_{\{Z_i \neq -1\}}$, per $i = 1, \dots, n$, abbiamo che

- $X_i \sim \text{Be}(q_i)$, $Y_i \sim \text{Pois}(q_i)$ e $P(X_i \neq Y_i) \leq q_i^2$, per ogni $i = 1, \dots, n$, come abbiamo visto nella dimostrazione del Lemma 4.3;
- X_1, \dots, X_n sono variabili aleatorie indipendenti, per la Proposizione 3.40, e analogamente anche Y_1, \dots, Y_n sono variabili aleatorie indipendenti;
- $W_n := Y_1 + Y_2 + \dots + Y_n \sim \text{Pois}(q_1 + q_2 + \dots + q_n)$, per la Proposizione 3.109.

Pertanto, se poniamo $S_n := X_1 + X_2 + \dots + X_n$,

$$d(p_{S_n}, p_{W_n}) \leq P(S_n \neq W_n) \leq P\left(\bigcup_{i=1}^n \{X_i \neq Y_i\}\right) \leq \sum_{i=1}^n P(X_i \neq Y_i) \leq \sum_{i=1}^n q_i^2,$$

e la dimostrazione è completa. \square

4.2 Un'applicazione alla finanza: il modello binomiale

Consideriamo un mercato finanziario molto semplificato, in cui sia presente un unico titolo rischioso, ad esempio un'azione, e in cui non vi sia inflazione e non vi siano costi di transazione. Sia $X_0 > 0$ il valore odierno del titolo, che viene aggiornato quotidianamente. Indichiamo con X_1, X_2, \dots i valori del titolo nei giorni successivi, e assumeremo che $X_n > 0$ per ogni n .

Gli istituti finanziari, ad esempio le banche, offrono ai loro clienti le cosiddette *opzioni*. Un esempio è l'*opzione call europea*, che dà all'acquirente il diritto di acquistare dall'istituto finanziario al giorno T una unità del titolo ad un prezzo $f(X_T)$, dove $f : (0, \infty) \rightarrow \mathbb{R}$ è la funzione definita da

$$f(x) := \begin{cases} K & \text{se } x \geq K \\ x & \text{se } x < K \end{cases}, \quad (4.4)$$

dove K è un prezzo massimo prefissato. In questo modo il cliente “si assicura” contro il rischio di dover acquistare il titolo a un prezzo troppo elevato. I parametri dell'opzione T e K sono detti rispettivamente *scadenza* e *prezzo di esercizio*.

Analizziamo più in dettaglio l'opzione europea (4.4). Se il valore del titolo al giorno T è minore di K , non c'è nessun problema per l'istituto finanziario: esso acquisterà sul mercato un'unità del titolo, che rivenderà al cliente allo stesso prezzo. Se invece il valore del titolo al giorno T è maggiore di K , l'istituto finanziario sarà costretto a vendere al cliente al prezzo K un'unità del titolo il cui valore di mercato è $X_T > K$, rimettendoci dunque $(X_T - K)$. Possiamo quindi affermare che il *valore dell'opzione* (per il cliente) è dato dalla funzione

$$\pi(X_T) = X_T - f(X_T) = (X_T - K)^+ = \begin{cases} (X_T - K) & \text{se } X_T \geq K \\ 0 & \text{se } X_T < K \end{cases}. \quad (4.5)$$

Per una fissata scadenza T si possono considerare molteplici opzioni, corrispondenti a diverse scelte della funzione $\pi(\cdot)$. Nel seguito, ci prenderemo la libertà di considerare una arbitraria funzione *positiva* $\pi : (0, \infty) \rightarrow [0, \infty)$ e scriveremo per brevità “l'opzione π ” invece di “l'opzione il cui valore è $\pi(X_T)$ ”. Suggeriamo comunque di tenere a mente l'esempio (4.5).

Una delle domande fondamentali in finanza è: qualè il *giusto prezzo* V_0 a cui l'istituto dovrebbe vendere un'opzione π ? La domanda può apparire vaga, ma un modo possibile per rispondere consiste nel considerare una *strategia di copertura*: l'istituto finanziario può cercare di investire il denaro V_0 al fine di produrre al giorno T un capitale esattamente uguale a $\pi(X_T)$, a prescindere da quale sia stata l'evoluzione del prezzo del titolo, “coprendo” in questo modo il valore dell'opzione. Non è affatto ovvio che ciò sia possibile! Tuttavia, nel caso in cui esista effettivamente una tale strategia di investimento, corrispondente ad un unico valore di V_0 , tale valore può essere considerato il “prezzo giusto” dell'opzione.

Descriviamo più precisamente in che cosa consiste una strategia d'investimento. Supponiamo che al momento di vendita dell'opzione l'istituto finanziario utilizzi un capitale V_0 per attivare un *portafoglio*, di cui una parte investita in a_0 unità del titolo rischioso, la parte restante c_0 investita in un titolo non rischioso, cioè a tasso di rendimento costante. Per semplicità di calcolo assumiamo che tale tasso sia uguale a zero. In altre parole

$$V_0 = a_0 X_0 + c_0.$$

Il giorno successivo, essendo variato da X_0 a X_1 il prezzo del titolo rischioso, il capi-

tale diventa

$$V_1 = a_0X_1 + c_0 = V_0 + a_0(X_1 - X_0).$$

A questo punto l'istituto finanziario può modificare la porzione di capitale investita nel titolo rischioso, passando da a_0 ad a_1 unità possedute, con il rimanente $c_1 = V_1 - a_1X_1$ mantenuta a rendimento costante. Dopo due giorni il capitale diventa:

$$V_2 = a_1X_2 + c_1 = V_1 + a_1(X_2 - X_1) = V_0 + a_0(X_1 - X_0) + a_1(X_2 - X_1).$$

Iterando il procedimento, si ottiene che il capitale dopo n giorni è

$$V_n = V_0 + \sum_{i=0}^{n-1} a_i(X_{i+1} - X_i). \quad (4.6)$$

I coefficienti a_0, a_1, \dots, a_{T-1} rappresentano la strategia di investimento. Affinché si tratti di una strategia effettivamente implementabile, la scelta di a_n può dipendere dall'evoluzione del prezzo solo fino all'istante n , cioè $a_n = a_n(X_0, X_1, \dots, X_n)$. Si parla in tal caso di una *strategia non-anticipativa*, dal momento che in ogni giorno la strategia si basa solo sull'informazione effettivamente disponibile. Analogamente, il capitale iniziale $V_0 = V_0(X_0)$ può dipendere solo dal valore iniziale X_0 del titolo. Sottolineiamo che, assegnati il capitale iniziale V_0 e la strategia non-anticipativa a_0, \dots, a_{T-1} , risulta determinato il capitale V_n per ogni giorno $n \leq T$, grazie alla relazione (4.6), e si ha $V_n = V_n(X_0, X_1, \dots, X_n)$.

Osservazione 4.6. La scelta di a_i può essere tale che la quantità $c_i = V_i - a_iX_i$ di capitale investita nel titolo non rischioso sia *negativa*: ciò corrisponde semplicemente a prendere a prestito del denaro. In principio anche le quantità a_i possono essere negative, il che corrisponde a una “vendita allo scoperto” del titolo. \square

Possiamo ora definire precisamente la nozione di strategia di copertura.

Definizione 4.7. Consideriamo un'opzione $\pi(\cdot) \geq 0$ di scadenza T . Per un dato capitale iniziale V_0 , si dice *strategia di copertura* una strategia non-anticipativa a_0, a_1, \dots, a_{T-1} tale che, per qualunque evoluzione X_1, \dots, X_T del prezzo, si abbia

$$\begin{aligned} V_n(X_0, X_1, \dots, X_n) &\geq 0, & \forall n \in \{0, \dots, T\}, \\ V_T(X_0, X_1, \dots, X_T) &= \pi(X_T). \end{aligned} \quad (4.7)$$

Nel caso in cui vi sia un'unica scelta del capitale iniziale $V_0 = V_0(X_0)$ per cui esiste una strategia di copertura, diremo che la funzione $V_0(\cdot)$ è il *prezzo* dell'opzione.

Stabilire l'esistenza della strategia per la copertura dell'opzione costituisce un problema di grande rilevanza economica. Per poterlo affrontare, è necessario fare qualche assunzione sull'evoluzione del titolo X_n . Definiamo le quantità

$$R_n := \frac{X_n}{X_{n-1}}, \quad \text{per } n = 1, 2, \dots, T, \quad (4.8)$$

che descrivono le variazioni *relative* del prezzo del titolo rispetto al giorno precedente. Faremo l'ipotesi (molto forte) che le quantità R_n possano assumere esattamente due valori c e C , con $0 < c < 1 < C$. Ciò significa che ogni giorno il prezzo del titolo può soltanto contrarsi di un fattore c , oppure dilatarsi di un fattore C . Si tratta naturalmente di un'assunzione poco realistica, che ha però il vantaggio di prestarsi ad una trattazione matematica elegante ed efficace. Inoltre, molti dei risultati che otterremo possono essere estesi a modelli di mercato più ricchi e interessanti. Si noti che, assumendo che $X_0 > 0$, segue che $X_n = R_n R_{n-1} \cdots R_1 X_0 > 0$ per ogni n .

Osservazione 4.8. Sarebbe naturale descrivere le quantità R_n in termini di variabili aleatorie, assegnandone un'opportuna distribuzione. Come vedremo, questa non gioca alcun ruolo nel nostro problema. \square

Possiamo enunciare il risultato principale di questo paragrafo.

Teorema 4.9. *Per ogni opzione $\pi(\cdot) \geq 0$ di scadenza T , c'è un'unico capitale iniziale $V_0(\cdot)$ per cui esiste una strategia di copertura a_0, \dots, a_{T-1} , ed essa è unica. In particolare, $V_0(\cdot)$ è il prezzo dell'opzione.*

Dimostrazione. Fissiamo un capitale iniziale $V_0(\cdot)$ e consideriamo una strategia non-anticipativa a_0, \dots, a_{T-1} . Noto il valore X_{T-1} del prezzo del titolo, il prezzo X_T del giorno successivo può assumere i due valori cX_{T-1} e CX_{T-1} . Ricordando (4.6), la strategia in esame copre l'opzione, ossia $\pi(X_T) = V_T$, se e solo se

$$\pi(X_T) = V_{T-1} + a_{T-1}(X_T - X_{T-1}),$$

sia nel caso $X_T = cX_{T-1}$ che nel caso $X_T = CX_{T-1}$. In altre parole, si deve avere

$$\begin{aligned}\pi(cX_{T-1}) &= V_{T-1} - a_{T-1}(1-c)X_{T-1} \\ \pi(CX_{T-1}) &= V_{T-1} + a_{T-1}(C-1)X_{T-1},\end{aligned}$$

dalla quale si ricava

$$a_{T-1} = \frac{\pi(CX_{T-1}) - \pi(cX_{T-1})}{(C-c)X_{T-1}} \quad (4.9)$$

$$V_{T-1} = \frac{1-c}{C-c}\pi(CX_{T-1}) + \frac{C-1}{C-c}\pi(cX_{T-1}). \quad (4.10)$$

Dunque, il capitale V_{T-1} al giorno $T-1$ e la strategia a_{T-1} sono determinati in modo *univoco* dalla richiesta di copertura dell'opzione. Sottolineiamo tuttavia che V_{T-1} non è un "parametro libero" della strategia: esso è infatti determinato dai valori V_0, a_0, \dots, a_{T-2} tramite la relazione (4.6). Quindi, mentre la relazione (4.9) fornisce una definizione esplicita di a_{T-1} come funzione di X_{T-1} , la relazione (4.10) su V_{T-1} costituisce piuttosto un *vincolo implicito* su V_0, a_0, \dots, a_{T-2} (che, per il momento, non è ovvio che sia possibile soddisfare!).

A questo punto il procedimento si può iterare. Per coprire l'opzione è necessario che il capitale V_{T-1} al giorno $T-1$ soddisfi la relazione (4.10). Con un abuso di notazione, d'ora in avanti *indichiamo con $V_{T-1}(X_{T-1})$ la funzione esplicita di X_{T-1}* data dal membro destro in (4.10). Ragionando come sopra, grazie a (4.6), il capitale V_{T-2} e la strategia a_{T-2} devono soddisfare le relazioni

$$\begin{aligned} V_{T-1}(cX_{T-2}) &= V_{T-2} - a_{T-2}(1-c)X_{T-2} \\ V_{T-1}(CX_{T-2}) &= V_{T-2} + a_{T-2}(C-1)X_{T-2}, \end{aligned}$$

da cui si ottiene

$$a_{T-2} = \frac{V_{T-1}(CX_{T-2}) - V_{T-1}(cX_{T-2})}{(C-c)X_{T-2}} \quad (4.11)$$

$$V_{T-2} = \frac{1-c}{C-c}V_{T-1}(CX_{T-2}) + \frac{C-1}{C-c}V_{T-1}(cX_{T-2}). \quad (4.12)$$

Dunque imporre la relazione $V_T = \pi(X_T)$ equivale a richiedere che a_{T-1} e a_{T-2} siano dati rispettivamente dalle formule (4.9), (4.11) e il capitale V_{T-2} soddisfi la relazione (4.12) (che è una condizione implicita su V_0, a_0, \dots, a_{T-3}); queste relazioni garantiscono inoltre che V_{T-1} soddisfi la relazione (4.10). D'ora in poi indicheremo con $V_{T-2}(X_{T-2})$ la funzione data dal membro destro di (4.12).

Iterando gli argomenti precedenti, concludiamo che una strategia è tale che $V_T = \pi(X_T)$ se e solo se per ogni $n \in \{0, 1, \dots, T-1\}$ sono soddisfatte le seguenti relazioni ricorsive:

$$a_n(X_n) = \frac{V_{n+1}(CX_n) - V_{n+1}(cX_n)}{(C-c)X_n} \quad (4.13)$$

$$V_n(X_n) = \frac{1-c}{C-c}V_{n+1}(CX_n) + \frac{C-1}{C-c}V_{n+1}(cX_n) \quad (4.14)$$

$$V_T(X_T) = \pi(X_T). \quad (4.15)$$

Si noti che la relazione (4.14) può essere effettivamente soddisfatta, partendo da (4.15) e procedendo "a ritroso", definendo le funzioni V_n per $n = T-1, T-2, \dots, 0$. Tali funzioni determinano univocamente i valori $a_n(X_n)$ grazie a (4.13). Per costruzione, il capitale V_n associato alla strategia a_0, a_1, \dots, a_{T-1} e al capitale iniziale V_0 , appena determinati, soddisfa le relazioni (4.14), (4.15). In particolare $V_T = \pi \geq 0$, da cui si mostra facilmente per induzione a ritroso che $V_n \geq 0$ per ogni $0 \leq n \leq T$. Questo significa che le relazioni (4.7) sono effettivamente verificate!

In definitiva, abbiamo mostrato che esiste un'unica scelta del capitale iniziale $V_0(X_0)$, determinata dalle relazioni (4.14), (4.15), per cui esiste una strategia a_0, a_1, \dots, a_{T-1} tale che $V_T = \pi(X_T)$. Tale strategia è univocamente determinata, grazie a (4.13), e si ha $V_n \geq 0$ per ogni $n \in \{0, \dots, T\}$; pertanto abbiamo determinato l'unica strategia di copertura, e V_0 è il prezzo dell'opzione. \square

Osservazione 4.10. Vale la pena di osservare che, se la funzione $\pi : (0, \infty) \rightarrow [0, \infty)$ è crescente, come per l'opzione call europea (4.5), la strategia di copertura non com-

porta vendite allo scoperto, ossia $a_n \geq 0$ per ogni $n \in \{0, \dots, T\}$. Si noti infatti che le equazioni (4.13), (4.14), (4.15) sono formule ricorsive, le cui soluzioni sono funzioni di una variabile $a_n(x), V_n(x)$, che riscriviamo per chiarezza come segue:

$$a_n(x) = \frac{V_{n+1}(Cx) - V_{n+1}(cx)}{(C - c)x} \quad (4.16)$$

$$V_n(x) = \frac{1 - c}{C - c} V_{n+1}(Cx) + \frac{C - 1}{C - c} V_{n+1}(cx) \quad (4.17)$$

$$V_T(x) = \pi(x). \quad (4.18)$$

Se $V_T(x) = \pi(x)$ è una funzione crescente di $x > 0$, è immediato verificare per induzione “all’indietro” su n che $V_n(x)$ è una funzione crescente di $x > 0$ per $n \leq T - 1$. Essendo $Cx > cx$ per ogni $x > 0$, segue $a_n(x) \geq 0$ per ogni $0 \leq n \leq T - 1$. \square

Abbiamo dunque determinato il prezzo dell’opzione, come funzione $V_0(x_0)$ del valore $X_0 = x_0$ del titolo al momento in cui l’opzione viene acquistata, e la strategia autofinanziante che copre l’opzione. Sottolineiamo che non è stato necessario fare nessuna ipotesi sulla distribuzione delle variabili aleatorie (X_1, X_2, \dots, X_T) : in effetti, non è nemmeno necessario descrivere tali quantità come variabili aleatorie! La sola ipotesi (fondamentale) alla base della dimostrazione del Teorema 4.9 è la richiesta che il valore X_{n+1} del titolo al giorno $n + 1$ sia necessariamente cX_n oppure CX_n .

Tuttavia, il prezzo $V_0(x_0)$ determinato ammette un’interpretazione probabilistica fondamentale, che si può estendere a modelli di mercato assai più complessi, e che ha grande rilevanza sia teorica che applicativa. Ricordiamo la definizione (4.8) delle “variabili di aggiornamento” R_n e osserviamo che, se $X_0 = x_0$, allora il valore del titolo al giorno n è dato da $X_n = R_n R_{n-1} \cdots R_1 x_0$.

Teorema 4.11. *Siano Q_1, Q_2, \dots, Q_T variabili aleatorie indipendenti, che assumono soltanto i valori c e C , e di media $E(Q_n) = 1$. Allora la soluzione $V_n(x)$ dell’equazione ricorsiva (4.17) è, per $n \leq T - 1$,*

$$V_n(x) = E[\pi(Q_T \cdot Q_{T-1} \cdots Q_{n+1} x)].$$

In particolare, il prezzo dell’opzione è

$$V_0(x_0) = E[\pi(Q_T \cdot Q_{T-1} \cdots Q_1 x_0)].$$

Osservazione 4.12. Posto $Z_n = Q_n Q_{n-1} \cdots Q_1 x$, il prezzo dell’opzione è dunque $V_0 = E[\pi(Z_T)]$. La variabile aleatoria Z_n si può interpretare come il valore al giorno n di un titolo avente lo stesso valore iniziale x_0 del titolo reale, ma che evolve in un mercato *fittizio* in cui le variabili aleatorie di “aggiornamento” sono indipendenti e hanno media 1 (*mercato neutrale*). Il prezzo dell’opzione è dato dal valor medio del valore dell’opzione $\pi(Z_T)$ calcolato non rispetto al mercato “reale”, ma rispetto al corrispondente mercato neutrale. \square

Dimostrazione (del Teorema 4.11). Si noti, anzitutto, che le richieste di assumere solo i valori c e C e di avere media 1 caratterizzano univocamente la distribuzione (comune) delle Q_n . Infatti dev'essere

$$\begin{aligned} E(Q_n) &= cP(Q_n = c) + CP(Q_n = C) = 1 \\ P(Q_n = c) + P(Q_n = C) &= 1 \end{aligned}$$

da cui segue che

$$P(Q_n = c) = \frac{C-1}{C-c}, \quad P(Q_n = C) = \frac{1-c}{C-c}.$$

In quanto segue, per una successione $(a_n)_{n \geq 0}$ a valori reali, scriveremo a_m^T in luogo di $(a_m, a_{m+1}, \dots, a_T)$. Per l'indipendenza delle Q_n

$$P(Q_{n+1}^T = q_{n+1}^T) = P(Q_{n+1} = q_{n+1})P(Q_{n+2}^T = q_{n+2}^T),$$

per ogni $q_{n+1}, q_{n+2}, \dots, q_T \in \{c, C\}$. Posto

$$W_n(x) := E[\pi(Q_T \cdot Q_{T-1} \cdots Q_{n+1}x)],$$

e convenendo che $\prod_{k=T+1}^T a_k \equiv 1$, per ogni $n \leq T-1$ si ha:

$$\begin{aligned} W_n(x) &= E \left[\pi \left(x \prod_{k=n+1}^T Q_k \right) \right] \\ &= \sum_{q_{n+1}, q_{n+2}, \dots, q_T \in \{c, C\}} \pi \left(x \prod_{k=n+1}^T q_k \right) P(Q_{n+1}^T = q_{n+1}^T) \\ &= \sum_{q_{n+1} \in \{c, C\}} P(Q_{n+1} = q_{n+1}) \sum_{q_{n+2}, \dots, q_T \in \{c, C\}} \pi \left(x \prod_{k=n+1}^T q_k \right) P(Q_{n+2}^T = q_{n+2}^T) \\ &= \frac{C-1}{C-c} \sum_{q_{n+2}, \dots, q_T \in \{c, C\}} \pi \left(cx \prod_{k=n+2}^T q_k \right) P(Q_{n+2}^T = q_{n+2}^T) \\ &\quad + \frac{1-c}{C-c} \sum_{q_{n+2}, \dots, q_T \in \{c, C\}} \pi \left(Cx \prod_{k=n+2}^T q_k \right) P(Q_{n+2}^T = q_{n+2}^T) \\ &= \frac{C-1}{C-c} W_{n+1}(cx) + \frac{1-c}{C-c} W_{n+1}(Cx). \end{aligned}$$

Dunque $W_n(x)$ risolve l'equazione ricorsiva (4.17) e, per definizione $W_T(x) \equiv V_T(x)$. Pertanto $W_n(x) \equiv V_n(x)$ per ogni $n = 0, 1, \dots, T$. \square

4.3 Il problema del collezionista di figurine

Completare un album di figurine, lo sappiamo per esperienza, è un'impresa difficile e a volte costosa. In questo paragrafo vedremo un modello semplificato di raccolta di figurine che si presta ad uno studio quantitativo assai preciso.

Supponiamo che un album consista di N figurine, che identifichiamo con gli elementi dell'insieme $\{1, 2, \dots, N\}$. Le figurine vengono vendute singolarmente, ed ogni figurina acquistata, indipendentemente dai precedenti acquisti, ha ugual probabilità di essere una delle N figurine possibili.

Rendiamo più precise queste specifiche. Sia $(X_n)_{n \geq 1}$ una successione di variabili aleatorie che rappresentano la sequenza delle figurine acquistate. Secondo le specifiche precedenti, le X_n sono indipendenti, con distribuzione uniforme discreta sull'insieme $\{1, 2, \dots, N\}$. L'insieme aleatorio

$$A_n := \{X_1, X_2, \dots, X_n\} \subseteq \{1, 2, \dots, N\},$$

consiste delle figurine distinte ottenute nei primi n acquisti. Si noti che i valori assunti X_1, \dots, X_n possono essere ripetuti, pertanto tipicamente A_n ha cardinalità strettamente minore di n . Definiamo $T_0 = 0$ e, per $k = 1, 2, \dots, N$,

$$T_k := \min\{n \geq 1 : |A_n| = k\}.$$

In altre parole, T_k è il numero di acquisti necessari per ottenere k figurine distinte; in particolare, T_N è il numero di acquisti necessari a completare l'album. Si noti che T_k è una variabile aleatoria, che a priori può assumere anche il valore $+\infty$, definita sullo stesso spazio di probabilità su cui sono definite le X_n . Infine, per $k = 1, 2, \dots, N$, sia

$$Y_k := T_k - T_{k-1}.$$

Il cuore dell'analisi è dato dal seguente risultato (che mostra, tra l'altro, che le variabili aleatorie T_k sono quasi certamente finite: $P(T_k < +\infty) = 1$ per ogni k).

Proposizione 4.13. *Le variabili aleatorie Y_1, Y_2, \dots, Y_N sono indipendenti, con distribuzioni marginali*

$$Y_k \sim \text{Geo}\left(1 - \frac{k-1}{N}\right). \quad (4.19)$$

Dimostrazione. L'idea della dimostrazione è piuttosto semplice. Supponiamo di avere appena acquistato la $(k-1)$ -esima figurina distinta, cioè di aver eseguito T_{k-1} acquisti. Rimangono dunque $N - (k-1)$ figurine mancanti. Considerato che ogni figurina viene acquistata con la stessa probabilità, il successivo acquisto fornirà una “nuova” figurina, cioè distinta dalle precedenti, con probabilità $\frac{N-k+1}{N}$. Perciò il numero Y_k di acquisti necessari per trovare una nuova figurina è il tempo di primo successo in uno schema di prove ripetute indipendenti con probabilità di successo $(1 - \frac{k-1}{N})$,

da cui segue (4.19). Forniamo ora un argomento più rigoroso, da cui segue anche l'indipendenza delle Y_k .

Si noti che, ovviamente $Y_1 \equiv 1$. Si tratta perciò di dimostrare che la densità congiunta di (Y_2, \dots, Y_N) è quella di $N - 1$ variabili aleatorie indipendenti con distribuzione data da (4.19), cioè

$$P(Y_2 = y_2, Y_3 = y_3, \dots, Y_N = y_N) = \prod_{k=2}^N \left(1 - \frac{k-1}{N}\right) \left(\frac{k-1}{N}\right)^{y_k-1}, \quad (4.20)$$

per ogni $y_2, y_3, \dots, y_N \geq 1$. Sia $\Sigma = (\Sigma_1, \dots, \Sigma_N)$ la sequenza delle N figurine “distinte”, ordinate nell'ordine cronologico in cui sono acquistate, ossia $\Sigma_i := X_{T_i}$. Osserviamo che Σ_N è una variabile aleatoria a valori in S_N , il gruppo delle permutazioni su $\{1, \dots, N\}$. Sommando sui valori σ assunti da Σ , otteniamo l'uguaglianza di eventi

$$\{Y_2 = y_2, \dots, Y_N = y_N\} = \bigcup_{\sigma \in S_N} G_\sigma, \quad (4.21)$$

dove $G_\sigma := \{Y_2 = y_2, \dots, Y_N = y_N\} \cap \{\Sigma = \sigma\}$. Notiamo che

$$G_\sigma = \bigcap_{k=1}^N G_{\sigma,k},$$

dove $G_{\sigma,1} := \{X_1 = \sigma(1)\}$ mentre per $k \geq 2$ definiamo

$$G_{\sigma,k} = \{X_n \in \{\sigma(1), \dots, \sigma(k-1)\} \text{ per } n = t_{k-1} + 1, \dots, t_k - 1; X_{t_k} = \sigma(k)\},$$

avendo posto $y_1 := 1$ e $t_h := y_1 + \dots + y_h$. Dalle ipotesi sulle variabili aleatorie X_k

$$P(G_{\sigma,k}) = \frac{1}{N} \left(\frac{k-1}{N}\right)^{y_k-1}$$

(si ricordi che $y_1 := 1$). Inoltre:

- gli eventi G_σ sono disgiunti al variare di σ ;
- per ogni $\sigma \in S_N$ fissato, gli eventi $(G_{\sigma,k})_{k=1}^N$ sono indipendenti, per la Proposizione 3.39, in quanto funzione di sequenze disgiunte della successione (X_n) .
Di conseguenza $P(G_\sigma) = \prod_{k=1}^N P(G_{\sigma,k})$.

Ricordando (4.21), otteniamo dunque

$$\begin{aligned} P(Y_2 = y_2, Y_3 = y_3, \dots, Y_N = y_N) &= \sum_{\sigma \in S_N} \frac{1}{N} \prod_{k=2}^N \frac{1}{N} \left(\frac{k-1}{N}\right)^{y_k-1} \\ &= N! \frac{1}{N} \prod_{k=2}^N \frac{1}{N} \left(\frac{k-1}{N}\right)^{y_k-1} = \prod_{k=2}^N \left(1 - \frac{k-1}{N}\right) \left(\frac{k-1}{N}\right)^{y_k-1}, \end{aligned}$$

che completa la dimostrazione. \square

A questo punto, essendo

$$T_N = 1 + Y_2 + \cdots + Y_N,$$

valor medio e varianza di T_N si possono calcolare facilmente: ricordando le formule (3.87), (3.88) per valor medio e varianza delle variabili aleatorie geometriche, si ha

$$E(T_N) = 1 + \sum_{k=2}^N E(Y_k) = 1 + \sum_{k=2}^N \frac{N}{N-k+1} = N \sum_{k=1}^N \frac{1}{k}.$$

Analogamente,

$$\begin{aligned} \text{Var}(T_N) &= \sum_{k=2}^N \text{Var}(Y_k) = N \sum_{k=2}^N \frac{k-1}{(N-k+1)^2} \\ &= N^2 \sum_{k=2}^N \frac{1}{(N-k+1)^2} - N \sum_{k=2}^N \frac{1}{N-k+1} = N^2 \sum_{k=1}^{N-1} \frac{1}{k^2} - N \sum_{k=1}^{N-1} \frac{1}{k}, \end{aligned} \quad (4.22)$$

avendo sfruttato, nella prima uguaglianza in (4.22), l'indipendenza delle Y_k . Al fine di ottenere il comportamento preciso di $E(T_N)$ e $\text{Var}(T_N)$ nel limite per $N \rightarrow +\infty$, enunciamo, senza dimostrare, i seguenti risultati sulle serie armoniche:

$$\sum_{k=1}^N \frac{1}{k} = \log N + \gamma + \frac{1}{2N} + o(1/N),$$

dove $\gamma \simeq 0.577$ è chiamata *costante di Eulero-Mascheroni*, e

$$\sum_{k=1}^N \frac{1}{k^2} = \frac{\pi^2}{6} + \frac{1}{2N^2} + o(1/N^2),$$

da cui si ottengono le seguenti espressioni asintotiche:

$$E(T_N) = N \log N + \gamma N + \frac{1}{2} + o(1)$$

$$\text{Var}(T_N) = \frac{\pi^2}{6} N^2 - N \log N - \gamma N + o(1).$$

Mostriamo infine un risultato che fornisce informazioni sulla distribuzione di T_N . La notazione $\lceil x \rceil$ denota il più piccolo intero maggiore o uguale al numero reale x .

Proposizione 4.14. *Per ogni $c > 1$ e per ogni $N \geq 1$*

$$P(T_N > \lceil N \log N + cN \rceil) \leq e^{-c}, \quad (4.23)$$

$$P(T_N < N \log N - cN) \leq e^{-c}, \quad (4.24)$$

Dimostrazione. Cominciamo con il dimostrare (4.23). Sia $k := \lceil N \log N + cN \rceil$ e, per $i \in \{1, 2, \dots, N\}$, consideriamo gli eventi

$$B_i := \{i \notin \{X_1, X_2, \dots, X_k\}\}.$$

In altre parole, l'evento B_i corrisponde all'affermazione “la figurina i non compare tra le prime k figurine acquistate”. Si noti che:

$$\{T_N > k\} = \bigcup_{i=1}^N B_i.$$

Usando il fatto che, per l'indipendenza delle X_n

$$P(B_i) = P\left(\bigcap_{n=1}^k \{X_n \neq i\}\right) = \left(\frac{N-1}{N}\right)^k,$$

otteniamo

$$P(T_N > k) \leq \sum_{i=1}^N P(B_i) = N \left(\frac{N-1}{N}\right)^k = N \left(1 - \frac{1}{N}\right)^k \leq N e^{-k/N}.$$

dove abbiamo anche usato il fatto che $(1-x) \leq e^{-x}$ per ogni $x \in \mathbb{R}$. Dunque, essendo $k \geq N \log N + cN$, si ha

$$P(T_N > k) \leq N e^{-\log N - c} = e^{-c}.$$

Passiamo ora alla dimostrazione di (4.24). Sia $s > 0$ e $t := N \log N - cN$. Applicando la disuguaglianza di Markov alla variabile aleatoria $X := e^{-sT_N}$, abbiamo

$$P(T_N < t) = P(e^{-sT_N} > e^{-st}) \leq e^{st} E(e^{-sT_N}). \quad (4.25)$$

Ricordiamo che T_N è somma di variabili aleatorie geometriche indipendenti:

$$T_N = 1 + Y_2 + \dots + Y_N, \quad Y_k \sim Ge\left(1 - \frac{k-1}{N}\right),$$

da cui, ricordando la formula (3.86) per la funzione generatrice dei momenti delle variabili Geometriche, segue che

$$E(e^{-sT_N}) = e^{-s} \prod_{k=2}^N E(e^{-sY_k}) = \prod_{k=1}^N \frac{e^{-s} - \frac{k-1}{N}}{e^{-s} - \frac{k-1}{N}} \quad (4.26)$$

Usiamo ora (4.25) e (4.26) con $s := 1/N$, per cui $e^{st} = Ne^{-c}$ (si ricordi che $t = N \log N - cN$). Applicando la disuguaglianza $e^{1/N} \geq 1 + \frac{1}{N}$, si ottiene

$$\begin{aligned} P(T_N < N \log N - cN) &\leq e^{-c} N \prod_{k=1}^N \frac{\frac{N-k+1}{N}}{e^{1/N} - \frac{k-1}{N}} \leq e^{-c} N \prod_{k=1}^N \frac{\frac{N-k+1}{N}}{1 + \frac{1}{N} - \frac{k-1}{N}} \\ &= e^{-c} N \prod_{k=1}^N \frac{N-k+1}{N-k+2} = e^{-c} \frac{N}{N+1} \leq e^{-c}, \end{aligned}$$

avendo usato nell'ultima uguaglianza il fatto che il prodotto è telescopico. □

Essendo $N \ll N \log N$ per N grande, le relazioni (4.23) e (4.24) esprimono il fatto che, con probabilità vicino ad uno, il numero di acquisti necessari a completare l'album è approssimativamente $N \log N$, e la probabilità di “deviazioni” di ordine N da tale valore viene stimata in modo esplicito.

Chiudiamo con un'interpretazione alternativa dei risultati visti. Supponiamo di formare un gruppo di persone scelte casualmente, una dopo l'altra, fermandosi non appena ci sia una persona nata in ciascuno degli $N = 365$ giorni dell'anno. Se i compleanni sono uniformemente distribuiti tra i giorni dell'anno, il gruppo è composto in media da $E(T_N) \approx N \log N \approx 2153$ persone.

4.4 Mescolare un mazzo di carte

Mescolare accuratamente un mazzo di carte è fondamentale per l'equità dei giochi che con le carte si svolgono. Ciò è di particolare rilevanza nei casinò, dove vengono usate macchine automatiche: se il mescolamento non fosse accurato, favorendo certi ordinamenti invece di altri, un giocatore abile che ne fosse a conoscenza potrebbe sfruttare a proprio vantaggio tale inaccuratezza. In questo paragrafo illustriamo e analizziamo un metodo di mescolamento estremamente semplice, in cui un “singolo mescolamento” consiste nell'estrarre una carta a caso dal mazzo, mettendola in cima al mazzo stesso. Tale modello consente un'analisi relativamente semplice; inoltre tanto le tecniche quanto, a livello qualitativo, i risultati, sono simili a quelli di modelli più realistici.

Come vedremo, la formulazione e l'analisi del modello sono strettamente legate al problema del collezionista di figurine. Sia N il numero di carte, che verranno etichettate con gli elementi dell'insieme $\{1, 2, \dots, N\}$. L'insieme dei possibili ordinamenti del mazzo verrà identificato con S_N , l'insieme delle permutazioni di $\{1, 2, \dots, N\}$: data $\sigma \in S_N$, la carta che si trova in cima al mazzo è $\sigma(1)$, quella successiva è $\sigma(2)$, e così via. Più formalmente, $\sigma(i) = j$ significa che nella posizione i (partendo dalla cima del mazzo) c'è la carta j , per ogni $1 \leq i, j \leq n$. Dunque la permutazione σ associa a ogni “posizione” la carta corrispondente, mentre la permutazione inversa σ^{-1} associa a ogni carta la posizione corrispondente.

Sia $(U_n)_{n \geq 1}$ una successione di variabili aleatorie indipendenti, con distribuzione uniforme discreta su $\{1, 2, \dots, N\}$. Esse rappresentano le *posizioni* delle carte estratte

dal mazzo nei successivi mescolamenti: all' n -esimo mescolamento, la carta in posizione U_n viene presa dal mazzo e posta in cima allo stesso. Denotiamo con Σ_n l'ordinamento del mazzo dopo n mescolamenti. Si ha che $(\Sigma_n)_{n \geq 0}$ è una successione di variabili aleatorie a valori in S_N , che possono essere definite ricorsivamente come segue, per $i = 1, 2, \dots, N$:

$$\Sigma_0(i) = i, \quad \Sigma_n(i) = \begin{cases} \Sigma_{n-1}(U_n) & \text{se } i = 1 \\ \Sigma_{n-1}(i-1) & \text{se } 1 < i \leq U_n, \\ \Sigma_{n-1}(i) & \text{se } U_n < i \leq N \end{cases} \quad \forall n \geq 1. \quad (4.27)$$

Si noti che (4.27) codifica proprio il mescolamento sopra descritto: se l'ordinamento del mazzo dopo $n-1$ mescolamenti è Σ_{n-1} , nell' n -esimo mescolamento si estrae la carta in posizione U_n , ossia la carta

$$X_n := \Sigma_{n-1}(U_n),$$

e la si mette in cima al mazzo, producendo l'ordinamento Σ_n .

Vale la pena di sottolineare che la successione $(X_n)_{n \geq 1}$ delle carte estratte successivamente dal mazzo è composta da variabili aleatorie *indipendenti con distribuzione uniforme discreta su* $\{1, 2, \dots, N\}$, analogamente alla successione delle loro posizioni $(U_n)_{n \geq 1}$. Infatti, qualunque sia l'ordinamento Σ_{n-1} del mazzo, scegliendo la posizione U_n in modo uniforme, la carta estratta $X_n = \Sigma_{n-1}(U_n)$ è una qualunque tra le N carte con la stessa probabilità, indipendentemente dalle carte precedentemente estratte. Lasciamo al lettore interessato la formalizzazione di questo argomento.

Introduciamo l'*insieme* aleatorio costituito dalle carte estratte nei primi n mescolamenti:

$$A_n := \{X_1, X_2, \dots, X_n\} \subseteq \{1, \dots, N\},$$

e osserviamo che l'ordinamento Σ_n può essere descritto nel modo seguente:

- le carte dell'insieme A_n occupano le prime posizioni del mazzo, nell'ordine inverso in cui sono state estratte;
- le carte non estratte, cioè appartenenti a $\{1, 2, \dots, N\} \setminus A_n$, occupano le ultime posizioni del mazzo, in ordine crescente (in quanto il loro ordine relativo iniziale non è stato modificato).

Come nel problema del collezionista di figurine, definiamo per $k \in \{1, \dots, N\}$ le variabili aleatorie

$$T_k := \min\{n \geq 1 : |A_n| = k\},$$

e

$$Y_k := T_k - T_{k-1}.$$

Tali variabili aleatorie soddisfano la Proposizione 4.13, che useremo a breve.

Ricordando che Σ_n è una variabile aleatoria a valori in S_N , indichiamone con Q_n la distribuzione e con q_n la relativa densità. Intuitivamente, possiamo dire che dopo n mescolamenti il mazzo è “ben mescolato” se Q_n è “vicina” alla probabilità uniforme U su S_N , ossia $U(\{\sigma\}) := u(s) = \frac{1}{N!}$ per ogni $\sigma \in S_N$. Più precisamente, usando le no-

tazioni introdotte nella Definizione 4.1, come misura della “bontà” del mescolamento useremo la distanza in variazione totale

$$d_n := d(q_n, u) = \frac{1}{2} \sum_{\sigma \in S_N} \left| q_n(\sigma) - \frac{1}{N!} \right|. \quad (4.28)$$

Nella parte restante di questo paragrafo dimostreremo il seguente risultato.

Teorema 4.15. *Siano $N \in \mathbb{N}$ e $c \in (0, \infty)$. Allora:*

- (i) *per $n > N \log N + cN$ si ha $d_n \leq e^{-c}$;*
- (ii) *per $n < N \log N - cN$ si ha $d_n \geq 1 - f_N(c)$, dove*

$$f_N(c) := \min_{m \in \{1, \dots, N\}} \left(\frac{1}{m!} + (m+1)e^{-c} \right), \quad (4.29)$$

da cui segue che $d_n \geq 1 - (c+5)e^{-c}$ se $c \geq 3$.

Dunque, se c è sufficientemente grande, $N \log N + cN$ mescolamenti sono sufficienti per essere vicini alla distribuzione uniforme, mentre $N \log N - cN$ non sono sufficienti. A titolo di esempio, per un mazzo da “Scala 40”, in cui $N = 108$, applicando il Teorema 4.15 si ha che $d_n \geq 0.6$ per $n \leq 181$, mentre $d_n \leq 0.05$ per $n \geq 830$. Tali stime sono in realtà abbastanza rozze, in quanto la transizione da valori “grandi” a valori “piccoli” per d_n avviene assai più rapidamente.

Il risultato del Teorema 4.15 diventa tuttavia assai significativo per valori grandi di N , per cui, ad esempio, $10N \ll N \log N$. In tal caso, si passa da valori di d_n vicini a 1 a valori di d_n vicini a 0 passando da $(N \log N - 5N)$ a $(N \log N + 5N)$ mescolamenti, quindi con una *variazione* del numero di mescolamenti pari a $10N$, che è molto piccola rispetto al numero totale di mescolamenti ($\sim N \log N$). Questo fenomeno di “transizione rapida”, detto *cut-off*, è comune a molte dinamiche stocastiche, in particolare a molti algoritmi di mescolamento effettivamente usati nei casinò.

La dimostrazione del Teorema 4.15 sarà preceduta da alcuni risultati intermedi. L’idea di base è piuttosto semplice, e può essere illustrata come segue. Se il numero di mescolamenti n è tale che $n < T_{N-m}$, le ultime m carte del mazzo sono disposte in ordine crescente; dato che la probabilità che ciò si verifichi per la distribuzione uniforme è $1/m!$, la distribuzione di Σ_n è “distante” da quella uniforme, se m è piccolo. Se invece $n > T_{N-1}$, l’ordine iniziale è stato del tutto “distrutto”, e vedremo che in questo caso le carte sono uniformemente disposte nel mazzo. Quindi, euristicamente, la transizione di d_n da valori vicini a 1 a valori vicini a 0 avviene tra T_{N-m} e T_{N-1} : sfruttando la Proposizione 4.14 si ottiene il Teorema 4.15.

Cominciamo con un risultato di carattere generale.

Lemma 4.16. *Siano P e Q due probabilità discrete su un insieme E , e siano $p(x) := P(\{x\})$, $q(x) := Q(\{x\})$ le relative densità. Allora*

$$d(p, q) = \sup_{A \subseteq E} |P(A) - Q(A)|.$$

Dimostrazione. Sia $A \subseteq E$. Si noti che $|P(A) - Q(A)| = |P(A^c) - Q(A^c)|$, perché $P(A^c) = 1 - P(A)$ e analogamente per Q , pertanto

$$\begin{aligned} |P(A) - Q(A)| &= \frac{1}{2} |P(A) - Q(A)| + \frac{1}{2} |P(A^c) - Q(A^c)| \\ &= \frac{1}{2} \left| \sum_{x \in A} [p(x) - q(x)] \right| + \frac{1}{2} \left| \sum_{x \in A^c} [p(x) - q(x)] \right| \\ &\leq \frac{1}{2} \sum_{x \in E} |p(x) - q(x)| = d(p, q), \end{aligned}$$

che dimostra

$$d(p, q) \geq \sup_{A \subseteq E} |P(A) - Q(A)|.$$

Per mostrare la disuguaglianza opposta, esibiamo $A \subseteq E$ per cui $P(A) - Q(A) = d(p, q)$. È sufficiente considerare $A := \{x \in E : p(x) \geq q(x)\}$, così che

$$\begin{aligned} P(A) - Q(A) &= \frac{1}{2} [P(A) - Q(A)] + \frac{1}{2} [Q(A^c) - P(A^c)] \\ &= \frac{1}{2} \sum_{x \in A} [p(x) - q(x)] + \frac{1}{2} \sum_{x \in A^c} [q(x) - p(x)] \\ &= \frac{1}{2} \sum_{x \in E} |p(x) - q(x)| = d(p, q), \end{aligned}$$

completando la dimostrazione. \square

Mostriamo ora rigorosamente che successivamente all'istante T_{N-1} l'ordine delle carte nel mazzo è uniforme.

Lemma 4.17. *Per ogni $\sigma \in S_N$ e per ogni $n \geq N - 1$*

$$P(\Sigma_n = \sigma \mid T_{N-1} \leq n) = \frac{1}{N!}.$$

Dimostrazione. Fissiamo $\sigma \in S_N$ e $n \geq N - 1$. Si noti che l'evento $\{\Sigma_n = \sigma, T_{N-1} = n\}$ è una funzione delle variabili aleatorie X_1, \dots, X_n , nel senso che conoscendo i valori assunti da tali variabili si sa con certezza se l'evento si è verificato oppure no.

Usando la notazione $\mathbf{x} = (x_1, x_2, \dots, x_n) \in \{1, 2, \dots, N\}^n$, possiamo dunque scrivere

$$\{\Sigma_n = \sigma, T_{N-1} \leq n\} = \bigcup_{\mathbf{x} \in A_n(\sigma)} \{X_1 = x_1, X_2 = x_2, \dots, X_n = x_n\},$$

dove $A_n(\sigma)$ denota l'insieme dei $\mathbf{x} \in \{1, 2, \dots, N\}^n$ per cui

$$\{X_1 = x_1, X_2 = x_2, \dots, X_n = x_n\} \subseteq \{\Sigma_n = \sigma, T_{N-1} \leq n\}.$$

Dato che la distribuzione congiunta di (X_1, \dots, X_n) è la probabilità uniforme sull'insieme $\{1, 2, \dots, N\}^n$ (perché?), si deduce che

$$P(\Sigma_n = \sigma, T_{N-1} \leq n) = \frac{|A_n(\sigma)|}{N^n}.$$

Se dimostriamo che $|A_n(\sigma)|$ non dipende da σ , segue che $P(\Sigma_n = \sigma, T_{N-1} \leq n)$ non dipende da σ , da cui la tesi segue facilmente (esercizio).

A tale scopo, denotando con id la permutazione identica, mostriamo che $|A_n(\sigma)| = |A_n(\text{id})|$ esibendo una corrispondenza biunivoca fra $A_n(\sigma)$ e $A_n(\text{id})$. Indicando con σ^{-1} è la permutazione inversa di σ , una biiezione $\varphi : A_n(\sigma) \rightarrow A_n(\text{id})$ è data da

$$\varphi(x_1, x_2, \dots, x_n) := (\sigma^{-1}(x_1), \sigma^{-1}(x_2), \dots, \sigma^{-1}(x_n)). \quad (4.30)$$

Per dimostrarlo, osserviamo che l'insieme $A_n(\sigma)$ consta di tutti e soli gli elementi $(x_1, \dots, x_n) \in \{1, \dots, N\}^n$ con le seguenti proprietà:

- (i) $|\{x_1, x_2, \dots, x_n\}| \geq N - 1$;
- (ii) per ogni $i \in \{1, \dots, N - 1\}$, $\sigma(i)$ è l' i -esima carta distinta nella sequenza “inversa” x_n, x_{n-1}, \dots, x_1 . Infatti, per le regole del mescolamento, la carta in cima al mazzo è l'ultima carta estratta, ossia $\sigma(1) = x_n$; la seconda carta del mazzo è la penultima carta estratta se essa è diversa dall'ultima, ossia $\sigma(2) = x_{n-1}$ se $x_{n-1} \neq x_n$; se invece $x_{n-1} = x_n$, si ha $\sigma(2) = x_{n-2}$ se $x_{n-2} \neq x_n$, ecc.

Si noti che, essendo σ una permutazione, se la proprietà (i) vale per \mathbf{x} , allora vale per $\varphi(\mathbf{x})$. Mostriamo quindi che se \mathbf{x} soddisfa la proprietà (ii) per σ , allora $\varphi(\mathbf{x})$ la soddisfa per id . Per ipotesi, l' i -esima carta distinta nella sequenza inversa x_n, x_{n-1}, \dots, x_1 è $\sigma(i) = x_{n-t_i}$, per un opportuno $t_i \in \{0, \dots, n\}$; ma allora l' i -esima carta distinta nella sequenza inversa $\sigma^{-1}(x_n), \dots, \sigma^{-1}(x_1)$ è necessariamente $\sigma^{-1}(x_{n-t_i}) = i = \text{id}(i)$, come segue da (4.30). Resta solo da mostrare che φ è effettivamente una biiezione: ciò segue dal fatto che ammette la seguente mappa inversa:

$$\varphi^{-1}(y_1, y_2, \dots, y_n) := (\sigma(y_1), \sigma(y_2), \dots, \sigma(y_n)),$$

come si verifica immediatamente. Questo conclude la dimostrazione. \square

Si può interpretare il Lemma 4.17 come segue. Supponiamo che il mescolamento sia eseguito da una macchina che è in grado di leggere e memorizzare le carte estratte (senza, ovviamente, che questa informazione sia accessibile ai giocatori). Se

la macchina smette di mescolare dopo aver estratto $N - 1$ carte distinte, cioè dopo T_{N-1} mescolamenti, la distribuzione risultante dell'ordinamento finale è *esattamente* uniforme, ossia la macchina ha eseguito un mescolamento *perfetto*.

Il Lemma 4.17 è il risultato chiave per dimostrare il seguente risultato.

Lemma 4.18. *Per ogni $n \geq N - 1$,*

$$d_n \leq P(T_{N-1} > n).$$

Dimostrazione. Usando il Lemma 4.17, otteniamo

$$\begin{aligned} P(\Sigma_n = \sigma) &= P(\Sigma_n = \sigma, T_{N-1} \leq n) + P(\Sigma_n = \sigma, T_{N-1} > n) \\ &= \frac{1}{N!} P(T_{N-1} \leq n) + P(\Sigma_n = \sigma, T_{N-1} > n), \end{aligned}$$

che possiamo riscrivere nella forma

$$P(\Sigma_n = \sigma) - \frac{1}{N!} = -\frac{1}{N!} P(T_{N-1} > n) + P(\Sigma_n = \sigma, T_{N-1} > n).$$

Per la disuguaglianza triangolare, da ciò segue che

$$\left| P(\Sigma_n = \sigma) - \frac{1}{N!} \right| \leq \frac{1}{N!} P(T_{N-1} > n) + P(\Sigma_n = \sigma, T_{N-1} > n),$$

e sommando su $\sigma \in S_N$, ricordando la definizione (4.28), otteniamo

$$2d_n = \sum_{\sigma \in S_N} \left| P(\Sigma_n = \sigma) - \frac{1}{N!} \right| \leq 2P(T_{N-1} > n),$$

che conclude la dimostrazione. \square

Lemma 4.19. *Per ogni $c > 0$ e $0 \leq m \leq N$*

$$P(T_{N-m} < N \log N - cN) \leq (m+1)e^{-c}.$$

Dimostrazione. Per $m = 0$ si tratta della disuguaglianza (4.24), dimostrata nella Proposizione 4.14. Per $m \geq 1$ la dimostrazione è identica, con la sola differenza che i prodotti in (4.26) si estendono per $k \leq N - m$ anziché $k \leq N$. Si trova perciò

$$P(T_{N-m} < N \log N - cN) \leq e^{-c} N \prod_{k=1}^{N-m} \frac{N-k+1}{N-k+2} = e^{-c} \frac{N(m+1)}{N+1} \leq (m+1)e^{-c},$$

come cercato. \square

Non abbiamo ora che da mettere assieme i risultati appena dimostrati.

Dimostrazione (del Teorema 4.15). Cominciamo col dimostrare la prima affermazione; prendiamo quindi $n > N \log N + cN$. Essendo $T_{N-1} < T_N$, applicando il Lemma 4.18 e la relazione (4.23), per $n > N \log N + cN$ si ha

$$d_n \leq P(T_{N-1} > n) \leq P(T_N > N \log N + cN) \leq e^{-c}.$$

Per dimostrare la seconda affermazione del Teorema 4.15, consideriamo gli ordinamenti del mazzo di carte in cui le ultime m carte sono in ordine crescente:

$$A_m := \{\sigma \in S_N : \sigma(N-m+1) < \sigma(N-m+2) < \dots < \sigma(N)\}.$$

Come abbiamo osservato in precedenza, se dopo n mescolamenti sono state estratte al più $N-m$ carte distinte, ossia se $T_{N-m} > n$ le ultime m carte del mazzo sono in ordine crescente, dunque

$$P(\Sigma_n \in A_m | T_{N-m} \geq n) = 1.$$

Perciò, ricordando che Q_n denota la distribuzione di Σ_n , per $n < N \log N - cN$

$$\begin{aligned} Q_n(A_m) &= P(\Sigma_n \in A_m) \geq P(\Sigma_n \in A_m | T_{N-m} \geq n) P(T_{N-m} \geq n) \\ &= P(T_{N-m} \geq n) \geq P(T_{N-m} \geq N \log N - cN) \geq 1 - (m+1)e^{-c}, \end{aligned}$$

dove l'ultima disuguaglianza segue dal Lemma 4.19. Inoltre, tramite un semplice argomento di conteggio, lasciato al lettore, si mostra che

$$U(A_m) = \frac{1}{m!},$$

dove denotiamo con U la probabilità uniforme su S_N . Perciò, per il Lemma 4.16,

$$d_n \geq Q_n(A_m) - U(A_m) \geq 1 - (m+1)e^{-c} - \frac{1}{m!}.$$

Questa disuguaglianza vale per ogni $m \in \{1, \dots, N\}$. Considerando il massimo su m nel membro destro, si ottiene $d_n \geq 1 - f_N(c)$, dove $f_N(c)$ è definita in (4.29).

Resta solo da dimostrare che $f_N(c) \leq (c+5)e^{-c}$ se $c \geq 3$, da cui segue che $d_n \geq 1 - (c+5)e^{-c}$, come annunciato. Per la formula di Stirling, nella versione della Proposizione 1.34, si ha $m! \geq m^m e^{-m} \sqrt{2\pi m}$. Inoltre, si può verificare la disuguaglianza

$$x^x e^{-x} \sqrt{2\pi x} \geq e^x/3 \quad \text{per ogni } x \geq 4, \quad (4.31)$$

mostrando che la differenza dei logaritmi dei due membri è positiva in $x = 4$ e che è una funzione crescente per $x \geq 4$.

Scegliamo ora $m = \lceil c \rceil$ in (4.29) (ossia, m è il più piccolo intero $\geq c$). Questo valore di m è legittimo, ossia è in $\{1, \dots, N\}$, perché nella seconda parte del Teorema 4.15 deve essere $c \leq \log N$ (in caso contrario $n < N \log N - cN < 0$ e d_n non è definita) e dunque $1 \leq m = \lceil c \rceil \leq$

$\lceil \log N \rceil \leq N$ per ogni $N \in \mathbb{N}$. Se $m \geq 4$, ossia $c > 3$, grazie a (4.31) si ha $m! = (\lceil c \rceil)! \geq e^{\lceil c \rceil} / 3 \geq e^c / 3$, e dato che $(m+1) = (\lceil c \rceil + 1) \leq c+2$, si ottiene

$$f_N(c) \leq \frac{1}{(\lceil c \rceil)!} + (\lceil c \rceil + 1)e^{-c} \leq \frac{3}{e^c} + (c+2)e^{-c} = (c+5)e^{-c}, \quad \forall c > 3.$$

Dato che $f_N(c)$ è una funzione continua di c , la disuguaglianza vale anche per $c = 3$. \square

4.5 Rivisitazione delle passeggiate aleatorie

Nel Paragrafo 2.2 abbiamo introdotto il modello della passeggiata aleatoria semplice, basato sullo spazio di probabilità (Ω_N, P_N) , dove $\Omega_N = \{-1, 1\}^N$ è lo spazio degli “incrementi” e P_N è la probabilità uniforme su Ω_N . Rivisitiamo questo modello usando le variabili aleatorie.

Se $\mathbf{x} = (x_1, x_2, \dots, x_N) \in \Omega_N$, per $i = 1, 2, \dots, N$ definiamo le variabili aleatorie

$$\begin{aligned} X_i(\mathbf{x}) &:= x_i \\ S_i(\mathbf{x}) &:= x_1 + \dots + x_i = \sum_{j=1}^i X_j(\mathbf{x}). \end{aligned} \quad (4.32)$$

Le definizioni (4.32) permettono di riformulare i risultati ottenuti nel Paragrafo 2.2 in termini di variabili aleatorie. Ad esempio, la proprietà di ricorrenza, dimostrata nel Teorema 2.7, si può esprimere come segue:

$$\lim_{N \rightarrow +\infty} P_N(S_n = 0 \text{ per qualche } n = 2, \dots, N) = 1. \quad (4.33)$$

Sottolineiamo il fatto che (4.33) è una semplice riscrittura del Teorema 2.7.

È utile però, a questo punto, notare quanto segue.

- La probabilità

$$P_N(S_n = 0 \text{ per qualche } n = 2, \dots, N)$$

dipende unicamente dalla *distribuzione congiunta* delle variabili aleatorie X_1, X_2, \dots, X_N , come si vede dalla relazione

$$P_N(S_n = 0 \text{ per qualche } n = 2, \dots, N) = \sum_{\mathbf{x} \in A_N} P_N(X_1 = x_1, X_2 = x_2, \dots, X_N = x_N),$$

$$\text{dove } A_N := \{\mathbf{x} \in \Omega_N : x_1 + \dots + x_n = 0 \text{ per qualche } n = 2, \dots, N\}.$$

- Essendo P_N la probabilità uniforme su Ω_N ,

$$P_N(X_1 = x_1, X_2 = x_2, \dots, X_N = x_N) = \frac{1}{2^N} = \prod_{i=1}^N P(X_i = x_i).$$

Si deduce quindi che X_1, X_2, \dots, X_N sono variabili aleatorie indipendenti e con

le stesse distribuzioni marginali, ossia

$$P_N(X_i = 1) = P(X_i = -1) = \frac{1}{2}. \quad (4.34)$$

Segue da queste considerazioni che se (Ω_N, P_N) è un *arbitrario* spazio di probabilità, in cui siano definite variabili aleatorie X_1, X_2, \dots, X_N indipendenti e con distribuzioni marginali (4.34), l'enunciato (4.33) continua a valere. Analoghe considerazioni possono essere fatte per la riformulazione in termini di variabili aleatorie di altri risultati relativi alla passeggiata aleatoria semplice.

Nello spirito, e nei limiti, di quanto affermato all'inizio di questo capitolo, possiamo andare oltre, considerando una *successione* $(X_n)_{n \geq 1}$ di variabili aleatorie indipendenti con distribuzione (4.34), in modo da evitare di indicizzare con N la probabilità P e poter considerare eventi più generali, che dipendono dall'intera successione. Al di là dell'esercizio di astrazione, la formulazione del modello in termini di successioni di variabili aleatorie indipendenti è utile anche per generalizzare il modello stesso, considerandone una versione *asimmetrica*, che ora definiamo.

Tirando le fila del discorso, d'ora in avanti supponiamo che $(X_n)_{n \geq 1}$ sia una successione di variabili aleatorie indipendenti, con distribuzioni marginali

$$P(X_i = 1) = p, \quad P(X_i = -1) = q := 1 - p, \quad \text{con } p \in (0, 1), \quad (4.35)$$

e definiamo *passeggiata aleatoria* la successione $(S_n)_{n \geq 0}$ definita da

$$S_n := \begin{cases} \sum_{k=1}^n X_k & \text{per } n \geq 1 \\ 0 & \text{per } n = 0 \end{cases}.$$

Se $p \neq \frac{1}{2}$, la passeggiata aleatoria ha dunque una “direzione preferenziale”. Si noti che il caso $p < \frac{1}{2}$ può essere ricondotto al caso $p > \frac{1}{2}$ mediante una riflessione, ossia considerando $-S_n$ invece di S_n . Pertanto, nel seguito supporremo che $p \in [\frac{1}{2}, 1)$.

Osservazione 4.20. Si noti che la semplice trasformazione lineare $Y_i := \frac{1}{2}(X_i + 1)$ produce una successione di variabili aleatorie indipendenti con distribuzione $\text{Be}(p)$, sicché $\frac{1}{2}(S_n + n) \sim \text{Bin}(n, p)$. Molte proprietà della passeggiata aleatoria possono pertanto essere formulate in termini di variabili aleatorie binomiali. In quanto segue, non avremo tuttavia convenienza ad operare questa trasformazione. \square

Il problema che affrontiamo è quello di studiare le proprietà del *tempo di primo passaggio al livello* $m \in \mathbb{Z}$, definito come la seguente variabile aleatoria

$$T_m := \min\{n \geq 0 : S_n = m\},$$

o, più esplicitamente, $T_m(\omega) := \min\{n \geq 0 : S_n(\omega) = m\}$, per ogni $\omega \in \Omega$, dove Ω è lo spazio di probabilità su cui sono definite le variabili X_k , e dunque anche le S_n . Si noti che $T_0 = 0$, mentre $T_m > 0$ per ogni $m \neq 0$.

Sottolineiamo che l'insieme $\{n \geq 0 : S_n(\omega) = m\}$ potrebbe essere vuoto, ed in tal caso poniamo $T_m(\omega) = +\infty$. Di conseguenza T_m è una variabile aleatoria discreta, a valori nell'insieme numerabile $\mathbb{N}_0 \cup \{+\infty\} = \{0, 1, 2, \dots, +\infty\}$. Il risultato principale che mostreremo è il seguente.

Teorema 4.21. *Se $p = \frac{1}{2}$, allora $P(T_m = +\infty) = 0$ per ogni $m \in \mathbb{Z}$. Se invece $p > \frac{1}{2}$,*

$$P(T_m = +\infty) = \begin{cases} 0 & \text{se } m \geq 0 \\ 1 - \left(\frac{1-p}{p}\right)^m > 0 & \text{se } m < 0 \end{cases}.$$

La dimostrazione è basata sullo studio della funzione reale $\zeta_m : (0, 1) \rightarrow \mathbb{R}$ definita da

$$\zeta_m(z) := \sum_{k \in \mathbb{N}_0} z^k P(T_m = k), \quad \text{per } z \in (0, 1), \quad (4.36)$$

con la convenzione $0^0 = 1$. Sottolineiamo che nella somma *non* è incluso $k = +\infty$. Si noti che se poniamo $z = e^t$ in (4.36) otteniamo formalmente la funzione generatrice dei momenti della variabile aleatoria T_m . Tuttavia, dato che T_m può assumere anche il valore $+\infty$, occorre un po' di cautela. Ritorneremo più avanti su questo punto.

La proprietà fondamentale che useremo della funzione $\zeta_m(z)$ è la seguente:

$$\zeta_m(1-) := \lim_{z \uparrow 1} \zeta_m(z) = 1 - P(T_m = +\infty), \quad (4.37)$$

che ora dimostriamo. Innanzitutto, essendo $P(T_m = k) \geq 0$, per monotonia

$$\zeta_m(z) \leq \sum_{k \in \mathbb{N}_0} P(T_m = k) = P(T_m \in \mathbb{N}_0) = 1 - P(T_m = +\infty), \quad \forall z \in (0, 1). \quad (4.38)$$

D'altro canto, per ogni $N \in \mathbb{N}$ si ha $\zeta_m(z) \geq \sum_{k=1}^N z^k P(T_m = k)$. Ma questo è un polinomio in z , quindi è continuo e per $z \uparrow 1$ otteniamo

$$\liminf_{z \uparrow 1} \zeta_m(z) \geq \lim_{z \uparrow 1} \left(\sum_{k=1}^N z^k P(T_m = k) \right) = \sum_{k=1}^N P(T_m = k).$$

Dato che questa stima vale per ogni $N \in \mathbb{N}$, prendendo l'estremo superiore su $N \in \mathbb{N}$ si ha che $\liminf_{z \uparrow 1} \zeta_m(z) \geq 1 - P(T_m = +\infty)$. Ricordando (4.38), si ottiene la relazione (4.37).

Vedremo ora come calcolare esplicitamente $\zeta_m(z)$.

Proposizione 4.22. Sia $p \geq \frac{1}{2}$. Per ogni $z \in (0, 1)$

$$\zeta_m(z) = \begin{cases} \left[\frac{1}{2(1-p)z} \left(1 - \sqrt{1 - 4p(1-p)z^2} \right) \right]^m & \text{se } m \geq 0 \\ \left[\frac{1}{2pz} \left(1 - \sqrt{1 - 4p(1-p)z^2} \right) \right]^m & \text{se } m \leq 0 \end{cases}. \quad (4.39)$$

Dimostrazione. Il calcolo sarà basato sulla seguente formula ricorsiva:

$$P(T_m = k) = p P(T_{m-1} = k-1) + q P(T_{m+1} = k-1), \quad \forall k \geq 1, m \neq 0, \quad (4.40)$$

la cui interpretazione è semplice: si decompone l'evento $\{T_m = k\}$ secondo il primo passo della passeggiata aleatoria, che vale $+1$ o -1 con probabilità p e q rispettivamente, dopodiché ci si ritrova con gli eventi $\{T_{m-1} = k-1\}$ e $\{T_{m+1} = k-1\}$. Formalmente, usando l'indipendenza delle variabili aleatorie X_k , $k \geq 1$, abbiamo

$$\begin{aligned} P(T_m = k) &= P(S_k = m, S_j < m \text{ per } j < k) \\ &= P(S_k = m, S_j < m \text{ per } j < k | X_1 = 1) P(X_1 = 1) \\ &\quad + P(S_k = m, S_j < m \text{ per } j < k | X_1 = -1) P(X_1 = -1) \\ &= p P(X_2 + \dots + X_k = m-1, X_2 + \dots + X_j < m-1 \text{ per } j < k) \\ &\quad + q P(X_2 + \dots + X_k = m+1, X_2 + \dots + X_j < m+1 \text{ per } j < k) \\ &= p P(X_1 + \dots + X_{k-1} = m-1, X_2 + \dots + X_j < m-1 \text{ per } j < k) \\ &\quad + q P(X_1 + \dots + X_{k-1} = m+1, X_2 + \dots + X_j < m+1 \text{ per } j < k) \\ &= p P(T_{m-1} = k-1) + q P(T_{m+1} = k-1), \end{aligned}$$

dove abbiamo usato il fatto che i vettori aleatori (X_2, \dots, X_k) e (X_1, \dots, X_{k-1}) hanno la stessa distribuzione.

Se $m \neq 0$ si ha necessariamente $T_m \geq 1$, pertanto nella relazione (4.36) possiamo restringere la somma a $k \in \mathbb{N}$. Inserendo quindi l'identità (4.40), otteniamo

$$\begin{aligned} \zeta_m(z) &= p \sum_{k \in \mathbb{N}} z^k P(T_{m-1} = k-1) + q \sum_{k \in \mathbb{N}} z^k P(T_{m+1} = k-1) \\ &= pz \sum_{k \in \mathbb{N}} z^{k-1} P(T_{m-1} = k-1) + qz \sum_{k \in \mathbb{N}} z^{k-1} P(T_{m+1} = k-1) \\ &= pz \zeta_{m-1}(z) + qz \zeta_{m+1}(z). \end{aligned}$$

Ciò mostra che, per ogni fissato $z \in (0, 1]$, la successione $(x_m = \zeta_m(z))_{m \in \mathbb{Z}}$ soddisfa l'equazione ricorsiva

$$x_m = pz x_{m-1} + qz x_{m+1}, \quad \forall m \in \mathbb{Z} \setminus \{0\}. \quad (4.41)$$

Cominciamo ad analizzare quest'equazione per $m \geq 1$, cercando soluzioni con indici positivi $(x_m)_{m \geq 0}$ a valori reali. La linearità dell'equazione (4.41) permette di dimostrare facilmente le seguenti affermazioni.

- Se $\mathbf{y} = (y_m)_{m \geq 0}$ e $\mathbf{z} = (z_m)_{m \geq 0}$ sono due soluzioni di (4.41), ogni loro combinazione lineare è anche una soluzione.
- Se $\mathbf{y} = (y_m)_{m \geq 0}$ e $\mathbf{z} = (z_m)_{m \geq 0}$ sono due soluzioni di (4.41) e (y_0, y_1) e (z_0, z_1) sono vettori di \mathbb{R}^2 linearmente indipendenti, ossia non proporzionali, ogni soluzione $\mathbf{x} = (x_m)_{m \geq 0}$ di (4.41) è combinazione lineare di \mathbf{y} e \mathbf{z} . Infatti esistono $\alpha, \beta \in \mathbb{R}$ tali che $(x_0, x_1) = \alpha(y_0, y_1) + \beta(z_0, z_1)$, e la relazione (4.41), soddisfatta da $\mathbf{x}, \mathbf{y}, \mathbf{z}$, per $m = 1, 2, \dots$ mostra che $x_i = \alpha y_i + \beta z_i$ per $i = 2, 3, \dots$.

Cerchiamo allora soluzioni di (4.41) della forma $x_m = \lambda^m$. Sostituendo in (4.41), si vede che λ^m risolve (4.41) se e solo se

$$qz\lambda^2 - \lambda + pz = 0.$$

Questo è un polinomio di secondo grado in λ , le cui soluzioni sono

$$\lambda_-(z) = \frac{1}{2qz} \left[1 - \sqrt{1 - 4pqz^2} \right], \quad \lambda_+(z) = \frac{1}{2qz} \left[1 + \sqrt{1 - 4pqz^2} \right]. \quad (4.42)$$

Si noti che $pq = p(1-p) \leq \frac{1}{4}$ per ogni $p \in [0, 1]$. Di conseguenza, per ogni $z < 1$ si ha $1 - 4pqz^2 > 0$ e dunque $\lambda_-(z) \neq \lambda_+(z)$, dunque i vettori $(1, \lambda_-)$ e $(1, \lambda_+)$ sono linearmente indipendenti. Possiamo dunque esprimere, per ogni $z \in (0, 1)$ fissato, la successione $(\zeta_m(z))_{m \geq 0}$ come combinazione lineare

$$\zeta_m(z) = \alpha(z) \lambda_-^m(z) + \beta(z) \lambda_+^m(z), \quad \forall m \geq 0, z \in (0, 1), \quad (4.43)$$

per opportuni coefficienti $\alpha(z), \beta(z) \in \mathbb{R}$.

Per ogni $z \in (0, 1)$ si ha $\lambda_+(z) > 1$, perché $2qz < 1$, mentre $\lambda_-(z) < 1$, come si verifica facilmente. Di conseguenza la successione $(\lambda_-^m(z))_{m \geq 0}$ è limitata, mentre $(\lambda_+^m(z))_{m \geq 0}$ è illimitata. Dal momento che $0 \leq \zeta_m(z) \leq 1$, cf. (4.38), in (4.43) si deve avere $\beta(z) \equiv 0$. Notando poi che $\zeta_0(z) = 1$ per ogni z , grazie a (4.36) (si ricordi che $T_0 = 0$), si deve avere $\alpha(z) \equiv 1$. In definitiva, abbiamo ottenuto che

$$\zeta_m(z) = \lambda_-^m(z), \quad \forall m \geq 0, z \in (0, 1),$$

dimostrando la prima riga in (4.39).

Analizziamo infine l'equazione (4.41) per $m \leq -1$, cercando soluzioni con indici negativi $(x_m)_{m \leq 0}$. L'analisi svolta sopra può essere ripetuta in modo pressoché identico, mostrando che anche per $m \leq 0$ la funzione $\zeta_m(z)$ può essere espressa mediante la relazione (4.43), per un'opportuna scelta dei coefficienti $\alpha(z), \beta(z) \in \mathbb{R}$, non necessariamente gli stessi di prima. In effetti, per $m \leq 0$ è la successione $(\lambda_-^m(z))_{m \geq 0}$ a essere illimitata, mentre $(\lambda_+^m(z))_{m \geq 0}$ è limitata, pertanto si deve avere $\alpha(z) \equiv 0$. La

condizione $\zeta_0(z) = 1$ per ogni z mostra quindi che $\beta(z) \equiv 1$, quindi

$$\zeta_m(z) = \lambda_+^m(z), \quad \forall m \leq 0, z \in (0, 1).$$

Ricordando (4.42) si ottiene

$$\lambda_+^{-1}(z) = \frac{2qz}{1 + \sqrt{1 - 4pqz^2}} = \frac{2qz(1 - \sqrt{1 - 4pqz^2})}{1 - (1 - 4pqz^2)} = \frac{1}{2pz} \left[1 - \sqrt{1 - 4pqz^2} \right],$$

dimostrando la seconda riga in (4.39). \square

Possiamo ora dimostrare facilmente il Teorema 4.21. Usando l'espressione esplicita per $\zeta_m(z)$ ottenuta nella Proposizione 4.22, si ha che

$$\zeta_m(1-) := \lim_{z \uparrow 1} \zeta_m(z) = \begin{cases} 1 & \text{se } p = \frac{1}{2}, \forall m \in \mathbb{Z} \\ 1 & \text{se } p > \frac{1}{2}, \forall m \geq 0 \\ \left(\frac{1-p}{p}\right)^m & \text{se } p > \frac{1}{2}, \forall m \leq 0 \end{cases}.$$

Ricordando la relazione (4.37), la dimostrazione del Teorema 4.21 è completa.

Concludiamo questo paragrafo analizzando più in dettaglio il caso $p > \frac{1}{2}$. Grazie al Teorema 4.21, la variabile aleatoria T_m per $m \geq 1$ è quasi certamente finita, ossia $P(T_m < +\infty) = 1$. Consideriamone allora la funzione generatrice dei momenti

$$M_{T_m}(t) := E[e^{tT_m}] = \sum_{k \in \mathbb{N}_0} e^{tk} P(T_m = k).$$

Come abbiamo già osservato, questa funzione coincide con $\zeta_m(e^t)$, cf. (4.36), quindi per $t < 0$ abbiamo un'espressione esplicita per $M_{T_m}(t)$, data dalla prima riga in (4.39):

$$M_{T_m}(t) = \left[\frac{e^{-t}}{2(1-p)} \left(1 - \sqrt{1 - 4p(1-p)e^t} \right) \right]^m, \quad \forall m \geq 1, t \in (-\infty, 0). \quad (4.44)$$

Per verifica diretta, quest'uguaglianza vale anche per $t = 0$, dato che $M_{T_m}(0) = 1$.

Mostriamo ora che $M_{T_m}(t)$ è finita in un intorno di $t = 0$.

Proposizione 4.23. *Se $p > \frac{1}{2}$, esiste $\delta > 0$ tale che $M_{T_m}(t) < +\infty$ per ogni $t \in (-\infty, \delta)$ e per ogni $m \geq 1$.*

Dimostrazione. Notiamo anzitutto che

$$\{T_m \geq n\} \subseteq \{S_n < m\}. \quad (4.45)$$

Cerchiamo allora un limite superiore per la probabilità $P(S_n < m)$. Sia $s > 0$. Per la

disuguaglianza di Markov, applicata alla variabile aleatoria $X := e^{-sS_n}$:

$$P(S_n < m) = P(e^{-sS_n} > e^{-sm}) \leq \frac{E[e^{-sS_n}]}{e^{-sm}} = e^{sm} M_{S_n}(-s).$$

Essendo S_n somma di variabili aleatorie indipendenti con distribuzione (4.35), si ha

$$M_{S_n}(-s) = (M_{X_1}(-s))^n = (pe^{-s} + qe^s)^n,$$

pertanto

$$P(S_n < m) \leq \exp[n \log(pe^{-s} + qe^s) + sm].$$

La funzione $\varphi(s) := \log(pe^{-s} + qe^s)$ è tale che $\varphi(0) = 0$, $\varphi'(0) = q - p < 0$. Quindi esiste $\bar{s} > 0$ tale che $\varphi(\bar{s}) < 0$. Allora, posto $\delta := -\varphi(\bar{s}) > 0$ e $C_m := \exp[\bar{s}m]$, abbiamo ottenuto

$$P(S_n < m) \leq C_m e^{-\delta n}. \quad (4.46)$$

Usando quindi (4.45) e (4.46),

$$P(T_m = n) \leq P(T_m \geq n) \leq C_m e^{-\delta n},$$

da cui segue che per ogni $t < \delta$

$$M_{T_m}(t) = \sum_{n \in \mathbb{N}_0} e^{tn} P(T_m = n) \leq \sum_{n \in \mathbb{N}_0} e^{tn} C_m e^{-\delta n} < +\infty,$$

completando la dimostrazione. \square

La Proposizione 4.23, assieme al Teorema 3.96, implica che la variabile aleatoria T_m ammette momenti di ogni ordine, che possono essere determinati calcolando le derivate nell'origine della funzione $M_{T_m}(t)$. Grazie alla relazione (4.44), abbiamo un'espressione esplicita per $M_{T_m}(t)$ se $t \leq 0$, pertanto possiamo calcolarne le derivate. A titolo di esempio, media e varianza di T_m valgono

$$E(T_m) = M'_m(0) = \frac{m}{p-q}, \quad \text{Var}(T_m) = M''_m(0) - [M'_m(0)]^2 = \frac{mpq}{(p-q)^3}.$$

Osservazione 4.24. Grazie alla Proposizione 4.23, con argomenti standard di analisi matematica si mostra che l'uguaglianza (4.44) si estende a $t \in (-\infty, +\delta)$. \square

Osservazione 4.25. L'espressione (4.39) per la funzione $\zeta_m(z)$ può essere sfruttata per calcolare altre quantità interessanti. Ad esempio, ricordando la relazione (4.36), la funzione $\zeta_m(z)$ si può espandere in serie di Taylor attorno a $z = 0$ e i coefficienti di tale serie coincidono con le probabilità $P(T_m = k)$. Per i dettagli, si consultino i riferimenti contenuti nelle note bibliografiche alla fine di questo capitolo. \square

4.6 La condensazione di Bose-Einstein

Nel Paragrafo 2.3, abbiamo definito le *statistiche di Bose-Einstein* P_{BE} e di *Fermi-Dirac* P_{FD} come le probabilità uniformi sugli insiemi $\widehat{\Omega}_{n,r}$ e $\widehat{\Omega}'_{n,r}$, che descrivono i numeri di occupazione di n particelle indistinguibili (bosoni o fermioni):

$$\begin{aligned}\widehat{\Omega}_{n,r} &:= \{(k_1, \dots, k_r) \in (\mathbb{N}_0)^r : k_1 + \dots + k_r = n\}. \\ \widehat{\Omega}'_{n,r} &:= \{(k_1, \dots, k_r) \in \{0, 1\}^r : k_1 + \dots + k_r = n\}.\end{aligned}\quad (4.47)$$

La scelta della probabilità uniforme sull'insieme delle possibili configurazioni è appropriata se l'energia totale del sistema è indipendente dagli stati delle particelle che lo compongono (o quando l'energia ha un effetto trascurabile).

Nel caso più realistico in cui l'energia del sistema invece dipenda dagli stati delle particelle componenti, è opportuno considerare una *misura di Gibbs*, come descritto nell'Esempio 1.14: se Ω è lo spazio delle configurazioni del sistema e $H : \Omega \rightarrow \mathbb{R}$ è la funzione energia, quando il sistema è in equilibrio termico alla temperatura assoluta T , ponendo $\beta := \frac{1}{k_B T}$ (dove k_B è la costante di Boltzmann), la probabilità di osservare il sistema in una configurazione $\omega \in \Omega$ è data da

$$P_\beta(\{\omega\}) := \frac{e^{-\beta H(\omega)}}{Z(\beta)}, \quad \text{dove} \quad Z(\beta) := \sum_{\omega \in \Omega} e^{-\beta H(\omega)}. \quad (4.48)$$

Il punto cruciale è la scelta dello spazio delle configurazioni Ω . Per un sistema composto da n particelle *indistinguibili*, si sceglierà $\Omega = \widehat{\Omega}_{n,r}$ se le particelle sono bosoni e $\Omega = \widehat{\Omega}'_{n,r}$ se le particelle sono fermioni. Nel caso in cui le particelle siano invece distinguibili – per esempio, se sono diverse tra loro – si sceglierà lo spazio diverso $\Omega = \Omega_{n,r}$ delle funzioni da $\{1, \dots, n\}$ in $\{1, \dots, r\}$, definito in (2.15).

Come abbiamo già osservato nell'Esempio 1.14, per $\beta = 0$ (limite di temperatura infinita) la probabilità $P_\beta(\{\omega\})$ non dipende più da ω , dunque P_0 non è altro che la probabilità uniforme su Ω . Ritroviamo in questo modo le statistiche di Bose-Einstein e di Fermi-Dirac: $P_0 = P_{BE}$ se $\Omega = \widehat{\Omega}_{n,r}$, mentre $P_0 = P_{FD}$ se $\Omega = \widehat{\Omega}'_{n,r}$. Sempre dall'Esempio 1.14 sappiamo che per $\beta \rightarrow \infty$ (limite di temperatura nulla) la probabilità P_β si concentra sull'insieme dei minimi assoluti della funzione H . In questo regime un sistema di bosoni indistinguibili dà origine a un fenomeno molto interessante, noto come *condensazione di Bose-Einstein*, in cui gli effetti quantistici diventano visibili su grande scala. Ci apprestiamo ora a studiare un modello quantitativo di tale fenomeno.

Consideriamo un sistema di n particelle (bosoni) indistinguibili. Conviene rimuovere l'assunzione che ogni particella possa assumere un numero *finito* di stati: assumeremo che gli stati possibili di ogni singola particella siano un insieme numerabile, che etichettiamo con gli elementi di $\mathbb{N}_0 = \{0, 1, 2, \dots\}$. Le configurazioni del sistema saranno sempre descritte dai numeri di occupazione, che ora non sono più vettori finiti bensì successioni $\mathbf{k} = (k_i)_{i \in \mathbb{N}_0} = (k_0, k_1, \dots)$, dove $k_i \in \mathbb{N}_0$ indica il numero di particelle nello stato i , tali che $k_0 + k_1 + \dots = n$. In particolare, una tale successio-

ne \mathbf{k} ha solo un numero finito (al più n) di componenti non nulle. Lo spazio delle configurazioni del sistema è dunque

$$\Omega_n := \left\{ \mathbf{k} = (k_i)_{i \in \mathbb{N}_0} : k_i \in \mathbb{N}_0 \forall i \in \mathbb{N}_0; \sum_{i \in \mathbb{N}_0} k_i = n \right\}. \quad (4.49)$$

È chiara l'analogia con lo spazio $\widehat{\Omega}_{n,r}$, definito in (2.17), ma sottolineiamo che l'insieme Ω_n è infinito (numerabile). In particolare, non è possibile considerare la probabilità uniforme su Ω_n . Tuttavia, per ogni $\beta > 0$ è comunque possibile definire una misura di Gibbs P_β mediante (4.48), purché $Z_\beta < \infty$. Questo è esattamente quello che faremo. Occorre innanzitutto definire la funzione energia $H : \Omega_n \rightarrow \mathbb{R}$.

A ogni stato $i \in \mathbb{N}_0$ corrisponde un'energia di singola particella $\varepsilon_i \in \mathbb{R}$. Assumiamo che $\varepsilon_0 < \varepsilon_1 \leq \varepsilon_2 \leq \dots \leq \varepsilon_i \leq \varepsilon_{i+1} \leq \dots$, ossia gli stati sono numerati in ordine crescente di energia, e lo stato 0, detto *stato fondamentale*, è l'unico con energia minima. I valori precisi delle energie ε_i , che dipendono dai dettagli del modello, verranno specificati più avanti. Se le particelle sono non interagenti (si parla in questo caso di un *gas perfetto*), l'energia totale del sistema è data dalla somma delle energie delle singole particelle: dunque a ogni configurazione $\mathbf{k} = (k_0, k_1, \dots) \in \Omega_n$ corrisponde un'energia

$$H(\mathbf{k}) := \sum_{i \in \mathbb{N}_0} k_i \varepsilon_i = k_0 \varepsilon_0 + k_1 \varepsilon_1 + \dots, \quad (4.50)$$

dove la somma ha solo un numero finito di termini non nulli. L'ipotesi che le particelle siano non interagenti è naturalmente solo una semplificazione, che tuttavia è accettabile se l'interazione tra le particelle è debole. Prendiamo quindi (4.50) come definizione della funzione energia $H : \Omega_n \rightarrow \mathbb{R}$.

Possiamo ora definire la *funzione di partizione* $Z_{n,\beta}$ come la seguente somma (numerabile) a termini positivi:

$$Z_{n,\beta} := \sum_{\mathbf{k} \in \Omega_n} e^{-H(\mathbf{k})} = \sum_{(k_0, k_1, \dots) : k_i \in \mathbb{N}_0, k_0 + k_1 + \dots = n} e^{-\beta(k_0 \varepsilon_0 + k_1 \varepsilon_1 + \dots)}. \quad (4.51)$$

Mostriamo più avanti, sotto opportune ipotesi sulle energie ε_i , che $Z_{n,\beta} < +\infty$ per ogni $n \in \mathbb{N}$ e $\beta \in (0, \infty)$. Questo è fondamentale, perché permette di definire una probabilità $P_{n,\beta}$ su Ω_n assegnandone la densità discreta, esattamente come in (4.48):

$$P_{n,\beta}(\{\mathbf{k}\}) := \frac{e^{-\beta H(\mathbf{k})}}{Z_{n,\beta}} = \frac{e^{-\beta(k_0 \varepsilon_0 + k_1 \varepsilon_1 + \dots)}}{Z_{n,\beta}}, \quad \forall \mathbf{k} = (k_0, k_1, \dots) \in \Omega_n. \quad (4.52)$$

Il nostro obiettivo è di studiare le proprietà della probabilità $P_{n,\beta}$ nel *limite termodinamico* $n \rightarrow \infty$, per ogni $\beta > 0$ fissato. Questo limite matematico è utile per descrivere sistemi con un numero elevato di particelle, che è il caso tipico di molti sistemi fisici, in cui n può essere dell'ordine del numero di Avogadro $\simeq 6 \cdot 10^{23}$.

La domanda che ci poniamo è la seguente: in media, quante delle n particelle che compongono il sistema sono nello stato fondamentale 0? Vedremo che la risposta dipende fortemente dal valore di β , ossia dalla temperatura: in particolare, se la tem-

peratura è sufficientemente bassa, mostreremo che mediamente *una frazione strettamente positiva delle n particelle è nello stato fondamentale*. Questo fenomeno, noto come *condensazione di Bose-Einstein*, ha la sua origine precisamente nella prescrizione quantistica di trattare le particelle come indistinguibili, e non si verifica se le particelle sono distinguibili. La dimostrazione procede in sei passi.

Passo 1. La prima osservazione, elementare ma non banale, è che possiamo assumere, senza perdita di generalità, che l'energia dello stato fondamentale sia nulla. Infatti, ponendo $\varepsilon'_i := \varepsilon_i - \varepsilon_0$, per ogni $\mathbf{k} = (k_0, k_1, \dots) \in \Omega_n$ possiamo scrivere

$$e^{-\beta(k_0\varepsilon_0 + k_1\varepsilon_1 + \dots)} = e^{-\beta(k_0 + k_1 + \dots)\varepsilon_0} e^{-\beta(k_1\varepsilon'_1 + k_2\varepsilon'_2 + \dots)} = e^{-\beta\varepsilon_0 n} e^{-\beta(k_1\varepsilon'_1 + k_2\varepsilon'_2 + \dots)},$$

poiché $k_0 + k_1 + \dots = n$. Ricordando (4.51), il fattore $e^{-\beta\varepsilon_0 n}$ si semplifica tra numeratore e denominatore nell'espressione (4.52). Il risultato netto è che nell'espressione (4.52) è possibile rimpiazzare ogni ε_i con ε'_i , purché si faccia lo stesso anche in (4.51), e si noti che $\varepsilon'_0 = 0$. Per non appesantire le notazioni, omettiamo gli apici e *d'ora in avanti supporremo che $\varepsilon_0 = 0$* , ricordando che le ε_i rappresentano le differenze di energia rispetto allo stato fondamentale.

Facciamo inoltre l'ipotesi che la successione $(\varepsilon_i)_{i \in \mathbb{N}}$ sia tale che

$$\sum_{i \in \mathbb{N}} e^{-\beta\varepsilon_i} < \infty, \quad \forall i \geq 1, \forall \beta > 0. \quad (4.53)$$

Intuitivamente, ciò significa che $\varepsilon_i \rightarrow +\infty$ più che logaritmicamente, una condizione naturale che è sempre soddisfatta negli esempi rilevanti per il nostro problema. Definiamo quindi il prodotto infinito

$$C_\beta := \prod_{i=1}^{\infty} (1 - e^{-\beta\varepsilon_i})^{-1} := \lim_{r \rightarrow \infty} \prod_{i=1}^r (1 - e^{-\beta\varepsilon_i})^{-1}, \quad (4.54)$$

osservando che il limite esiste: infatti, essendo $\varepsilon_i \geq \varepsilon_1 > 0$ per ogni $i \in \mathbb{N}$, segue che $(1 - e^{-\beta\varepsilon_i})^{-1} > 1$ e dunque la successione dei prodotti è crescente in r . Inoltre

$$1 < C_\beta < \infty, \quad \forall \beta \in (0, \infty). \quad (4.55)$$

La prima disuguaglianza è immediata. Per quanto riguarda la seconda, notiamo che per ogni $x_0 \in (0, 1)$ vale la stima $-\log(1 - x) \leq cx$, per ogni $x \in [x_0, 1]$, grazie alla concavità del logaritmo, dove $c := -\log(x_0)/x_0 > 0$. Applicando questa stima per $x_0 = e^{-\beta\varepsilon_1}$, segue che esiste $c(\beta) > 0$ tale che per ogni $r \in \mathbb{N}$

$$\prod_{i=1}^r (1 - e^{-\beta\varepsilon_i})^{-1} = e^{-\sum_{i=1}^r \log(1 - e^{-\beta\varepsilon_i})} \leq e^{c(\beta) \sum_{i=1}^r e^{-\beta\varepsilon_i}} \leq e^{c(\beta) \sum_{i=1}^{\infty} e^{-\beta\varepsilon_i}} < \infty,$$

proprio grazie alla condizione (4.53).

Passo 2. Introduciamo lo spazio Ω delle successioni di numeri naturali *definitivamente nulle*, ossia in cui solo un numero finito di componenti sono diverse da zero – o, equivalentemente, in cui la somma delle componenti è finita:

$$\begin{aligned}\Omega &:= \{ \mathbf{k} = (k_0, k_1, \dots) : k_i \in \mathbb{N}_0 \forall i \in \mathbb{N}_0; \exists r \in \mathbb{N} : k_i = 0 \forall i > r \} \\ &= \left\{ \mathbf{k} = (k_0, k_1, \dots) : k_i \in \mathbb{N}_0 \forall i \in \mathbb{N}_0; \sum_{i \in \mathbb{N}_0} k_i < \infty \right\} \subseteq (\mathbb{N}_0)^{\mathbb{N}_0}. \quad (4.56)\end{aligned}$$

Intuitivamente, Ω rappresenta lo spazio delle configurazioni di un sistema con un numero finito, ma non fissato, di particelle. Chiaramente Ω è l'unione *disgiunta* degli insiemi Ω_n , introdotti in (4.49):

$$\Omega = \bigcup_{n=0}^{\infty} \Omega_n, \quad \Omega_n \cap \Omega_{n'} = \emptyset \quad \text{per } n \neq n'.$$

Una rappresentazione alternativa di Ω , altrettanto utile, si ha notando che ogni $\mathbf{k} \in \Omega$ è della forma $\mathbf{k} = (k_0, k_1, \dots, k_r, 0, 0, 0, \dots)$ per un opportuno $r \in \mathbb{N}_0$, pertanto

$$\Omega = \bigcup_{r=0}^{\infty} \Lambda_r, \quad \text{con } \Lambda_r := (\mathbb{N}_0)^{\{0, \dots, r\}} \times \{\mathbf{0}\}, \quad (4.57)$$

dove $\mathbf{0} = (0, 0, \dots)$ indica la successione nulla. Gli insiemi Λ_r sono numerabili e crescenti, ossia $\Lambda_r \subseteq \Lambda_{r+1}$ per ogni $r \in \mathbb{N}_0$. Quindi *l'insieme Ω – e a maggior ragione $\Omega_n \subseteq \Omega$ – è numerabile*, in quanto unione numerabile di insiemi numerabili.

Definiamo ora, per ogni $n \in \mathbb{N}$ e $\beta > 0$, una probabilità $q_{n,\beta}$ su Ω , assegnandone la densità discreta $q_{n,\beta}$ nel modo seguente:

$$q_{n,\beta}(\mathbf{k}) := \frac{C_\beta^{-1}}{n+1} \mathbb{1}_{\{0, \dots, n\}}(k_0) e^{-\beta \sum_{i=1}^{\infty} k_i \varepsilon_i}, \quad \forall \mathbf{k} = (k_0, k_1, \dots) \in \Omega. \quad (4.58)$$

Ricordando la definizione (4.57) di Λ_r , per ogni $r \in \mathbb{N}$ si ha

$$\begin{aligned}\sum_{\mathbf{k} \in \Lambda_r} q_{n,\beta}(\mathbf{k}) &= \frac{C_\beta^{-1}}{n+1} \sum_{k_0, \dots, k_r \in \mathbb{N}_0} \mathbb{1}_{\{0, \dots, n\}}(k_0) e^{-\beta k_1 \varepsilon_1} \dots e^{-\beta k_r \varepsilon_r} \\ &= C_\beta^{-1} \prod_{i=1}^r (1 - e^{-\beta \varepsilon_i})^{-1},\end{aligned}$$

avendo usato la somma della serie geometrica (0.4). Ricordando che una somma infinita è per definizione l'estremo superiore delle possibili somme finite, ed essendo Ω l'unione crescente degli insiemi Λ_r , per la relazione (4.54) si ha

$$\sum_{\mathbf{k} \in \Omega} q_{n,\beta}(\mathbf{k}) = \sup_{r \in \mathbb{N}} \sum_{\mathbf{k} \in \Lambda_r} q_{n,\beta}(\mathbf{k}) = C_\beta^{-1} \lim_{r \rightarrow \infty} \prod_{i=1}^r (1 - e^{-\beta \varepsilon_i})^{-1} = 1,$$

dunque $q_{n,\beta}(\cdot)$ è effettivamente una densità discreta su Ω .

Sottolineiamo che $Q_{n,\beta}$ è una probabilità sull'insieme fissato Ω , che non dipende da n , mentre la probabilità di Bose-Einstein $P_{n,\beta}$, che abbiamo definito in (4.52) e che è l'oggetto del nostro studio, è una probabilità sullo spazio Ω_n delle configurazioni di un sistema con n particelle. Dato che $\Omega_n \subseteq \Omega$, e vista la somiglianza tra le espressioni (4.58) e (4.52), ci si può aspettare qualche legame tra $Q_{n,\beta}$ e $P_{n,\beta}$. In effetti, ricordando la definizione (4.51) della funzione di partizione $Z_{n,\beta}$ (in cui poniamo $\varepsilon_0 = 0$, come detto), possiamo scrivere

$$Z_{n,\beta} = \sum_{\mathbf{k} \in \Omega_n} e^{-\beta \sum_{i=1}^{\infty} k_i \varepsilon_i} = (n+1) C_\beta \sum_{\mathbf{k} \in \Omega_n} q_{n,\beta}(\mathbf{k}) = (n+1) C_\beta Q_{n,\beta}(\Omega_n). \quad (4.59)$$

Notiamo incidentalmente che questa relazione mostra che $Z_{n,\beta} \leq (n+1) C_\beta < \infty$, per ogni $n \in \mathbb{N}$ e $\beta > 0$, come annunciato in precedenza. Ricordando (4.52), si ha

$$P_{n,\beta}(\{\mathbf{k}\}) = \frac{e^{-\beta \sum_{i=1}^{\infty} k_i \varepsilon_i}}{Z_{n,\beta}} = \frac{q_{n,\beta}(\mathbf{k})}{Q_{n,\beta}(\Omega_n)} = Q_{n,\beta}(\{\mathbf{k}\} \mid \Omega_n), \quad \forall \mathbf{k} \in \Omega_n. \quad (4.60)$$

Questo mostra che la probabilità di Bose-Einstein $P_{n,\beta}$ non è altro che il *condizionamento* della probabilità $Q_{n,\beta}$ rispetto all'evento Ω_n . L'utilità di questa rappresentazione, come sarà chiaro a breve, sta nel fatto che la dipendenza da n della probabilità $Q_{n,\beta}$ è molto speciale, dal momento che n compare in (4.58) solo nel termine relativo alla prima componente k_0 .

Passo 3. Riformuliamo ora il problema in termini di variabili aleatorie. Sebbene si tratti sostanzialmente di una traduzione, ciò fornisce una rappresentazione più intuitiva e “probabilistica” delle probabilità (4.52). Introduciamo sullo spazio Ω le variabili aleatorie reali $(X_i)_{i \in \mathbb{N}_0}$, definite da

$$X_i(\mathbf{k}) := k_i, \quad \forall \mathbf{k} = (k_0, k_1, \dots) \in \Omega.$$

Intuitivamente, se \mathbf{k} rappresenta una configurazione del sistema, la variabile aleatoria X_i conta quante particelle ci sono nello stato i . In particolare, X_0 indica quante particelle sono nello stato fondamentale, ossia la quantità principale a cui siamo interessati. Risulta comodo studiare la quantità complementare $S_{>}$, che conta il numero totale di particelle *al di fuori* dello stato fondamentale:

$$S_{>} := \sum_{i=1}^{\infty} X_i. \quad (4.61)$$

Si noti che la funzione $S_{>} : \Omega \rightarrow \mathbb{N}_0$ è ben definita e *finita*: infatti, nonostante la somma in (4.61) sia infinita, sappiamo che $X_i(\mathbf{k}) = k_i \neq 0$ solo per un numero finito di indici i , per ogni $\mathbf{k} \in \Omega$.

Possiamo ora rappresentare gli eventi di Ω in termini delle variabili aleatorie X_i . Per esempio, l'evento Ω_n , ossia “il numero di particelle del sistema è pari a n ”, può essere rappresentato come

$$\Omega_n = \left\{ \sum_{i=0}^{\infty} X_i = n \right\} = \{X_0 + S_{>} = n\}, \quad \forall n \in \mathbb{N}.$$

dove sottolineiamo che, data qualunque variabile aleatoria Y definita su Ω , la scrittura $\{Y = n\}$ significa $\{\mathbf{k} \in \Omega : Y(\mathbf{k}) = n\}$, per la Definizione 3.5. Notiamo che, tautologicamente, per ogni $m \in \mathbb{N}_0$ possiamo scrivere l'evento $\{S_{>} = m\}$ come unione disgiunta dei singoletti corrispondenti ai suoi elementi:

$$\{S_{>} = m\} = \bigcup_{\mathbf{k} \in \{S_{>} = m\}} \{\mathbf{k}\}.$$

Pertanto, per l'additività della probabilità, otteniamo da (4.60)

$$P_{n,\beta}(S_{>} = m) = Q_{n,\beta}(S_{>} = m \mid X_0 + S_{>} = n), \quad \forall n \in \mathbb{N}, \forall m \in \{0, \dots, n\}. \quad (4.62)$$

Sfrutteremo questa relazione tra poco. Prima studiamo più precisamente $Q_{n,\beta}$.

Passo 4. Vogliamo ora determinare le distribuzioni delle variabili aleatorie $(X_i)_{i \in \mathbb{N}_0}$ rispetto alla probabilità $Q_{n,\beta}$, da cui deriveremo in particolare proprietà importanti di X_0 e $S_{>}$. Mostriamo il seguente risultato.

Proposizione 4.26. *Le variabili aleatorie $(X_i)_{i \in \mathbb{N}_0}$ rispetto alla probabilità $Q_{n,\beta}$ sono indipendenti, con le seguenti distribuzioni marginali:*

$$X_0 \sim \text{Unif}(\{0, 1, \dots, n\}), \quad X_i \sim \text{Geo}_0(1 - e^{-\beta \varepsilon_i}), \quad \forall i \geq 1. \quad (4.63)$$

Dimostrazione. Ricordando la definizione di indipendenza per una famiglia infinita di variabili aleatorie, e la densità discreta (3.85) di una variabile $\text{Geo}_0(p)$, dobbiamo mostrare quanto segue: per ogni $\ell \in \mathbb{N}$ e per ogni $k_0, k_1, \dots, k_\ell \in \mathbb{N}_0$

$$\begin{aligned} Q_{n,\beta}(X_0 = k_0, X_1 = k_1, \dots, X_\ell = k_\ell) &= p_{\text{Unif}\{0, \dots, n\}}(k_0) \prod_{i=1}^{\ell} p_{\text{Geo}_0(1 - e^{-\beta \varepsilon_i})}(k_i) \\ &= \frac{1}{n+1} \mathbb{1}_{\{0, \dots, n\}}(k_0) \cdot \prod_{i=1}^{\ell} (1 - e^{-\beta \varepsilon_i})(e^{-\beta \varepsilon_i})^{k_i}. \end{aligned} \quad (4.64)$$

Osserviamo che

$$Q_{n,\beta}(X_0 = k_0, \dots, X_\ell = k_\ell) = \sum_{(k_{\ell+1}, k_{\ell+2}, \dots) \in \Omega} q_{n,\beta}((k_0, \dots, k_\ell, k_{\ell+1}, k_{\ell+2}, \dots)).$$

Si ricordi la definizione (4.58) di $q_{n,\beta}(\cdot)$ e la serie geometrica (0.4). Ogni somma su k_i , con $i > \ell$, produce il termine $\sum_{k_i \in \mathbb{N}_0} e^{-\beta k_i \varepsilon_i} = (1 - e^{-\beta \varepsilon_i})^{-1}$, che si semplifica con l'analogo termine in C_β , cf. (4.54). In questo modo si ottiene (4.64).[†] \square

[†] Dato che le variabili k_i su cui si somma sono in numero infinito, per giustificare rigorosamente questo procedimento occorrerebbe procedere per troncamento, sommando innanzitutto su un numero finito r di componenti $k_{\ell+1}, \dots, k_{\ell+r}$, e poi prendendo l'estremo superiore su r , analogamente a quanto abbiamo fatto nelle righe successive a (4.58). I dettagli sono lasciati al lettore.

Osservazione 4.27. La Proposizione 4.26 fornisce un esempio di una successione infinita $(X_i)_{i \in \mathbb{N}}$ di variabili aleatorie *indipendenti*, definite su uno spazio di probabilità discreto $(\Omega, \mathcal{Q}_{n,\beta})$. Si noti che le distribuzioni marginali $X_i \sim \text{Geo}_0(1 - e^{-\beta \varepsilon_i})$ dipendono da i e tendono a diventare “degeneri” per $i \rightarrow \infty$: infatti $\varepsilon_i \rightarrow \infty$, quindi (esercizio) la densità discreta di $\text{Geo}_0(1 - e^{-\beta \varepsilon_i})$ converge verso $\mathbb{1}_{\{0\}}(\cdot)$.

In effetti, se le variabili indipendenti $(X_i)_{i \in \mathbb{N}}$ avessero tutte la stessa distribuzione marginale (non concentrata in un punto), esse non potrebbero essere definite su uno spazio di probabilità discreto, per quanto visto nell’Osservazione 3.42 \square

La Proposizione 4.26 è di fondamentale importanza. Notiamo in particolare che

- la distribuzione marginale della variabile aleatoria X_i , per $i \geq 1$, *non dipende da n* : l’unica dipendenza da n è in X_0 , la cui distribuzione marginale è $\text{Unif}\{0, \dots, n\}$;
- le variabili aleatorie X_0 e $S_{>}$ sono *indipendenti*, dal momento che $S_{>}$ è funzione delle variabili $(X_i)_{i \geq 1}$, cf. (4.61).

Volendo essere precisi, l’indipendenza di X_0 e $S_{>}$ può essere dedotta dalla Proposizione 3.40, notando che $S_{>}$ è una funzione della variabile aleatoria $\mathbf{X}_{>} = (X_1, X_2, \dots) : \Omega \rightarrow \Omega$, definita da $\mathbf{X}_{>}((k_0, k_1, \dots)) := (k_1, k_2, \dots)$, e che X_0 e $\mathbf{X}_{>}$ sono variabili aleatorie indipendenti. Quest’ultimo fatto segue intuitivamente dalla Proposizione 3.39, che però è stata enunciata e dimostrata per famiglie finite di variabili aleatorie. In alternativa, per mostrare l’indipendenza di X_0 e $\mathbf{X}_{>}$ è sufficiente determinarne le distribuzioni congiunte e marginali, a partire da (4.58), con calcoli analoghi a quelli svolti nella dimostrazione della Proposizione 4.26.

Dimostriamo ora il seguente risultato, elementare ma forse non intuitivo.

Lemma 4.28. *Siano X e S due variabili aleatorie indipendenti, a valori in \mathbb{N}_0 , definite su uno spazio di probabilità (Ω, \mathcal{Q}) . Se $X \sim \text{Unif}\{0, \dots, n\}$, con $n \in \mathbb{N}_0$, allora*

$$\mathbb{Q}(S = m | X + S = n) = \mathbb{Q}(S = m | S \leq n), \quad \forall m \in \{0, \dots, n\}.$$

Dimostrazione. Si tratta di un semplice calcolo: per l’indipendenza di X e S e il fatto che $\mathbb{Q}(X = k) = \frac{1}{n+1} \mathbb{1}_{\{0, \dots, n\}}(k)$,

$$\mathbb{Q}(X + S = n) = \sum_{\ell=0}^n \mathbb{Q}(S = \ell) \mathbb{Q}(X = n - \ell) = \frac{1}{n+1} \sum_{\ell=0}^n \mathbb{Q}(S = \ell) = \frac{\mathbb{Q}(S \leq n)}{n+1},$$

pertanto, per ogni $m \in \{0, \dots, n\}$

$$\begin{aligned} \mathbb{Q}(S = m | X + S = n) &= \frac{\mathbb{Q}(S = m, X + S = n)}{\mathbb{Q}(X + S = n)} = \frac{\mathbb{Q}(S = m, X = n - m)}{\mathbb{Q}(S \leq n)/(n+1)} \\ &= \frac{\mathbb{Q}(S = m) \mathbb{Q}(X = n - m)}{\mathbb{Q}(S \leq n)/(n+1)} = \frac{\mathbb{Q}(S = m)}{\mathbb{Q}(S \leq n)} = \mathbb{Q}(S = m | S \leq n), \end{aligned}$$

e la dimostrazione è completa. \square

Possiamo finalmente riscrivere la relazione (4.62) nel modo seguente:

$$P_{n,\beta}(S_{>} = m) = Q_{\beta}(S_{>} = m \mid S_{>} \leq n), \quad \forall n \in \mathbb{N}, \forall m \in \{0, \dots, n\}. \quad (4.65)$$

Si noti che abbiamo volutamente omissso la dipendenza da n nella probabilità $Q_{\beta} = Q_{n,\beta}$, perché il membro destro di questa relazione dipende solo dalla distribuzione della variabile aleatoria $S_{>}$ rispetto a $Q_{n,\beta}$, la quale è *indipendente da n* . Infatti, per la relazione (4.61), $S_{>}$ è funzione delle $(X_i)_{i \geq 1}$, che per la Proposizione 4.26 sono variabili aleatorie indipendenti con distribuzioni marginali $\text{Geo}_0(1 - e^{-\beta \epsilon_i})$, che dipendono dunque solo da β e non da $n \in \mathbb{N}$. La dipendenza da n è concentrata soltanto nel condizionamento rispetto all'evento $\{S_{>} \leq n\}$.

Abbiamo ottenuto un risultato fondamentale: grazie alla relazione (4.65), possiamo affermare che *il numero $S_{>}$ di particelle al di fuori dello stato fondamentale, in un gas perfetto di n bosoni a temperatura inversa β , è distribuito come la somma di (un numero infinito di) variabili aleatorie geometriche indipendenti, con parametri espliciti che dipendono solo da β , condizionata a valere meno di n* . Questa descrizione permette di studiare in modo molto preciso le proprietà di $S_{>}$, rispetto alla probabilità $P_{n,\beta}$, nel limite termodinamico $n \rightarrow \infty$.[†] Noi ci accontenteremo di dedurre alcune conseguenze, semplici ma molto rilevanti.

Passo 5. Sfruttiamo ora la relazione (4.65), in modo da ottenere una stima esplicita sul numero medio di particelle al di fuori dello stato fondamentale $E_{n,\beta}(S_{>})$ nel limite termodinamico $n \rightarrow \infty$, dove $E_{n,\beta}$ indica il valore atteso rispetto alla probabilità $P_{n,\beta}$. Per chiarezza, il valore atteso rispetto alla probabilità $Q_{\beta} = Q_{n,\beta}$ verrà indicato con $E_{Q_{\beta}}$. Deduciamo allora da (4.65) che

$$E_{n,\beta}(S_{>}) = E_{Q_{\beta}}(S_{>} \mid S_{>} \leq n), \quad \forall n \in \mathbb{N}. \quad (4.66)$$

dove naturalmente $E_{Q_{\beta}}(\cdot \mid S_{>} \leq n)$ indica il valore atteso rispetto alla probabilità condizionale $Q_{\beta}(\cdot \mid S_{>} \leq n)$.

Intuitivamente, se si rimuove il condizionamento rispetto all'evento $\{S_{>} \leq n\}$, nel membro destro di (4.66), si ottiene qualcosa di più grande. Ciò segue in effetti dalla Proposizione 3.81: dato che la funzione $g(x) := x$ è crescente, mentre $h(x) := \mathbb{1}_{(-\infty, n]}(x)$ è decrescente, si ha

$$E_{Q_{\beta}}(S_{>} \mathbb{1}_{(-\infty, n]}(S_{>})) \geq E_{Q_{\beta}}(S_{>}) E_{Q_{\beta}}(\mathbb{1}_{(-\infty, n]}(S_{>})) = E_{Q_{\beta}}(S_{>}) Q_{\beta}(S_{>} \leq n),$$

e deduciamo dunque dalla relazione (4.66) che

$$E_{n,\beta}(S_{>}) \leq E_{Q_{\beta}}(S_{>}), \quad \forall n \in \mathbb{N}, \forall \beta > 0. \quad (4.67)$$

[†] Da un punto di vista tecnico, il passaggio dalla relazione (4.62) alla (4.65) costituisce una semplificazione notevole. Infatti, l'evento condizionante nel membro destro è passato dal vincolo “microscopico” $\{X_0 + S_{>} = n\}$, in cui si specifica un singolo valore di una variabile aleatoria, al vincolo “macroscopico” $\{S_{>} \leq n\}$, in cui si specifica un'ampia famiglia di valori.

Per calcolare il valore di $E_{Q_\beta}(S_>)$, ricordiamo che $X_i \sim \text{Geo}_0(1 - e^{-\beta \varepsilon_i})$ rispetto alla probabilità Q_β , pertanto, ricordando l'Osservazione 3.114,

$$E_{Q_\beta}(X_i) = \frac{1}{1 - e^{-\beta \varepsilon_i}} - 1 = \frac{e^{-\beta \varepsilon_i}}{1 - e^{-\beta \varepsilon_i}} = \frac{1}{e^{\beta \varepsilon_i} - 1}, \quad \forall i \in \mathbb{N}.$$

Applicando la linearità del valor medio alla somma *infinita* $S_> = \sum_{i=1}^{\infty} X_i$, otteniamo

$$E_{Q_\beta}(S_>) = \sum_{i \in \mathbb{N}} E_{Q_\beta}(X_i) = \sum_{i=1}^{\infty} \frac{1}{e^{\beta \varepsilon_i} - 1}. \quad (4.68)$$

Trattandosi di una somma infinita, la linearità del valor medio va giustificata. È possibile farlo, ad esempio, esprimendo la densità discreta di $S_>$ con una procedura di troncamento, usando la decomposizione di Ω in (4.57), con argomenti analoghi a quelli usati nelle righe successive a (4.58). I dettagli (piuttosto noiosi) sono lasciati al lettore interessato.

Combinando le relazioni (4.67) e (4.68), otteniamo infine la stima esplicita cercata:

$$E_{n,\beta}(S_>) \leq \sum_{i=1}^{\infty} \frac{1}{e^{\beta \varepsilon_i} - 1}, \quad \forall n \in \mathbb{N}, \forall \beta > 0. \quad (4.69)$$

Sottolineiamo che abbiamo maggiorato $E_{n,\beta}(S_>)$ con una quantità che non dipende esplicitamente da n (anche se, come vedremo, le energie ε_i possono dipendere da n).

Passo 6. L'ultimo passo della dimostrazione consiste nello specificare le energie ε_i e nel dedurre, finalmente, la condensazione di Bose-Einstein annunciata.

Se le particelle del sistema, di massa μ , sono contenute in una scatola, diciamo cubica per semplicità, di volume V , i livelli energetici sono esplicitamente noti dalla meccanica quantistica: usando per comodità l'insieme di indici (numerabile) $(\mathbb{Z})^3$ invece di \mathbb{N}_0 , le energie $(\tilde{\varepsilon}_{k,l,m})_{k,l,m \in \mathbb{Z}}$ sono date da

$$\tilde{\varepsilon}_{k,l,m} = \frac{h^2}{2\mu V^{2/3}}(k^2 + l^2 + m^2), \quad (4.70)$$

dove $h \simeq 6.6 \cdot 10^{-34} Js$ indica la costante di Planck. Si noti che lo stato fondamentale, quello con energia minima, è $(k, l, m) = (0, 0, 0)$ e l'energia corrispondente è $\varepsilon_{0,0,0} = 0$, quindi non c'è bisogno di sottrarre nessuna costante.

Ordinando le energie $(\tilde{\varepsilon}_{k,l,m})_{k,l,m \in \mathbb{Z}}$ in modo crescente, si ottiene una successione $(\varepsilon_i)_{i \in \mathbb{N}_0}$ tale che

$$\varepsilon_0 = 0, \quad \varepsilon_i \sim \frac{h^2}{2\mu V^{2/3}} \left(\frac{3i}{4\pi} \right)^{2/3} \quad \text{per } i \rightarrow +\infty, \quad (4.71)$$

dunque la condizione (4.53) è soddisfatta. Per semplicità, nel seguito utilizzeremo le energie $(\varepsilon_i)_{i \in \mathbb{N}_0}$ date precisamente dall'espressione nel membro destro in (4.71), in

modo da limitarci all'uso dell'integrale di Riemann in \mathbb{R} , invece che in \mathbb{R}^3 . I risultati asintotici che otterremo sono identici a quelli dati dall'espressione esatta (4.70).

Indichiamo con $\rho := n/V$ la *densità* delle particelle nella scatola. Per avere risultati fisicamente rilevanti, quando si considera il limite termodinamico $n \rightarrow \infty$ occorre contemporaneamente mandare anche il volume V all'infinito, in modo che la densità di particelle ρ resti costante. In questo modo, i risultati ottenuti nel limite, che dipenderanno dal valore di ρ , descriveranno con ottima approssimazione le proprietà di un sistema con un numero di particelle finito ma molto elevato, con densità ρ . Fissiamo dunque $\rho \in (0, \infty)$ e poniamo $V := n/\rho$, così che

$$\varepsilon_i = \frac{C}{n^{2/3}} i^{2/3}, \quad \text{dove} \quad C := \frac{h^2}{2\mu} \left(\frac{3\rho}{4\pi} \right)^{2/3}. \quad (4.72)$$

Sostituiamo quindi questa espressione in (4.69), ottenendo

$$E_{n,\beta}(S_>) \leq \sum_{i=1}^{\infty} g\left(\frac{i}{n}\right), \quad \text{dove} \quad g(x) := \frac{1}{e^{\beta C x^{2/3}} - 1}.$$

La funzione positiva g è integrabile secondo Riemann (in senso generalizzato) in $[0, \infty)$, perché il decadimento all'infinito è esponenziale. Pertanto, le somme di Riemann convergono verso l'integrale:

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^{\infty} f\left(\frac{i}{n}\right) = \int_0^{\infty} f(x) dx = \int_0^{\infty} \frac{1}{e^{\beta C x^{2/3}} - 1} dx = \frac{3}{(\beta C)^{3/2}} \int_0^{\infty} \frac{r^2}{e^{r^2} - 1} dr,$$

dove per l'ultima uguaglianza abbiamo usato il cambio di variabili $x = (\beta C)^{-3/2} r^3$. L'ultimo integrale, per quanto non del tutto esplicito, può essere espresso come

$$\int_0^{\infty} \frac{r^2}{e^{r^2} - 1} dr = \frac{\sqrt{\pi}}{4} \sum_{\ell=1}^{\infty} \frac{1}{\ell^{3/2}} = \frac{\sqrt{\pi}}{4} \zeta\left(\frac{3}{2}\right), \quad \text{dove} \quad \zeta(a) := \sum_{\ell=1}^{\infty} \frac{1}{\ell^a}.$$

Non dimostreremo questa uguaglianza. Ad ogni modo, quel che è rilevante è che l'integrale sia una costante finita.

Dalle stime precedenti, ricordando che $\beta = 1/k_B T$, dove indichiamo con $k_B \simeq 1.38 \cdot 10^{-23} J/(mol K)^{-1}$ costante di Boltzmann e con T la temperatura assoluta, si ha

$$\limsup_{n \rightarrow \infty} E_{n,\beta} \left(\frac{S_>}{n} \right) \leq \frac{3}{(\beta C)^{3/2}} \frac{\sqrt{\pi}}{4} \zeta\left(\frac{3}{2}\right) = \left(\frac{(2\pi\mu k_B)^{3/2} \zeta(\frac{3}{2})}{h^3} \right) \frac{T^{3/2}}{\rho}.$$

Consideriamo ora il numero di occupazione X_0 dello stato fondamentale. Si noti che $P_{n,\beta}(X_0 + S_> = n) = 1$, dal momento che $X_0 + S_>$ è il numero totale di particelle nel sistema. Di conseguenza $E_{n,\beta}(X_0) = n - E_{n,\beta}(S_>)$ e dunque

$$\liminf_{n \rightarrow \infty} E_{n,\beta} \left(\frac{X_0}{n} \right) \geq 1 - C \frac{T^{3/2}}{\rho} \quad \text{dove} \quad C := \frac{(2\pi\mu k_B)^{3/2} \zeta(\frac{3}{2})}{h^3}. \quad (4.73)$$

Abbiamo finalmente ottenuto il risultato cercato. Se è soddisfatta la condizione $1 - CT^{3/2}/\rho > 0$, il numero medio di particelle nello stato fondamentale è una frazione strettamente positiva del numero totale n di particelle nel sistema. Sottolineiamo che si tratta di una condizione esplicita, funzione della temperatura T e della densità di particelle ρ , che è verificata se T è sufficientemente bassa e/o se ρ è sufficientemente elevata. Con un po' più di lavoro, è possibile mostrare (non lo faremo) che la stima dal basso ottenuta in (4.73) è precisa, nel senso che

$$\lim_{n \rightarrow \infty} E_{n,\beta} \left(\frac{X_0}{n} \right) = \left(1 - C \frac{T^{3/2}}{\rho} \right)^+, \quad (4.74)$$

dove ricordiamo che $x^+ := \max\{x, 0\}$ indica la parte positiva. Ciò significa in particolare che, per ogni valore fissato $\rho > 0$ della densità, c'è una temperatura critica $T_c = T_c(\rho) := (\rho/C)^{2/3}$ tale che

$$\lim_{n \rightarrow \infty} E_{n,\beta} \left(\frac{X_0}{n} \right) = \begin{cases} 0 & \text{se } T > T_c(\rho) \\ \left(1 - C \frac{T^{3/2}}{\rho} \right) > 0 & \text{se } T \leq T_c(\rho) \end{cases}. \quad (4.75)$$

Si dice che il sistema esibisce una *transizione di fase*: una piccola variazione della temperatura, da T_c a $T_c - \varepsilon$, con $\varepsilon > 0$, produce l'effetto macroscopico di far “collapsare” una frazione delle n particelle nello stesso stato fondamentale.

Osservazione 4.29. Il fenomeno appena dimostrato non si osserverebbe se le particelle fossero distinguibili. Come accennato all'inizio del paragrafo, in questo caso lo spazio delle configurazioni appropriato per un sistema di n particelle, ciascuna delle quali può assumere r stati, è dato dall'insieme $\Omega_{n,r}$ delle funzioni da $\{1, \dots, n\}$ in $\{1, \dots, r\}$, definito in (2.15). Rimuovendo l'ipotesi il numero di stati sia finito ed etichettando gli stati con i numeri $0, 1, 2, \dots$, consideriamo l'insieme Ω_n delle funzioni da $\{1, \dots, n\}$ in $\mathbb{N}_0 = \{0, 1, 2, \dots\}$, che possiamo rappresentare come

$$\Omega_n := (\mathbb{N}_0)^n = \{ \mathbf{s} = (s_1, \dots, s_n) : s_i \in \mathbb{N}_0, \forall 1 \leq i \leq n \},$$

dove $\mathbf{s} = (s_1, \dots, s_n)$ significa che l' i -esima particella è nello stato s_i .

Se le particelle non sono interagenti (gas perfetto), l'energia totale del sistema è data dalla funzione $H_n : \Omega_n \rightarrow \mathbb{R}$ definita da

$$H_n(\mathbf{s}) := \sum_{i=1}^n \varepsilon_{s_i}, \quad \forall \mathbf{s} = (s_1, \dots, s_n) \in \Omega_n,$$

dove $(\varepsilon_i)_{i \in \mathbb{N}_0}$ indica come prima la successione crescente delle energie, in cui assumiamo che $\varepsilon_0 = 0$. La corrispondente misura di Gibbs su Ω_n è data, per ogni $\beta > 0$, dalla probabilità

$$P_{n,\beta}^{MB}(\{\mathbf{s}\}) := \frac{e^{-\beta H_n(\mathbf{s})}}{Z_{n,\beta}^{MB}}, \quad \forall \mathbf{s} = (s_1, \dots, s_n) \in \Omega_n,$$

dove “MB” sta per Maxwell-Boltzmann.

A differenza di quanto accade per particelle indistinguibili, in questo caso la funzione di partizione può essere calcolata esplicitamente:

$$Z_{n,\beta}^{MB} := \sum_{\mathbf{s} \in \Omega_n} e^{-\beta H_n(\mathbf{s})} = \left(\sum_{i=0}^{\infty} e^{-\beta \varepsilon_i} \right)^n = \left(1 + \sum_{i=1}^{\infty} e^{-\beta \varepsilon_i} \right)^n,$$

dove l'espressione ottenuta è finita per ogni $\beta > 0$, grazie all'assunzione (4.53). Indicando sempre con $X_0 : \Omega_n \rightarrow \mathbb{N}_0$ la variabile aleatoria che conta il numero di particelle nello stato fondamentale, che in questo caso è data da $X_0(\mathbf{s}) := \sum_{i=1}^n \mathbb{1}_{\{s_i=0\}}$, lasciamo al lettore il compito di dimostrare che, per ogni $k \in \{0, \dots, n\}$,

$$P_{n,\beta}^{MB}(X_0 = k) = \frac{\binom{n}{k} \left(\sum_{i=1}^{\infty} e^{-\beta \varepsilon_i} \right)^{n-k}}{\left(1 + \sum_{i=1}^{\infty} e^{-\beta \varepsilon_i} \right)^n} = \binom{n}{k} q^k (1-q)^{n-k},$$

avendo posto

$$q := \frac{1}{1 + \sum_{i=1}^{\infty} e^{-\beta \varepsilon_i}}.$$

Quindi, per un gas perfetto di particelle distinguibili, il numero di particelle X_0 nello stato fondamentale ha distribuzione $\text{Bin}(n, q)$. Pertanto $E_{n,\beta}^{MB}(X_0) = nq$ e

$$E_{n,\beta}^{MB} \left(\frac{X_0}{n} \right) = q = \frac{1}{1 + \sum_{i=1}^{\infty} e^{-\beta \varepsilon_i}}.$$

Ricordando l'espressione (4.72) delle energie ε_i , in particolare il fatto che esse dipendono dal numero totale di particelle (attraverso il volume del sistema) e $\lim_{n \rightarrow \infty} \varepsilon_i = 0$, per ogni i fissato. Da ciò segue facilmente che

$$\lim_{n \rightarrow \infty} E_{n,\beta}^{MB} \left(\frac{X_0}{n} \right) = 0, \quad \forall \beta > 0,$$

in contrasto con (4.75). □

4.7 Note bibliografiche

Il nome *legge dei piccoli numeri* per i risultati relativi all'approssimazione di Poisson venne suggerito, alla fine del XIX secolo, dallo statistico russo L. von Bortkiewicz, in [9]. La versione della legge dei piccoli numeri qui proposta è tratta da un articolo di J.L. Hodges e L. Le Cam [30].

Il modello binomiale, presentato nel Paragrafo 4.2, venne proposto da J.C Cox, S.A. Ross, e M. Rubinstein [14] come una semplificazione del celebre modello, a tempo continuo, di Black & Scholes [7]. La monografia [43] di A. Pascucci e W.J. Runggaldier fornisce un'introduzione accessibile a numerosi modelli di finanza matematica a tempo discreto.

I Paragrafi 4.3 e 4.4 sono tratti in parte dal testo [1] di M. Aigner e G.M. Ziegler. La ricerca su questo argomento ebbe un grande sviluppo negli anni '80 del novecento, con alcuni fondamentali lavori di D. Aldous e P. Diaconis, fra i quali [2].

La formulazione delle passeggiate aleatorie in termini di successioni di variabili aleatorie è standard e si può trovare in molti testi di probabilità, introduttivi e non. In particolare, quanto presentato nel Paragrafo 4.5 è tratto dal testo [23] di W. Feller. Per approfondimenti sull'argomento, consigliamo il testo classico di F. Spitzer [51], e quello più recente di G.F. Lawler e V. Limic [35].

La condensazione di Bose-Einstein venne prevista negli anni '20 del novecento da S.N. Bose [10] e A. Einstein [19]. La sua osservazione sperimentale, a opera di E. Cornell e C. Wieman (1995), valse loro il Premio Nobel per la Fisica nel 2001.

Spazi di probabilità e variabili aleatorie generali

Sommario. Questo capitolo rappresenta una sorta di intermezzo. Esso contiene nozioni fondamentali, alla base dell'analisi matematica e del calcolo delle probabilità, ma il cui impatto in questo libro, di natura introduttiva, è limitato. Tuttavia, molti modelli probabilistici che tratteremo nel seguito non possono essere definiti a partire da spazi di probabilità discreti. Lo scopo di questo capitolo è quindi di fornire le nozioni “minime” affinché la trattazione risulti coerente e rigorosa. Contrariamente ai capitoli precedenti, gran parte delle dimostrazioni saranno omesse.

5.1 σ -algebre e misure di probabilità

Con gli spazi di probabilità discreti, abbiamo visto molti aspetti importanti del calcolo delle probabilità, senza dover ricorrere a strumenti di analisi matematica troppo sofisticati. La teoria fin qui sviluppata non ci permette, però, di affrontare alcuni argomenti fondamentali, sia dal punto di vista teorico che applicativo.

Un primo problema riguarda la definizione di variabili aleatorie il cui insieme dei valori sia più che numerabile. Molte grandezze che trattiamo quotidianamente (tempi, masse, lunghezze, ...) sono descritte in modo naturale come quantità che possono assumere qualunque valore di un intervallo di \mathbb{R} . È ovviamente impossibile, in uno spazio di probabilità discreto, definire una variabile aleatoria X per cui $X(\Omega)$ sia un intervallo di \mathbb{R} , dato che $X(\Omega)$ è sempre al più numerabile.

Un'altra questione importante riguarda lo studio delle *successioni* di variabili aleatorie. Come abbiamo mostrato nell'Osservazione 3.42, uno spazio di probabilità discreto è troppo “povero” perché in esso si possano definire successioni di variabili aleatorie discrete interessanti, ad esempio indipendenti e con la stessa distribuzione non concentrata in un punto.

Risulta quindi naturale cercare di estendere la nozione di spazio di probabilità, ammettendo che lo spazio campionario possa essere più che numerabile. Per avere un'idea del tipo di problemi che sorgono, vediamo due esempi significativi.

Esempio 5.1. Il concetto di probabilità uniforme (Esempio 1.13) è chiaro e naturale se lo spazio campionario è finito. È possibile estendere tale nozione ad uno spazio campionario continuo, ad esempio un intervallo limitato di \mathbb{R} ? Per fissare le idee, si può formalizzare l'idea di "scegliere a caso" un punto di $\Omega = [0, 1]$?

Se P è la probabilità "uniforme" su Ω che stiamo cercando di definire, è naturale richiedere che

$$P([a, b]) = b - a, \quad \forall 0 \leq a \leq b \leq 1. \quad (5.1)$$

In tal modo, oltre al fatto che $P(\Omega) = 1$, si ha che la probabilità di un intervallo dipende solo dalla sua lunghezza geometrica, e non dalla sua "posizione" in $[0, 1]$. A questo punto, è naturale chiedersi se sia possibile estendere la funzione P , definita in (5.1) sugli intervalli, a tutti i sottoinsiemi di $[0, 1]$, in modo che l'estensione sia σ -additiva (ossia verifichi l'assioma (P2) della Definizione 1.5). Si noti che da (5.1) segue che $P(\{x\}) = P([x, x]) = 0$ per ogni $x \in [0, 1]$, quindi una tale P non può essere una probabilità discreta (si ricordi la Definizione 3.7). \square

Esempio 5.2. Nel Paragrafo 1.3.4 abbiamo definito un modello probabilistico $(\Omega_n = \{0, 1\}^n, P_{n,p})$ per n prove ripetute indipendenti con probabilità di successo p . Pensiamo ora di effettuare una successione infinita di prove ripetute. La scelta naturale per lo spazio campionario è allora

$$\Omega = \{0, 1\}^{\mathbb{N}} = \{\omega = (\omega_1, \omega_2, \dots) : \omega_i \in \{0, 1\} \text{ per ogni } i \in \mathbb{N}\}.$$

È ben noto che Ω non è numerabile (è in corrispondenza biunivoca con l'insieme delle parti di \mathbb{N} , che ha la stessa cardinalità di \mathbb{R}). La probabilità P che vogliamo costruire sui sottoinsiemi di Ω dovrà soddisfare un requisito del tutto naturale: se consideriamo un evento che dipende solo dagli esiti di un numero finito n di prove, la sua probabilità dovrà essere uguale a quella calcolata nel Paragrafo 1.3.4. In altre parole, ricordando la formula (1.62), se $x_1, x_2, \dots, x_n \in \{0, 1\}$, dovrà essere

$$P(\{\omega \in \Omega : \omega_1 = x_1, \dots, \omega_n = x_n\}) = p^{\sum_{i=1}^n x_i} (1-p)^{n-\sum_{i=1}^n x_i}. \quad (5.2)$$

Come nell'esempio precedente, il problema è di stabilire se sia possibile estendere P a tutti i sottoinsiemi di Ω , in modo che soddisfi l'assioma di σ -additività. Se tale estensione esiste, allora necessariamente $P(\{\omega\}) = 0$ per ogni $\omega \in \Omega$, come abbiamo già mostrato nell'Osservazione 3.42. \square

In entrambi gli esempi, la funzione P è definita dapprima su una famiglia di insiemi "semplici" e ci si chiede se sia possibile estenderla a $\mathcal{P}(\Omega)$, in modo che l'estensione sia σ -additiva. Si può dimostrare che, in entrambi i casi, la risposta è *negativa*. La soluzione a questo problema è quella di ridimensionare l'obiettivo iniziale di estendere P a *tutto* $\mathcal{P}(\Omega)$, accontentandosi di estenderla a una classe di sottoinsiemi $\mathcal{A} \subseteq \mathcal{P}(\Omega)$ sufficientemente ampia. Le proprietà che è opportuno richiedere sulla classe \mathcal{A} sono descritte nella seguente definizione.

Definizione 5.3 (σ -algebra). Sia Ω un insieme non vuoto. Una famiglia $\mathcal{A} \subseteq \mathcal{P}(\Omega)$ di sottoinsiemi di Ω si dice σ -algebra se soddisfa le seguenti proprietà.

- (i) $\emptyset \in \mathcal{A}$.
- (ii) Se $A \in \mathcal{A}$, allora $A^c \in \mathcal{A}$.
- (iii) Se $(A_n)_{n \in \mathbb{N}}$ è una successione numerabile di elementi di \mathcal{A} , allora anche la loro unione $\bigcup_{n \in \mathbb{N}} A_n$ è un elemento di \mathcal{A} .

Gli elementi $A \in \mathcal{A}$ vengono detti sottoinsiemi *misurabili* di Ω .

Si noti che, se \mathcal{A} è una σ -algebra su Ω , necessariamente $\Omega \in \mathcal{A}$, come segue dalle proprietà (i) e (ii). Inoltre \mathcal{A} è chiusa per unioni finite, oltre che numerabili: infatti, ogni famiglia finita (A_1, \dots, A_n) di elementi di \mathcal{A} può essere completata a una successione, ponendo $A_k = \emptyset$ per $k > n$, senza modificarne l'unione. Infine, dall'identità di de Morgan $\bigcap_n A_n = (\bigcup_n A_n^c)^c$ segue che una σ -algebra è anche chiusa per intersezione, sia di una famiglia finita sia di una successione numerabile.

Definizione 5.4 (Spazio misurabile). Una coppia (Ω, \mathcal{A}) formata da un insieme e da una σ -algebra di suoi sottoinsiemi si dice *spazio misurabile*.

Definizione 5.5 (Spazio di probabilità generale). Sia (Ω, \mathcal{A}) uno spazio misurabile. Una funzione

$$P : \mathcal{A} \rightarrow [0, 1]$$

si dice *probabilità* (o *misura di probabilità*) se valgono le seguenti proprietà:

- (P1) $P(\Omega) = 1$.
- (P2) (σ -additività) Per ogni successione numerabile $(A_n)_{n \in \mathbb{N}}$ di elementi disgiunti di \mathcal{A} , si ha

$$P\left(\bigcup_{n \in \mathbb{N}} A_n\right) = \sum_{n \in \mathbb{N}} P(A_n).$$

La terna (Ω, \mathcal{A}, P) , dove (Ω, \mathcal{A}) è uno spazio misurabile e P è una probabilità su (Ω, \mathcal{A}) , è detta *spazio di probabilità (generale)*.

In altre parole, la nozione di probabilità su un insieme generale Ω , non necessariamente numerabile, è definita dallo stesso sistema di assiomi della probabilità visti nel Capitolo 1, con la differenza che la probabilità è definita su una σ -algebra di sottoinsiemi di Ω , che non coincide necessariamente con $\mathcal{P}(\Omega)$.

Osservazione 5.6. Se (Ω, P) è uno spazio di probabilità discreto, introdotto nella Definizione 1.5, esso si può identificare con lo spazio di probabilità $(\Omega, \mathcal{P}(\Omega), P)$. Più in generale, ogni probabilità discreta P su un insieme generico Ω (si ricordi la Definizione 3.7) determina uno spazio di probabilità $(\Omega, \mathcal{P}(\Omega), P)$. \square

In analogia a quanto fatto nel caso discreto, Ω sarà chiamato *spazio campionario* e gli elementi di \mathcal{A} saranno chiamati *eventi*. Le proprietà di una σ -algebra servono proprio a garantire che, lavorando sugli eventi mediante unioni e intersezioni, finite o numerabili, e passaggio al complementare, si ottengano ancora eventi.

Esattamente come nel caso discreto, in uno spazio di probabilità (Ω, \mathcal{A}, P) vale l'addittività finita: se $A_1, A_2, \dots, A_N \in \mathcal{A}$ sono eventi disgiunti,

$$P\left(\bigcup_{n=0}^N A_n\right) = \sum_{n=0}^N P(A_n). \quad (5.3)$$

Più in generale, tutti i risultati del Paragrafo 1.1.4 continuano a valere nel contesto generale ora introdotto, a patto di restringere i risultati alla σ -algebra degli eventi, e *senza che le dimostrazioni richiedano alcuna modifica*. Pertanto, nel seguito, ci riferiremo alle Proposizioni 1.16, 1.18, 1.19 e al Corollario 1.20 anche quando questi risultati verranno usati in spazi di probabilità generali. Anche le nozioni di probabilità condizionale e di indipendenza di eventi, studiate nel Paragrafo 1.3, si estendono senza difficoltà al contesto degli spazi di probabilità generali.

Torniamo ora al problema enunciato negli Esempi 5.1 e 5.2. In entrambi i casi, avevamo definito una funzione $P: \mathcal{C} \rightarrow [0, 1]$, dove \mathcal{C} è la famiglia degli sottointervalli chiusi di $[0, 1]$ nell'Esempio 5.1, mentre \mathcal{C} è la famiglia degli insiemi del tipo $\{\omega: \omega_1 = x_1, \dots, \omega_N = x_N\}$ nell'Esempio 5.2. È facile vedere che, in entrambi i casi, \mathcal{C} non è una σ -algebra. Il problema diventa allora di trovare una σ -algebra $\mathcal{A} \supseteq \mathcal{C}$ e una probabilità su (Ω, \mathcal{A}) che estenda la P originaria. La scelta della σ -algebra \mathcal{A} si può fare in modo “canonico”, grazie al seguente risultato.

Proposizione 5.7. *Sia Ω un insieme arbitrario.*

- (i) *Se $(\mathcal{A}_\alpha)_{\alpha \in I}$ è una famiglia di σ -algebre di sottoinsiemi di Ω , indicizzata da un insieme arbitrario I , allora la loro intersezione*

$$\bigcap_{\alpha \in I} \mathcal{A}_\alpha := \{A \subseteq \Omega: A \in \mathcal{A}_\alpha \forall \alpha \in I\}$$

è una σ -algebra di sottoinsiemi di Ω .

- (ii) *Per ogni famiglia $\mathcal{C} \subseteq \mathcal{P}(\Omega)$ di sottoinsiemi di Ω , è ben definita la più piccola σ -algebra \mathcal{A} contenente \mathcal{C} , con la seguente proprietà:*

$$\text{per ogni } \sigma\text{-algebra } \mathcal{A}' \supseteq \mathcal{C}, \text{ si ha che } \mathcal{A}' \supseteq \mathcal{A} \supseteq \mathcal{C}.$$

Tale σ -algebra, denotata con $\sigma(\mathcal{C})$, è detta σ -algebra generata da \mathcal{C} .

Dimostrazione. La dimostrazione del punto (i) è lasciata come esercizio. Veniamo al punto (ii). Definiamo

$$\Xi := \{\mathcal{A}' \subseteq \mathcal{P}(\Omega): \mathcal{A}' \text{ è una } \sigma\text{-algebra contenente } \mathcal{C}\}.$$

Notiamo che $\Xi \neq \emptyset$, perché $\mathcal{P}(\Omega) \in \Xi$. Per il punto (i), la famiglia \mathcal{A} definita da

$$\mathcal{A} := \bigcap_{\mathcal{A}' \in \Xi} \mathcal{A}'$$

è una σ -algebra contenente \mathcal{C} ; inoltre, per definizione, si ha $\mathcal{A} \subseteq \mathcal{A}'$ per ogni σ -algebra \mathcal{A}' contenente \mathcal{C} . \square

Per quanto riguarda gli Esempi 5.1 e 5.2, è possibile dimostrare che *esiste un'unica probabilità* che estende P a $\sigma(\mathcal{C})$. Si tratta di un risultato altamente non banale, al di là degli scopi di questo libro, su cui ritorneremo nel Paragrafo 5.4. In entrambi i casi, è possibile estendere P ad una σ -algebra strettamente più grande di $\sigma(\mathcal{C})$ (ma non a tutto $\mathcal{P}(\Omega)$, come già detto).

5.2 Variabili aleatorie generali

L'aver ammesso che una probabilità possa essere definita solo su una classe \mathcal{A} di sottoinsiemi di Ω , e non su tutto l'insieme delle parti $\mathcal{P}(\Omega)$, impone una maggiore attenzione alla definizione di variabile aleatoria. Infatti, se (Ω, \mathcal{A}, P) è uno spazio di probabilità e $X : \Omega \rightarrow E$ una funzione qualunque, a valori in un insieme generico E , la scrittura $P(X \in C)$ potrebbe non essere definita per tutti i sottoinsiemi $C \subseteq E$, in quanto non necessariamente $\{X \in C\} := X^{-1}(C) \in \mathcal{A}$. Per evitare questo problema, si richiede che anche l'insieme di arrivo E sia munito di una σ -algebra \mathcal{E} , e non tutte le funzioni $X : \Omega \rightarrow E$ verranno chiamate variabili aleatorie.

Definizione 5.8 (Variabile aleatoria generale). Sia (Ω, \mathcal{A}, P) uno spazio di probabilità, e sia (E, \mathcal{E}) uno spazio misurabile. Una funzione

$$X : \Omega \rightarrow E$$

si dice *variabile aleatoria* se per ogni $C \in \mathcal{E}$ si ha

$$\{X \in C\} := X^{-1}(C) \in \mathcal{A}.$$

Osservazione 5.9. Nel caso in cui $\mathcal{A} = \mathcal{P}(\Omega)$ (in particolare, quando lo spazio di probabilità è discreto, per l'Osservazione 5.6), *tutte* le funzioni $X : \Omega \rightarrow E$ sono variabili aleatorie, coerentemente con la Definizione 3.1. \square

Si noti che la nozione di variabile aleatoria a valori in un insieme E dipende dalla scelta della σ -algebra \mathcal{E} su E . Nel caso in cui $E = \mathbb{R}^n$ (in particolare, per $n = 1$, se $E = \mathbb{R}$) sceglieremo sempre la cosiddetta σ -algebra di Borel $\mathcal{B}(\mathbb{R}^n)$, definita come la σ -algebra generata dai sottoinsiemi aperti, nel senso della Proposizione 5.7. Tale scelta verrà sempre sottointesa nel seguito. Come nel caso discreto, chiameremo *reali* le variabili aleatorie a valori in \mathbb{R} , mentre le variabili aleatorie a valori in \mathbb{R}^n saranno dette *vettoriali*, o *vettori aleatori*, di dimensione n .

Osservazione 5.10. La σ -algebra $\mathcal{B}(\mathbb{R}^n)$ contiene tutti i sottoinsiemi di \mathbb{R}^n che si incontrano tipicamente: vi appartengono tutti gli insiemi aperti e tutti i chiusi, le loro unioni, intersezioni numerabili e i rispettivi complementari, ecc. Tuttavia esistono sottoinsiemi di \mathbb{R}^n che non appartengono a $\mathcal{B}(\mathbb{R}^n)$, ossia $\mathcal{B}(\mathbb{R}^n) \subsetneq \mathcal{P}(\mathbb{R}^n)$.

Non esiste una descrizione “esplicita” di tutti gli elementi di $\mathcal{B}(\mathbb{R}^n)$, ma questo non costituisce un problema, perché ci si può ricondurre a lavorare con insiemi “semplici”. Ad esempio, dato uno spazio di probabilità (Ω, \mathcal{A}, P) , si può mostrare che, affinché un’applicazione $X : \Omega \rightarrow \mathbb{R}^n$ sia una variabile aleatoria, è sufficiente che $\{X \in C\} \in \mathcal{A}$ per ogni sottoinsieme $C \subseteq \mathbb{R}^n$ chiuso (oppure aperto). Per variabili aleatorie reali $X : \Omega \rightarrow \mathbb{R}$ basta anzi considerare le semirette $C = (-\infty, t]$, ossia X è una variabile aleatoria se (e solo se)

$$\{X \leq t\} = \{\omega \in \Omega : X(\omega) \leq t\} \in \mathcal{A}, \quad \forall t \in \mathbb{R}.$$

Ad ogni modo, si tratta di questioni avanzate, con cui non avremo direttamente a che fare in questo libro. \square

Se X è un vettore aleatorio di dimensione n e $g : \mathbb{R}^n \rightarrow \mathbb{R}^m$ è una funzione generica, non è detto che la funzione $g(X) : \Omega \rightarrow \mathbb{R}^m$ sia una variabile aleatoria, cioè non è detto che $\{g(X) \in C\} \in \mathcal{A}$ per ogni $C \in \mathcal{B}(\mathbb{R}^m)$. Per averne la garanzia, assumeremo che g sia una *funzione misurabile*, secondo la seguente definizione.

Definizione 5.11 (Funzione misurabile). Una funzione $g : \mathbb{R}^n \rightarrow \mathbb{R}^m$ si dice *misurabile* (rispetto alle σ -algre di Borel) se per ogni $C \in \mathcal{B}(\mathbb{R}^m)$ si ha che $g^{-1}(C) \in \mathcal{B}(\mathbb{R}^n)$.

La misurabilità è una proprietà molto debole, soddisfatta da tutte le funzioni con cui si lavora abitualmente e, in particolare, da tutte le funzioni che incontreremo in questo libro. Ad esempio, *sono misurabili tutte le funzioni continue tranne al più in un insieme finito o numerabile di punti* (e dunque i polinomi, l’esponenziale, ecc.). Inoltre, ogni funzione $g : \mathbb{R} \rightarrow \mathbb{R}$ monotona (crescente o decrescente) è misurabile, come si verifica facilmente usando l’Osservazione 5.10.

A ogni variabile aleatoria è associato un oggetto di fondamentale importanza, la sua distribuzione, che ora introduciamo, in analogia con la Definizione 3.10.

Definizione 5.12 (Distribuzione di una variabile aleatoria). Sia X una variabile aleatoria, definita sullo spazio di probabilità (Ω, \mathcal{A}, P) , a valori in (E, \mathcal{E}) . Si dice *distribuzione (o legge) di X* l’applicazione $\mu_X : \mathcal{E} \rightarrow [0, 1]$ definita da

$$\mu_X(C) := P(X \in C).$$

(La probabilità $P(X \in C)$ è ben definita perché $\{X \in C\} = X^{-1}(C) \in \mathcal{A}$ è un evento per ogni $C \in \mathcal{E}$, per definizione di variabile aleatoria.)

La dimostrazione del seguente risultato è lasciata per esercizio al lettore.

Proposizione 5.13. *La distribuzione μ_X di una variabile aleatoria X a valori in (E, \mathcal{E}) è una probabilità su (E, \mathcal{E}) .*

In particolare, se X è una variabile aleatoria reale, la sua distribuzione μ_X è una probabilità su \mathbb{R} (o, più precisamente, sullo spazio misurabile $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$).

Un'oggetto di grande importanza per lo studio delle variabili aleatorie reali è la *funzione di ripartizione*, già introdotta a suo tempo per variabili aleatorie discrete:

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

Si noti che $F_X(x) = \mu_X((-\infty, x])$, ossia la funzione di ripartizione è determinata dalla distribuzione. Il fatto cruciale è che vale anche il viceversa, ossia la funzione di ripartizione di una variabile aleatoria reale *ne caratterizza la distribuzione*.

Proposizione 5.14. *Due variabili aleatorie reali X e Y con la stessa funzione di ripartizione $F_X = F_Y$ hanno la stessa distribuzione: $\mu_X = \mu_Y$.*

Le proprietà della funzione di ripartizione contenute nella Proposizione 3.89 continuano a valere per variabili aleatorie generali, senza alcuna modifica nella dimostrazione: per ogni variabile aleatoria reale X , la funzione di ripartizione F_X

$$\text{è una funzione da } \mathbb{R} \text{ in } [0, 1] \text{ crescente, continua da destra, ha} \quad (5.4) \\ \text{limite 0 per } x \rightarrow -\infty \text{ e limite 1 per } x \rightarrow +\infty.$$

Come vedremo, queste proprietà *caratterizzano* le funzioni di ripartizione.

Una variabile aleatoria reale discreta è stata definita, nel Capitolo 3, come una funzione $X : \Omega \rightarrow \mathbb{R}$, dove (Ω, \mathcal{P}) è uno spazio di probabilità discreto; in particolare, dato che Ω è numerabile, l'immagine $X(\Omega)$ è un sottoinsieme numerabile di \mathbb{R} . È quest'ultima proprietà quella davvero rilevante. Pertanto, estendiamo la nozione di variabile aleatoria reale discreta nel modo seguente.

Definizione 5.15. Sia X una variabile aleatoria reale, definita su uno spazio di probabilità (Ω, \mathcal{A}, P) . Diremo che X è una *variabile aleatoria discreta* se l'insieme $X(\Omega) \subseteq \mathbb{R}$ dei valori assunti da X è finito o numerabile.

Se X è una variabile aleatoria discreta, in questo senso più generale, definendo la densità discreta $p_X(x) := P(X = x)$ nel modo abituale, per ogni $A \subseteq \mathbb{R}$ si ha

$$\begin{aligned} \mu_X(A) &= P(X \in A) = P(X \in A \cap X(\Omega)) = P\left(\bigcup_{x \in A \cap X(\Omega)} \{X = x\}\right) \\ &= \sum_{x \in A \cap X(\Omega)} P(X = x) = \sum_{x \in A} P(X = x) = \sum_{x \in A} p_X(x), \end{aligned} \quad (5.5)$$

perché l'immagine $X(\Omega)$ è un insieme numerabile. In particolare, prendendo $A = \mathbb{R}$,

$$\sum_{x \in \mathbb{R}} p_X(x) = \sum_{x \in \mathbb{R}} P(X = x) = 1. \quad (5.6)$$

In altri termini, la Proposizione 3.11 continua a valere: la distribuzione μ_X di una variabile aleatoria reale discreta X è una probabilità discreta su \mathbb{R} , associata alla densità discreta p_X . Di conseguenza, *tutte le proprietà viste nel Capitolo 3 che dipendono solo dalla densità discreta p_X continuano a valere per variabili aleatorie discrete nel senso della Definizione 5.15.*

Il fatto è che, su uno spazio di probabilità generale (Ω, \mathcal{A}, P) , ci possono essere variabili aleatorie reali *non* discrete, per cui le relazioni (5.5), (5.6) cessano di valere. In effetti, come vedremo nel Capitolo 6, per una vasta classe di variabili aleatorie reali X si ha addirittura $P(X = x) = 0$ per ogni $x \in \mathbb{R}$. Di conseguenza, *per variabili aleatorie generali, la densità discreta $P(X = x)$ non identifica la distribuzione di X .* Questa è una delle ragioni per cui assume grande importanza la funzione di ripartizione F_X , che diventa la nuova “carta d'identità” della distribuzione di X .

Concludiamo osservando che per *ogni* variabile aleatoria reale X si ha

$$F_X(x) - F_X(x^-) = P(X = x), \quad \forall x \in \mathbb{R}, \quad (5.7)$$

come mostra, senza modifiche nella dimostrazione, la Proposizione 3.90.

5.3 Indipendenza e valor medio

Come principio generale, tutti i risultati del Capitolo 3 nel cui enunciato *non* si fa riferimento esplicito alla densità discreta si possono estendere a variabili aleatorie generali (talvolta senza modifiche nelle dimostrazioni, altre volte con adattamenti non banali). Non ripeteremo tutti gli enunciati, limitandoci a richiamare brevemente quelli più importanti, e nel seguito adotteremo la convenzione di riferirci ad essi anche quando ne useremo l'estensione a spazi di probabilità generali.

L'indipendenza di variabili aleatorie generali è definita in completa analogia col caso discreto (si ricordino le Definizioni 3.28 e 3.30).

Definizione 5.16 (Indipendenza di variabili aleatorie). Siano X_1, X_2, \dots, X_n variabili aleatorie, definite sullo stesso spazio di probabilità (Ω, \mathcal{A}, P) a valori rispettivamente negli spazi misurabili $(E_1, \mathcal{E}_1), (E_2, \mathcal{E}_2), \dots, (E_n, \mathcal{E}_n)$. Esse si dicono *indipendenti* se per ogni scelta di $A_1 \in \mathcal{E}_1, A_2 \in \mathcal{E}_2, \dots, A_n \in \mathcal{E}_n$ si ha

$$P(X_1 \in A_1, X_2 \in A_2, \dots, X_n \in A_n) = \prod_{i=1}^n P(X_i \in A_i). \quad (5.8)$$

Data una famiglia arbitraria $(X_i)_{i \in I}$ di variabili aleatorie, tutte definite sullo stesso spazio di probabilità, esse si dicono indipendenti se ogni sottofamiglia finita è formata da variabili aleatorie indipendenti.

L'interpretazione è sempre la stessa: l'indipendenza di una famiglia $(X_i)_{i \in I}$ di variabili aleatorie significa intuitivamente che avere informazioni sul valore di alcune di esse non modifica la probabilità degli eventi che dipendono dalle altre.

Le proprietà dell'indipendenza descritte nel Paragrafo 3.2.4 continuano a valere nel contesto generale. Segnaliamo in particolare la Proposizione 3.39 ("indipendenza a blocchi") e la Proposizione 3.40 ("funzioni di variabili aleatorie indipendenti sono indipendenti"), di cui si fa uso frequente, spesso in modo implicito.

Veniamo ora alla nozione di *valor medio*. Per una variabile aleatoria reale discreta X (nel senso generale della Definizione 5.15), esso è definito esattamente come nella Definizione 3.43 data a suo tempo, ossia

$$E(X) := \sum_{x \in \mathbb{R}} x p_X(x) = \sum_{x \in \mathbb{R}} x P(X = x) \in [-\infty, +\infty], \quad (5.9)$$

a condizione che la famiglia di numeri reali $(x p_X(x))_{x \in \mathbb{R}}$ ammetta somma; in caso contrario, diremo che la variabile aleatoria X non ammette valor medio.

Per una variabile aleatoria reale X generale, non necessariamente discreta, osserviamo che $X^+ := \max(X, 0)$ e $X^- := -\min(X, 0)$ sono variabili aleatorie positive, e che si ha $X = X^+ - X^-$ e $|X| = X^+ + X^-$. Si dà quindi la definizione seguente.

Definizione 5.17 (Valor medio di una variabile aleatoria generale). Sia X una variabile aleatoria reale, definita su uno spazio di probabilità (Ω, \mathcal{A}, P) .

- Se $X \geq 0$, si definisce *valor medio di X* la quantità

$$E(X) := \sup \left\{ E(Y) : Y : \Omega \rightarrow \mathbb{R} \text{ è una variabile aleatoria discreta tale che } 0 \leq Y \leq X \right\} \in [0, +\infty].$$

(In particolare, $E(X)$ coincide con la formula (5.9) se X è discreta.)

- In generale, si dice che X ammette valor medio se *almeno una* tra le quantità $E(X^+)$ e $E(X^-)$ è finita, e in tal caso si definisce

$$E(X) := E(X^+) - E(X^-) \in [-\infty, +\infty].$$

Di conseguenza, X ammette valor medio *finito* se e solo se *entrambe* le quantità $E(X^+)$ e $E(X^-)$ sono finite, o equivalentemente (per linearità, vedi sotto) se la quantità $E(|X|) = E(X^+) + E(X^-)$ è finita.

Molte proprietà del valor medio, viste nel Capitolo 3, continuano a valere per variabili aleatorie generali. Ricordiamo in particolare la Proposizione 3.55, che esprime la *monotonia* e la *linearità* del valor medio:

- se $X \leq Y$ allora $E(X) \leq E(Y)$, purché X e Y ammettano valor medio;
- $E(aX + bY) = aE(X) + bE(Y)$ se X e Y ammettono valor medio finito, per ogni $a, b \in \mathbb{R}$; oppure se X e Y sono positive, per ogni $a, b \geq 0$.

Menzioniamo anche la Proposizione 3.57 (“variabili aleatorie limitate ammettono valor medio finito”) e la Proposizione 3.58 (“variabili aleatorie positive con valor medio nullo sono quasi certamente nulle”).

Osservazione 5.18. Una proprietà importante, che non dimostreremo, è che *il fatto che una variabile aleatoria reale X ammetta valor medio e , in caso affermativo, il valore di $E(X)$, dipendono solo dalla distribuzione μ_X di X* . Per variabili aleatorie discrete, questo è chiaro dalla formula (5.9), che tuttavia (come ogni formula in cui compare la densità discreta) non è più valida per variabili aleatorie generali.

Se X è un vettore aleatorio di dimensione n e $g: \mathbb{R}^n \rightarrow \mathbb{R}$ è una funzione misurabile, $g(X)$ è una variabile aleatoria reale, la cui distribuzione è una funzione di g e di μ_X , come si può verificare. Pertanto, *se X e Y sono vettori aleatori di dimensione n con la stessa distribuzione, allora $g(X)$ ammette valor medio se e solo se $g(Y)$ ammette valor medio, e in caso affermativo i due valori medi sono uguali.* \square

Dato che i polinomi e l'esponenziale sono funzioni continue, quindi misurabili, si possono definire varianza, momenti e funzione generatrice dei momenti per una variabile aleatoria reale X come nel caso discreto e, per l'Osservazione 5.18, tali quantità dipendono solo dalla distribuzione di X . Ad esempio,

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

purché $E(X^2) < \infty$. Come nel caso discreto, si indica con $L^p = L^p(\Omega, \mathcal{A}, P)$ l'insieme delle variabili aleatorie reali definite su (Ω, \mathcal{A}, P) tali che $E(|X|^p) < \infty$. Per la Proposizione 3.60, si ha $L^q \subseteq L^p$ se $q \geq p$ e, inoltre, $XY \in L^1$ se X e Y sono in L^2 .

Analogamente si definisce la covarianza di due variabili aleatorie reali X e Y definite sullo stesso spazio di probabilità:

$$\text{Cov}(X, Y) := E((X - E(X))(Y - E(Y))) = E(XY) - E(X)E(Y),$$

che dipende solo dalla distribuzione del vettore (X, Y) . Si ricordino le proprietà di covarianza e varianza, descritte dalle Proposizioni 3.68 e 3.69, e in particolare

$$\begin{aligned} \text{Var}(aX + b) &= a^2 \text{Var}(X), \\ \text{Var}(X + Y) &= \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y), \end{aligned} \tag{5.10}$$

per ogni $X, Y \in L^2$ e $a, b \in \mathbb{R}$.

Ricordiamo che variabili aleatorie X e Y indipendenti e in L^1 sono scorrelate: $\text{Cov}(X, Y) = 0$, per il Corollario 3.74. In particolare, se X e Y sono indipendenti e in L^2 , segue da (5.10) che $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$.

Concludiamo osservando che tutte le disuguaglianze del Paragrafo 3.3.5, così come i risultati del Paragrafo 3.3.6 sul coefficiente di correlazione, continuano a valere nel caso generale, senza alcuna modifica nelle dimostrazioni.

5.4 Costruzione di modelli probabilistici

Riassumendo per sommi capi il contenuto di questo capitolo, su uno spazio campionario Ω più che numerabile una probabilità P soddisfa gli stessi assiomi di uno spazio di probabilità discreto: la differenza è che P non è necessariamente definita su tutti i sottoinsiemi di Ω ma solo su una sottoclasse (una σ -algebra) \mathcal{A} . Nel seguito, lo spazio di probabilità (Ω, \mathcal{A}, P) resterà per lo più “sullo sfondo” e lavoreremo direttamente con le variabili aleatorie definite su Ω .

Una variabile aleatoria reale X , definita su uno spazio di probabilità generale, può assumere una quantità più che numerabile di valori di \mathbb{R} e, dunque, non essere discreta. La distribuzione di X è sempre identificata dalla funzione di ripartizione $F_X(x) := P(X \leq x)$, la cui conoscenza permette in linea di principio di calcolare tutte le quantità di interesse relative a X (vedremo diversi esempi nel prossimo capitolo). Le nozioni di indipendenza e valor medio possono essere estese a variabili aleatorie generali e valgono proprietà analoghe a quelle viste nel caso discreto.

Concludiamo il capitolo enunciando alcuni risultati che garantiscono l'esistenza dei modelli probabilistici che incontreremo nel seguito.

Osserviamo che *ogni probabilità μ su \mathbb{R} è la distribuzione di una variabile aleatoria*: infatti è sufficiente considerare lo spazio di probabilità $(\Omega, \mathcal{A}, P) := (\mathbb{R}, \mathcal{B}(\mathbb{R}), \mu)$ e si verifica immediatamente che l'applicazione $X(\omega) := \omega$, ossia l'identità, è una variabile aleatoria con distribuzione $\mu_X = \mu$. Questo procedimento è detto *costruzione canonica di una variabile aleatoria con distribuzione assegnata*, in analogia con il caso di variabili aleatorie discrete, descritto nel Paragrafo 3.1.6.

La nozione di funzione di ripartizione è associata in modo naturale a ogni probabilità μ su \mathbb{R} , definendo

$$F_\mu(x) := \mu((-\infty, x]), \quad \forall x \in \mathbb{R},$$

e valgono le proprietà (5.4). Il fatto interessante è che tali proprietà *caratterizzano* le funzioni di ripartizione, come mostra il seguente (profondo) risultato.

Teorema 5.19. *Per ogni funzione $F : \mathbb{R} \rightarrow [0, 1]$ con le proprietà (5.4), esiste una probabilità μ su \mathbb{R} tale che $F_\mu = F$ (e dunque, per la costruzione canonica, esiste una variabile aleatoria reale X tale che $F_X = F$).*

Esempio 5.20. Si consideri la funzione $F : \mathbb{R} \rightarrow [0, 1]$ definita da

$$F(x) := \begin{cases} 0 & \text{se } x < 0 \\ x & \text{se } 0 \leq x \leq 1 \\ 1 & \text{se } x > 1 \end{cases},$$

che soddisfa le proprietà (5.4). Per il Teorema 5.19, esiste una probabilità λ su \mathbb{R} tale che $\lambda((-\infty, x]) = F(x)$. Segue che $\lambda((a, b]) = b - a$ per ogni intervallo $(a, b] \subseteq [0, 1]$,

e questa relazione si estende a $\lambda([a, b]) = b - a$. Questa è dunque la probabilità cercata nell'Esempio 5.1, detta *misura di Lebesgue* (ristretta a $[0, 1]$). \square

Una variabile aleatoria reale Y , la cui distribuzione è la probabilità λ dell'Esempio 5.20, è detta *uniforme continua* in $[0, 1]$. Ne studieremo le proprietà nel prossimo capitolo. È interessante notare che, *assumendo l'esistenza di tale variabile aleatoria Y* , è possibile “completare” la dimostrazione del Teorema 5.19, costruendo una variabile aleatoria con funzione di ripartizione F generica mediante un'opportuna funzione $X = h(Y)$, come mostreremo nella Proposizione 6.27.

Enunciamo infine il seguente risultato.

Teorema 5.21. *Per ogni $k \in \mathbb{N}$ sia assegnata una probabilità μ_k su \mathbb{R}^d . Esiste uno spazio di probabilità $(\Omega, \mathcal{A}, \mathbb{P})$ su cui sono definite variabili aleatorie reali indipendenti $(X_k)_{k \in \mathbb{N}}$, tali che X_k ha distribuzione μ_k .*

Ricordiamo che, nel caso speciale $d = 1$, assegnare una probabilità μ_k su \mathbb{R} equivale ad assegnare una funzione $F_k : \mathbb{R} \rightarrow [0, 1]$ con le proprietà (5.4), per il Teorema 5.19.

Il Teorema 5.21 garantisce, in particolare, l'esistenza di una successione di variabili aleatorie reali *indipendenti e identicamente distribuite* (i.i.d.), ossia con la stessa distribuzione, qualunque essa sia. Questa classe di modelli probabilistici di fondamentale importanza ricorrerà frequentemente nel seguito. In particolare, abbiamo l'esistenza di una successione di variabili i.i.d. di Bernoulli, che descrivono una successione *infinita* di prove ripetute e indipendenti (si ricordi il Paragrafo 3.2.3).

Nel Teorema 5.21, si può scegliere $\Omega = (\mathbb{R}^d)^{\mathbb{N}} = \{\omega = (\omega_i)_{i \in \mathbb{N}} : \omega_i \in \mathbb{R}^d\}$, analogamente all'Esempio 5.2, con $X_i(\omega) := \omega_i$. Questa costruzione è una generalizzazione, altamente non banale, di quanto visto nel Paragrafo 3.2.5 per un numero finito di variabili aleatorie discrete indipendenti.

5.5 Note bibliografiche

In questo capitolo abbiamo presentato una sintesi, largamente incompleta, della formulazione assiomatica del calcolo delle probabilità, dovuta a A.N. Kolmogorov (1933) in [32], che utilizzò strumenti di *teoria della misura* sviluppati negli anni precedenti da E. Borel e H. Lebesgue. Per uno studio sistematico di questi argomenti, suggeriamo la lettura di monografie ad essi dedicate: due “classici” sono il testo di G.B. Folland [25] e quello di P. Billingsley [6].

Variabili aleatorie assolutamente continue

Sommario. In questo capitolo, dopo alcuni richiami sull'integrale di Riemann, introduciamo le variabili aleatorie assolutamente continue, reali e vettoriali, e ne studiamo le proprietà salienti. Presentiamo quindi le distribuzioni notevoli più importanti e analizziamo diversi esempi rilevanti, tra cui il processo di Poisson. I paragrafi e gli esercizi segnalati con un asterisco * richiedono la conoscenza dell'integrale di Riemann multidimensionale.

6.1 Richiami sull'integrale di Riemann

In quanto segue daremo per nota la nozione di *integrale di Riemann* per funzioni reali di una variabile reale, limitandoci a richiamare alcuni fatti fondamentali e alcune estensioni meno standard.

6.1.1 L'integrale in senso proprio

L'integrale di Riemann in senso “proprio”

$$\int_a^b f(x) dx,$$

è definito per una classe di funzioni $f : [a, b] \rightarrow \mathbb{R}$ *limitate*, definite su un intervallo *limitato*, dette Riemann-integrabili. Nel seguito scriveremo spesso, semplicemente, integrabili. Tale classe include tutte le funzioni monotone (crescenti o decrescenti) e quelle continue. Più in generale, ogni funzione $f : [a, b] \rightarrow \mathbb{R}$ *limitata*, e continua tranne al più in un insieme finito di punti, è integrabile.

L'integrale è additivo rispetto al dominio di integrazione: per ogni funzione $f : [a, b] \rightarrow \mathbb{R}$ integrabile e per ogni punto $c \in [a, b]$ si ha

$$\int_a^b f(x) dx = \int_a^c f(x) dx + \int_c^b f(x) dx. \quad (6.1)$$

Ricordiamo anche la *linearità* dell'integrale: se $f, g : [a, b] \rightarrow \mathbb{R}$ sono integrabili e $\alpha, \beta \in \mathbb{R}$, la funzione $\alpha f + \beta g$ è integrabile e si ha

$$\int_a^b (\alpha f(x) + \beta g(x)) dx = \alpha \int_a^b f(x) dx + \beta \int_a^b g(x) dx. \quad (6.2)$$

Un'altra proprietà fondamentale è la *monotonia*: se $f, g : [a, b] \rightarrow \mathbb{R}$ sono integrabili,

$$f(x) \leq g(x) \quad \forall x \in [a, b] \quad \implies \quad \int_a^b f(x) dx \leq \int_a^b g(x) dx. \quad (6.3)$$

Se $f : [a, b] \rightarrow \mathbb{R}$ è integrabile, per ogni $x \in [a, b]$ possiamo definire

$$F(x) := \int_a^x f(t) dt.$$

La funzione $F : [a, b] \rightarrow \mathbb{R}$ è detta *funzione integrale* di f . Per il *teorema fondamentale del calcolo integrale*, F è una funzione *continua* in tutto l'intervallo $[a, b]$ ed è derivabile in ogni punto $x \in (a, b)$ in cui f è continua, con $F'(x) = f(x)$.

Un corollario importante, molto utile per il calcolo di integrali, è il seguente: se $f : [a, b] \rightarrow \mathbb{R}$ è una funzione continua e $G : [a, b] \rightarrow \mathbb{R}$ è una *primitiva* di f , ossia G è continua in $[a, b]$ e derivabile per ogni $x \in (a, b)$, con $G'(x) = f(x)$, si ha

$$\int_a^x f(t) dt = G(t) \Big|_a^x := G(x) - G(a), \quad \text{per ogni } x \in [a, b]. \quad (6.4)$$

In altri termini, la funzione G coincide con la funzione integrale di f a meno di una costante additiva (infatti le due funzioni hanno la stessa derivata).

6.1.2 L'integrale in senso improprio

Risulta importante estendere la nozione di integrale di Riemann a funzioni che possono essere non definite in punti isolati (tipicamente si tratta di funzioni illimitate nell'intorno di tali punti) e/o che sono definite su un dominio illimitato.

Cominciamo a considerare domini limitati: supponiamo che $-\infty < a < b < +\infty$ ed estendiamo l'integrale ad una classe di funzioni che possono non essere definite in un singolo punto dell'intervallo $[a, b]$.

- Se $f : (a, b] \rightarrow \mathbb{R}$ è una funzione Riemann-integrabile su $[a + \varepsilon, b]$, per ogni $\varepsilon > 0$, e se esiste finito il limite

$$\lim_{\varepsilon \downarrow 0} \int_{a+\varepsilon}^b f(x) dx, \quad (6.5)$$

f si dice *integrabile in senso improprio* (o generalizzato) su $[a, b]$ e si pone

$$\int_a^b f(x) dx := \lim_{\varepsilon \downarrow 0} \int_{a+\varepsilon}^b f(x) dx.$$

In modo del tutto analogo si definisce, quando esiste, l'integrale improprio su $[a, b]$ di una funzione $f : [a, b] \rightarrow \mathbb{R}$.

- Se c è un punto interno dell'intervallo $[a, b]$, una funzione $f : [a, b] \setminus \{c\} \rightarrow \mathbb{R}$ si dice integrabile in senso improprio su $[a, b]$ se essa lo è sia su $[a, c]$ che su $[c, b]$, nel senso appena visto, e in tal caso si pone

$$\int_a^b f(x) dx := \int_a^c f(x) dx + \int_c^b f(x) dx.$$

Estendiamo quindi la definizione di integrale in senso improprio su $[a, b]$ a una classe di funzioni $f : [a, b] \setminus N \rightarrow \mathbb{R}$, dove $N = \{c_1, \dots, c_n\}$ è un insieme finito di punti.

- Suddividendo l'intervallo $[a, b]$ in n sottointervalli $[t_0, t_1], [t_1, t_2], \dots, [t_{n-1}, t_n]$, con $t_0 := a$ e $t_n := b$, che contengano ciascuno un solo punto di N (ad esempio, ponendo $t_j := \frac{1}{2}(c_j + c_{j+1})$ per $1 \leq j \leq n-1$), si dice che f è integrabile in senso improprio su $[a, b]$ se lo è su ciascun sottointervallo, nel qual caso si pone

$$\int_a^b f(x) dx := \int_{t_0}^{t_1} f(x) dx + \dots + \int_{t_{n-1}}^{t_n} f(x) dx.$$

Consideriamo ora il caso di domini illimitati. Si dice che un sottoinsieme $N \subseteq \mathbb{R}$ è *formato da punti isolati* (o, equivalentemente, *non ha punti di accumulazione*) se N ha un numero finito di punti in ogni intervallo limitato, ossia se $N \cap [a, b]$ è un insieme finito per ogni $-\infty < a < b < +\infty$.

- Se $f : (-\infty, b] \setminus N \rightarrow \mathbb{R}$, dove $N \subseteq \mathbb{R}$ è un insieme di punti isolati, è integrabile in senso improprio su $[c, b]$, per ogni $c < b$, e se esiste finito il limite

$$\lim_{c \rightarrow -\infty} \int_c^b f(x) dx,$$

si dice che f è integrabile in senso improprio su $(-\infty, b]$ e si pone

$$\int_{-\infty}^b f(x) dx := \lim_{c \rightarrow -\infty} \int_c^b f(x) dx.$$

In modo analogo si definisce, quando esiste, l'integrale in senso improprio su $[a, +\infty)$ per una classe di funzioni $f : [a, +\infty) \setminus N \rightarrow \mathbb{R}$.

- Infine, se $f : \mathbb{R} \setminus N \rightarrow \mathbb{R}$, con $N \subseteq \mathbb{R}$ insieme di punti isolati, è una funzione integrabile in senso improprio sia su $(-\infty, 0]$ che su $[0, +\infty)$, si dice che f è integrabile in senso improprio su \mathbb{R} e si pone

$$\int_{-\infty}^{+\infty} f(x) dx := \int_{-\infty}^0 f(x) dx + \int_0^{+\infty} f(x) dx.$$

(Naturalmente si può considerare un punto $c \in \mathbb{R}$ arbitrario invece di 0.)

Osservazione 6.1. Sottolineiamo che le funzioni integrabili in senso proprio su un intervallo limitato $[a, b]$ lo sono anche in senso improprio, perché l'integrale è una funzione continua degli estremi di integrazione. \square

Abbiamo dunque definito l'integrale improprio su \mathbb{R} (o su un intervallo) per funzioni che possono essere non definite in un sottoinsieme $N \subseteq \mathbb{R}$ di punti isolati. Per alleggerire le notazioni, d'ora in avanti converremo di definire arbitrariamente i valori della funzione sull'insieme N : ad esempio, parleremo di funzioni $f: \mathbb{R} \rightarrow \mathbb{R}$ integrabili in senso improprio, intendendo che esiste un sottoinsieme di punti isolati $N \subseteq \mathbb{R}$ tale che $f: \mathbb{R} \setminus N \rightarrow \mathbb{R}$ è integrabile in senso improprio. (Si noti che modificare il valore di f in un insieme di punti isolati non cambia l'integrale.)

6.1.3 Alcuni esempi

Esempio 6.2. La funzione $f(x) := \frac{1}{x^\alpha}$, definita in $(0, \infty)$, è integrabile in senso improprio su $[0, 1]$ se $\alpha < 1$, mentre non lo è se $\alpha \geq 1$. Infatti, per $\alpha \neq 1$ la funzione $G(x) := \frac{1}{1-\alpha} x^{1-\alpha}$ è una primitiva di f , quindi

$$\int_{\varepsilon}^1 f(x) dx = G(1) - G(\varepsilon) = \frac{1}{1-\alpha} (1 - \varepsilon^{1-\alpha}) \xrightarrow{\varepsilon \downarrow 0} \begin{cases} \frac{1}{1-\alpha} < +\infty & \text{se } \alpha < 1 \\ +\infty & \text{se } \alpha > 1 \end{cases}.$$

Analogamente, per $\alpha = 1$ la funzione $G(x) := \log x$ è una primitiva di f , quindi

$$\int_{\varepsilon}^1 f(x) dx = G(1) - G(\varepsilon) = -\log \varepsilon \xrightarrow{\varepsilon \downarrow 0} +\infty.$$

Gli stessi risultati valgono rimpiazzando l'intervallo $[0, 1]$ con $[0, a]$, per $a > 0$. \square

Esempio 6.3. La funzione $f(x) = \frac{1}{x^\alpha}$ è integrabile in senso improprio su $[1, +\infty)$ se $\alpha > 1$, mentre non lo è se $\alpha \leq 1$. Come nell'Esempio 6.2, per $\alpha \neq 1$ si ha

$$\int_1^c f(x) dx = \frac{1}{1-\alpha} (c^{1-\alpha} - 1) \xrightarrow{c \uparrow +\infty} \begin{cases} \frac{1}{\alpha-1} < +\infty & \text{se } \alpha > 1 \\ +\infty & \text{se } \alpha < 1 \end{cases}.$$

Analogamente, per $\alpha = 1$

$$\int_1^c f(x) dx = \log c \xrightarrow{c \uparrow +\infty} +\infty.$$

Gli stessi risultati valgono per la semiretta $[a, \infty)$ invece di $[1, \infty)$, con $a > 0$. \square

Sottolineiamo la “simmetria” tra gli Esempi 6.2 e 6.3: la funzione $f(x) := \frac{1}{x^\alpha}$ è integrabile in senso improprio su $[1, \infty)$ se e solo se $\alpha > 1$, mentre lo è su $[0, 1]$ se e solo se $\alpha < 1$. Nel caso “limite” $\alpha = 1$, la funzione $\frac{1}{x}$ non è integrabile in senso improprio né su $[1, \infty)$, né su $[0, 1]$.

Si noti inoltre l'analogia tra l'integrabilità in senso improprio della funzione $\frac{1}{x^\alpha}$ su $[1, \infty)$ e la convergenza/divergenza della serie $\sum_{n \in \mathbb{N}} \frac{1}{n^\alpha}$, cf. la relazione (0.2).

Esempio 6.4. Per ogni $\varepsilon > 0$ si ha

$$\int_{-1}^{-\varepsilon} \frac{1}{x} dx + \int_{\varepsilon}^1 \frac{1}{x} dx = 0,$$

per simmetria. Prendendo il limite per $\varepsilon \rightarrow 0$, si potrebbe essere tentati di concludere che $\int_{-1}^{+1} \frac{1}{x} dx = 0$, commettendo un errore: infatti la funzione $\frac{1}{x}$ *non* è integrabile in senso improprio su $[-1, +1]$, perché non lo è su $[0, 1]$, per quanto visto nell'Esempio precedente (né lo è su $[-1, 0]$, per simmetria). \square

Esempio 6.5. La funzione $f(x) := e^{\lambda x}$ è integrabile in senso improprio su $[a, \infty)$, con $a \in \mathbb{R}$, se e solo se $\lambda < 0$; simmetricamente, lo è su $(-\infty, a]$ se e solo se $\lambda > 0$. Per mostrarlo, notiamo che la funzione $G(x) := \lambda^{-1} e^{\lambda x}$ è una primitiva di f per $\lambda \neq 0$, mentre $G(x) := x$ lo è per $\lambda = 0$, pertanto

$$\int_a^c f(x) dx = G(c) - G(a) \xrightarrow{c \uparrow +\infty} \begin{cases} \frac{1}{(-\lambda)} e^{\lambda a} < +\infty & \text{se } \lambda < 0 \\ +\infty & \text{se } \lambda \geq 0 \end{cases}.$$

Il caso in cui $c \downarrow -\infty$ è del tutto analogo. \square

In tutti gli esempi precedenti, per mostrare che una funzione f è integrabile in senso improprio, abbiamo calcolato esplicitamente l'integrale di f , trovandone una primitiva e applicando il teorema fondamentale del calcolo integrale, mediante la relazione (6.4). Questo metodo non è di vasta applicabilità, perché le funzioni di cui è nota una primitiva sono relativamente poche.

Un metodo generale consiste nel dedurre l'integrabilità di una funzione a partire da un'altra funzione di cui è nota l'integrabilità, sfruttando la monotonia dell'integrale, espressa dalla relazione (6.3). A tal fine, gli Esempi 6.2, 6.3 e 6.5 si rivelano molto utili come "termini di confronto". Un esempio chiarirà la tecnica.

Esempio 6.6 (Integrale di Gauss). Mostriamo che la funzione $f(x) := e^{-x^2}$ è integrabile in senso improprio su \mathbb{R} . Dato che f è continua, essa è integrabile in senso proprio su ogni intervallo chiuso e limitato $[a, b]$. Quindi basta mostrare che

$$\lim_{a \downarrow -\infty} \int_a^0 e^{-x^2} dx < \infty, \quad \lim_{b \uparrow +\infty} \int_0^b e^{-x^2} dx < \infty. \quad (6.6)$$

Dato che $\int_{-b}^0 e^{-x^2} dx = \int_0^b e^{-x^2} dx$, per simmetria, è sufficiente concentrarsi sul secondo limite. Sottolineiamo che il limite esiste, perché l'integrale è una funzione crescente dell'estremo di integrazione b , essendo la funzione integranda positiva. Quindi basta mostrare che l'integrale $\int_0^b e^{-x^2} dx$ è limitato superiormente.

Per $x \geq 1$ si ha $x^2 \geq x$, quindi $e^{-x^2} \leq e^{-x}$. Per monotonia dell'integrale,

$$\forall b \geq 1 : \quad \int_1^b e^{-x^2} dx \leq \int_1^b e^{-x} dx \leq \int_1^{+\infty} e^{-x} dx = e^{-1},$$

avendo applicato l'Esempio 6.5. Dato che $\int_0^b(\dots) = \int_0^1(\dots) + \int_1^b(\dots)$, concludiamo che la funzione $b \mapsto \int_0^b e^{-x^2} dx$ è limitata dall'alto, come richiesto.

Il valore dell'integrale di f su \mathbb{R} , detto *integrale di Gauss*, è noto:

$$\int_{-\infty}^{+\infty} e^{-x^2} dx = \sqrt{\pi}, \quad (6.7)$$

come mostreremo più avanti, nell'Esempio 6.36. □

6.1.4 Approfondimenti sull'integrabilità

L'integrale in senso improprio su \mathbb{R} (o su un intervallo) è stato definito per un sottoinsieme di una classe speciale di funzioni, che potremmo definire “Riemann-integrabili in senso proprio al di fuori di un insieme di punti isolati”. Più precisamente, affinché una funzione $f: \mathbb{R} \rightarrow \mathbb{R}$ sia integrabile in senso improprio su \mathbb{R} , è necessario (ma non sufficiente!) che essa soddisfi la seguente proprietà:

esiste un sottoinsieme $N \subseteq \mathbb{R}$ di punti isolati tale che f è Riemann-integrabile in senso proprio su ogni intervallo limitato $[a, b]$ che non contiene punti di N . (6.8)

Non c'è una notazione standard per le funzioni con questa proprietà, e non la introdurremo. Sottolineiamo che si tratta di una classe piuttosto vasta, che include tutti gli esempi che si incontrano tipicamente nelle applicazioni ed è chiusa per le operazioni abituali (se f, g hanno questa proprietà, anche $f \pm g, f \cdot g, f^+, f^-, |f|$ la hanno). Suggeriamo di tenere a mente che sono in questa classe, in particolare, *tutte le funzioni* $f: \mathbb{R} \rightarrow \mathbb{R}$ continue al di fuori di un insieme di punti isolati N .

Supponiamo d'ora in avanti che $f: \mathbb{R} \rightarrow \mathbb{R}$ sia una funzione che soddisfi la proprietà (6.8). Riassumendo a grandi linee il contenuto del Paragrafo 6.1.2, f è integrabile in senso improprio su \mathbb{R} se e solo se vale quanto segue.

- Per ogni punto $x_0 \in N$ e per ogni intervallo limitato (a, b) che contiene x_0 (e nessun altro punto di N), l'integrale di f su $[a, x_0 - \varepsilon]$ ha *limite finito* per $\varepsilon \downarrow 0$; lo stesso per l'integrale di f su $[x_0 + \varepsilon, b]$. Questo permette di definire l'integrale di f in senso improprio su ogni intervallo limitato $[a, b]$.
- L'integrale di f in senso improprio sull'intervallo limitato $[a, b]$, con b fissato, ha *limite finito* per $a \downarrow -\infty$; lo stesso per $b \uparrow +\infty$, con a fissato. Questo permette di definire l'integrale di f in senso improprio su \mathbb{R} .

La richiesta che tutti i limiti siano finiti è al fine di evitare forme di indeterminazione del tipo $+\infty - \infty$, quando si sommano gli integrali su intervalli disgiunti.

Supponiamo ora che la funzione f sia *positiva*. In questo caso l'integrale di f su un intervallo è un numero positivo, che cresce se si ingrandisce l'intervallo, quindi i

limiti di cui sopra esistono sempre in $[0, +\infty]$. Di conseguenza, *per funzioni positive (con la proprietà (6.8)) si può sempre definire l'integrale in senso improprio su \mathbb{R} , ammettendo che possa assumere il valore $+\infty$.*

Per funzioni f non necessariamente positive, l'integrale in senso improprio su \mathbb{R} delle funzioni positive f^+ , f^- , $|f|$ è sempre ben definito in $[0, +\infty]$. *Quando l'integrale di $|f|$ è finito* – equivalentemente, quando sono finiti entrambi gli integrali di f^+ e f^- – si mostra facilmente che f è integrabile in senso improprio su \mathbb{R} e si ha

$$\int_{-\infty}^{+\infty} f(x) dx = \int_{-\infty}^{+\infty} f^+(x) dx - \int_{-\infty}^{+\infty} f^-(x) dx.$$

(Si noti l'analogia con la definizione di somma infinita, nel capitolo introduttivo “Nozioni Preliminari”).

In conclusione, per funzioni $f: \mathbb{R} \rightarrow \mathbb{R}$ con la proprietà (6.8), conviene tenere a mente quanto segue:

- se f è positiva, è sempre ben definito l'integrale $\int_{-\infty}^{+\infty} f(x) dx \in [0, +\infty]$;
- se f è positiva e se $\int_{-\infty}^{+\infty} f(x) dx < +\infty$, diremo che f è integrabile in senso improprio su \mathbb{R} , mantenendo la notazione del Paragrafo 6.1.2;
- per funzioni f non necessariamente positive, se $|f|$ è integrabile in senso improprio su \mathbb{R} (ossia, se $\int_{-\infty}^{+\infty} |f(x)| dx < +\infty$), allora f è integrabile in senso improprio su \mathbb{R} , dunque $\int_{-\infty}^{+\infty} f(x) dx \in \mathbb{R}$ è ben definito.

Queste nozioni sono sufficienti per gli scopi di questo libro. Seguono alcuni approfondimenti per il lettore interessato.

Per funzioni $f: \mathbb{R} \rightarrow \mathbb{R}$ che soddisfano la proprietà (6.8), la condizione che $|f|$ sia integrabile in senso improprio su \mathbb{R} è *strettamente più forte* della condizione che f sia integrabile in senso improprio su \mathbb{R} . Ad esempio, la funzione $f: \mathbb{R} \rightarrow \mathbb{R}$ definita da

$$f(x) := \begin{cases} 0 & \text{se } x \leq 0 \\ \frac{(-1)^n}{n} & \text{se } n-1 < x \leq n, \text{ con } n \in \mathbb{N} \end{cases}$$

è integrabile in senso improprio su \mathbb{R} , perché $\sum_{n=1}^{\infty} \frac{(-1)^n}{n} = \lim_{N \rightarrow +\infty} \sum_{n=1}^N \frac{(-1)^n}{n} = -\log 2$, ma $|f|$ non lo è, perché $\sum_{n=1}^{\infty} \frac{1}{n} = +\infty$. Un altro esempio classico è fornito dalla funzione $f(x) := \frac{1}{x} \sin(x)$. Si possono costruire esempi nello stesso spirito anche su intervalli $[a, b]$ limitati.

Come vedremo nel seguito, la semplice condizione che f sia integrabile in senso improprio su \mathbb{R} è troppo debole per i nostri scopi: richiederemo pertanto che l'integrale di $|f|$ sia finito. Sottolineiamo che facciamo sempre l'ipotesi che f soddisfi la proprietà (6.8) (altrimenti non ha nemmeno senso chiedersi se f sia integrabile). Se non si vuole menzionare esplicitamente la proprietà (6.8), occorre richiedere che entrambe le funzioni f e $|f|$ siano integrabili in senso improprio su \mathbb{R} o, equivalentemente, che entrambe le funzioni f^+ e f^- lo siano.

Esempio 6.7. Segnaliamo che, se non si richiede che la funzione $f: \mathbb{R} \rightarrow \mathbb{R}$ soddisfi la proprietà (6.8), possono accadere “patologie”: ad esempio, la funzione $|f|$ può essere integrabile

senza che f lo sia. Un esempio è dato dalla funzione $f : \mathbb{R} \rightarrow \mathbb{R}$ definita da

$$f(x) := \begin{cases} 1 & \text{se } x \in [0, 1] \cap \mathbb{Q} \\ -1 & \text{se } x \in [0, 1] \setminus \mathbb{Q} \\ 0 & \text{se } x \notin [0, 1] \end{cases},$$

non è Riemann-integrabile, mentre $|f(x)| = \mathbb{1}_{[0,1]}(x)$ naturalmente lo è. \square

6.1.5 Proprietà dell'integrale

Molte proprietà dell'integrale di Riemann in senso proprio continuano a valere anche per l'integrale in senso improprio. Notiamo innanzitutto che le proprietà di additività rispetto al dominio di integrazione, di linearità e di monotonia, espresse dalle relazioni (6.1), (6.2) e (6.3), continuano a valere se le funzioni sono integrabili in senso improprio, come si verifica facilmente.

Anche il teorema fondamentale del calcolo integrale si può estendere a funzioni integrabili in senso improprio. Ne presentiamo una versione che sarà utile nel seguito. Cominciamo con un'utile definizione.

Definizione 6.8. Una funzione $g : \mathbb{R} \rightarrow \mathbb{R}$ si dice \mathcal{C}^1 a tratti se

- g è continua in ogni punto $x \in \mathbb{R}$;
- esiste un sottoinsieme di punti isolati $N \subseteq \mathbb{R}$ tale che g è derivabile in ogni punto $x \in \mathbb{R} \setminus N$;
- la derivata g' è una funzione continua in ogni punto $x \in \mathbb{R} \setminus N$.

(Sottolineiamo che non si richiede niente sulla derivata g' nei punti di N .)

Vale quindi il seguente risultato.

Teorema 6.9. (i) Sia $f : \mathbb{R} \rightarrow \mathbb{R}$ integrabile in senso improprio su $(-\infty, b]$, per ogni $b \in \mathbb{R}$. Allora la funzione $F : \mathbb{R} \rightarrow \mathbb{R}$ definita da

$$F(x) := \int_{-\infty}^x f(t) dt, \quad (6.9)$$

è continua su tutto \mathbb{R} . Inoltre, per ogni punto $x \in \mathbb{R}$ in cui f è continua, la funzione F è derivabile in x e si ha $F'(x) = f(x)$.

- (ii) Sia $F : \mathbb{R} \rightarrow \mathbb{R}$ una funzione \mathcal{C}^1 a tratti. Allora la funzione $f(x) := F'(x)$ (definita arbitrariamente nei punti in cui F non è derivabile) è integrabile in senso improprio su ogni intervallo limitato $[a, b]$ e si ha

$$F(b) - F(a) = \int_a^b f(t) dt.$$

Se inoltre $\lim_{a \rightarrow -\infty} F(a) = 0$, allora la funzione f è integrabile in senso improprio su $(-\infty, b]$, per ogni $b \in \mathbb{R}$, e vale la relazione (6.9).

Un'importante corollario del Teorema 6.9 è la formula di *integrazione per parti*. Se $[a, b] \subseteq \mathbb{R}$ è un intervallo chiuso e limitato e $f, g : [a, b] \rightarrow \mathbb{R}$ sono funzioni \mathcal{C}^1 a tratti, allora le funzioni $f'(x)g(x)$ e $f(x)g'(x)$ sono integrabili in senso improprio su $[a, b]$ e vale la formula:

$$\int_a^b f'(x)g(x) dx = f(x)g(x)\Big|_a^b - \int_a^b f(x)g'(x) dx, \quad (6.10)$$

dove si pone

$$f(x)g(x)\Big|_a^b := f(b)g(b) - f(a)g(a). \quad (6.11)$$

Questo segue immediatamente dal Teorema 6.9 (ii) con $F(x) := f(x)g(x)$.

Osservazione 6.10. Se $f, g : \mathbb{R} \rightarrow \mathbb{R}$ sono funzioni \mathcal{C}^1 a tratti, la formula (6.10) vale per ogni $a, b \in \mathbb{R}$ con $a \leq b$. Se il limite per $a \rightarrow -\infty$ e/o $b \rightarrow +\infty$ dei tre termini che appaiono nella relazione esiste finito[†], la formula (6.10) continua a valere con $a = -\infty$ e/o $b = +\infty$, per definizione di integrale improprio (estendendo la notazione (6.11) in modo naturale, ossia ponendo $f(\pm\infty) := \lim_{x \rightarrow \pm\infty} f(x)$). \square

Ricordiamo infine la formula di *cambio di variabili*. Siano $(a, b), (c, d) \subseteq \mathbb{R}$ intervalli aperti (limitati o illimitati) e sia $\psi : (c, d) \rightarrow (a, b)$ una funzione biunivoca e derivabile, con $\psi'(t) > 0$ per ogni $t \in (c, d)$ (oppure $\psi'(t) < 0$ per ogni $t \in (c, d)$). Una funzione $g : (a, b) \rightarrow \mathbb{R}$ è integrabile in senso improprio su $[a, b]$ se e solo se la funzione $g(\psi(t))|\psi'(t)|$ è integrabile in senso improprio su $[c, d]$, e in tal caso

$$\int_a^b g(x) dx = \int_c^d g(\psi(t))|\psi'(t)| dt. \quad (6.12)$$

6.2 Variabili aleatorie reali assolutamente continue

Le variabili aleatorie discrete descrivono quantità casuali che possono assumere un insieme finito o numerabile di valori. Ci sono però quantità che assumono un'infinità continua di valori, tipicamente un intervallo $(a, b) \subseteq \mathbb{R}$ (ad esempio lunghezze, tempi, ...). Per descrivere tali quantità casuali, introduciamo ora una famiglia importante di variabili aleatorie reali X , dette *assolutamente continue*, caratterizzate dal fatto che la loro distribuzione è identificata da una funzione $f_X : \mathbb{R} \rightarrow \mathbb{R}$, detta *densità*, che permette il calcolo di tutte le quantità di interesse.

Vedremo che molte proprietà delle variabili aleatorie discrete hanno un corrispettivo per le variabili aleatorie assolutamente continue, in cui si sostituisce la densità discreta p_X con la densità f_X e le somme con gli integrali. Occorre tuttavia un po' di cautela, per non trarre generalizzazioni azzardate.

La nozione di integrabilità a cui facciamo riferimento è quella secondo Riemann in senso improprio, descritta nel paragrafo precedente. Ricordiamo che le funzioni

[†] Basta in realtà che esista finito il limite di due qualunque dei tre termini: in questo caso, il limite del terzo termine esiste finito grazie alla relazione (6.10).

integrabili in senso proprio lo sono anche in senso improprio; d'ora in avanti, per alleggerire le notazioni, scriveremo semplicemente “integrabili”.

Lo spazio di probabilità (Ω, \mathcal{A}, P) su cui sono definite le variabili aleatorie che considereremo verrà di norma sottinteso.

6.2.1 Definizione e prime proprietà

Definizione 6.11 (Variabile aleatoria assolutamente continua). Una variabile aleatoria reale X si dice *assolutamente continua* se esiste una funzione positiva $f_X : \mathbb{R} \rightarrow [0, +\infty)$, integrabile su \mathbb{R} , tale che la funzione di ripartizione di X si può esprimere come

$$F_X(t) = P(X \leq t) = \int_{-\infty}^t f_X(x) dx, \quad \forall t \in \mathbb{R}. \quad (6.13)$$

Una tale funzione f_X viene detta *densità* di X .

Si veda la Figura 6.1 a pagina 254 per una rappresentazione grafica.

Prendendo il limite $t \rightarrow +\infty$ nella relazione (6.13) e ricordando la Proposizione 3.89, segue che la densità f_X deve soddisfare la seguente relazione:

$$\int_{-\infty}^{+\infty} f_X(x) dx = 1. \quad (6.14)$$

Si noti l'analogia con la proprietà $\sum_{x \in \mathbb{R}} p_X(x) = 1$, espressa dalla relazione (3.7), per la densità discreta p_X di una variabile aleatoria reale discreta X .

Osservazione 6.12. Analogamente a quanto visto nel Paragrafo 3.1.6 per variabili aleatorie discrete, è naturale chiedersi se ogni funzione positiva $f : \mathbb{R} \rightarrow [0, +\infty)$, integrabile su \mathbb{R} e tale che $\int_{-\infty}^{+\infty} f(x) dx = 1$, sia la densità di una variabile aleatoria assolutamente continua. *La risposta è affermativa.* Infatti, per una tale f , alla funzione $F(x) := \int_{-\infty}^x f(t) dt$ si può applicare il Teorema 5.19, che garantisce l'esistenza di una variabile aleatoria X per cui $F_X = F$. \square

Esempio 6.13. Consideriamo la funzione $f : \mathbb{R} \rightarrow \mathbb{R}$ definita da

$$f(x) := 2 \cdot \mathbb{1}_{(0, \frac{1}{2})}(x) = \begin{cases} 2 & \text{se } 0 < x < \frac{1}{2} \\ 0 & \text{altrimenti} \end{cases}.$$

Chiaramente f è integrabile su \mathbb{R} e il suo integrale vale

$$\int_{-\infty}^{+\infty} f(x) dx = \int_0^{\frac{1}{2}} 2 dx = 1.$$

Per quanto detto, esiste dunque una variabile aleatoria X assolutamente continua con densità $f_X(x) = f(x)$. Come vedremo nel seguito, una tale variabile aleatoria descrive un numero reale “scelto uniformemente” nell'intervallo $(0, \frac{1}{2})$. \square

È importante sottolineare che, se X è una variabile aleatoria assolutamente continua, la sua densità f_X non è identificata in modo unico. Infatti, se f_X è una densità di X , ogni funzione g per cui

$$\int_{-\infty}^t g(x) dx = \int_{-\infty}^t f_X(x) dx, \quad \forall t \in \mathbb{R}, \quad (6.15)$$

è una densità di X . Ad esempio, se g è ottenuta da f_X modificandone il valore in un insieme di punti isolati, allora (6.15) vale e dunque anche g è una densità di X .

Questa ambiguità nella nozione di densità di una variabile aleatoria assolutamente continua non costituisce di norma un problema. Come vedremo, in molti casi esiste una versione “canonica” della densità che è regolare, ad esempio continua. Spesso diremo, impropriamente, che una certa funzione f è la densità di X .

Se X è una variabile aleatoria assolutamente continua, la sua funzione di ripartizione $F_X(x) = \int_{-\infty}^x f_X(t) dt$ è una funzione *continua*, per il Teorema 6.9, pertanto si ha $F_X(x) = F_X(x^-)$ per ogni $x \in \mathbb{R}$. Dato che $F_X(x) - F_X(x^-) = P(X = x)$, per la relazione (5.7), segue che

$$P(X = x) = 0, \quad \forall x \in \mathbb{R}. \quad (6.16)$$

In altri termini, *la probabilità che una variabile aleatoria assolutamente continua X assuma un valore fissato x è nulla, qualunque sia $x \in \mathbb{R}$.*

La proprietà (6.16) può apparire paradossale, dal momento che $P(X \in \mathbb{R}) = 1$. La soluzione del paradosso è che *non vale la relazione* $P(X \in \mathbb{R}) = \sum_{x \in \mathbb{R}} P(X = x)$, perché l'insieme \mathbb{R} dei numeri reali ha cardinalità più che numerabile e la proprietà di σ -additività della probabilità vale soltanto per famiglie numerabili di eventi.

Osservazione 6.14. La relazione $\sum_{x \in \mathbb{R}} P(X = x) = 1$ vale invece se X è una variabile aleatoria reale discreta, nel senso generale della Definizione 5.15, come mostrato in (5.6). Di conseguenza, *una variabile aleatoria assolutamente continua non è discreta, né viceversa.* \square

Notiamo ora che per ogni $a, b \in [-\infty, +\infty]$, con $a \leq b$, si ha

$$P(X \in (a, b]) = P(X \leq b) - P(X \leq a) = F_X(b) - F_X(a) = \int_a^b f_X(t) dt.$$

Inoltre, grazie alla relazione (6.16),

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

In modo analogo, si mostra che

$$\begin{aligned} P(X \in [a, b]) &= P(X \in (a, b]) = P(X \in [a, b)) = P(X \in (a, b)) \\ &= F_X(b) - F_X(a) = \int_a^b f_X(x) dx. \end{aligned} \quad (6.17)$$

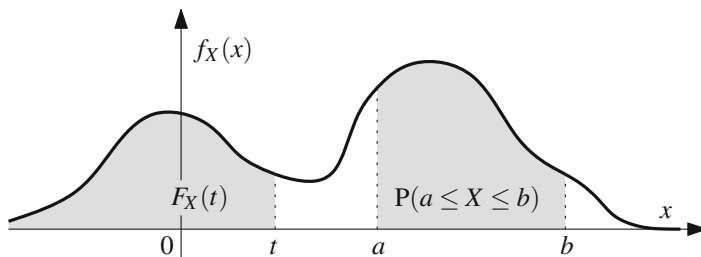


Figura 6.1 La funzione di ripartizione $F_X(t) = P(X \leq t)$ di una variabile aleatoria reale assolutamente continua X coincide con l'area sottesa dal grafico della densità f_X nella semiretta $(-\infty, t]$. Analogamente, la probabilità $P(a \leq X \leq b)$ è l'area sottesa dal grafico di f_X nell'intervallo $[a, b]$

Dunque, la probabilità che una variabile aleatoria assolutamente continua assuma valori in un intervallo non dipende dal fatto che gli estremi dell'intervallo siano inclusi o esclusi, contrariamente a quanto accade per le variabili aleatorie discrete. Tale probabilità può essere calcolata integrando la densità della variabile aleatoria sull'intervallo in questione (si veda la Figura 6.1).

In particolare, se $f_X(x) = 0$ per ogni $x \in (a, b)$, segue che $P(X \in (a, b)) = 0$, ossia gli intervalli in cui la densità è nulla costituiscono valori che (quasi certamente) non vengono assunti da X .

Osservazione 6.15. La proprietà (6.16) mostra, tra le altre cose, che la densità $f_X(x)$ di una variabile aleatoria assolutamente continua X non coincide con la probabilità che la variabile aleatoria assuma il valore x , ossia $f_X(x) \neq P(X = x)$.

Per dare un'interpretazione intuitiva alla densità f_X , sia x un punto in cui f_X è continua. Allora, per il Teorema 6.9 (i) e la relazione (6.17) si ha

$$f_X(x) = F'_X(x) = \lim_{\varepsilon \downarrow 0} \frac{F_X(x + \varepsilon) - F_X(x - \varepsilon)}{2\varepsilon} = \lim_{\varepsilon \downarrow 0} \frac{P(X \in [x - \varepsilon, x + \varepsilon])}{2\varepsilon}.$$

Questo mostra che $f_X(x)$ può essere interpretata come una *densità* di probabilità, ossia come il (limite del) rapporto tra la probabilità che X assuma valori in un piccolo intervallo centrato in x e l'ampiezza di tale intervallo.

Osserviamo infine che si può benissimo avere $f_X(x) > 1$ per qualche $x \in \mathbb{R}$, come abbiamo già visto nell'Esempio 6.13. Quello che conta è che l'integrale di f_X su ogni intervallo sia minore o uguale a 1, come segue da (6.17). \square

6.2.2 Determinare la densità

Un problema tipico consiste nel determinare se una variabile aleatoria X è assolutamente continua e, in caso affermativo, nel calcolarne la densità. In molti casi è possibile calcolare esplicitamente la funzione di ripartizione F_X , dunque si tratta di capire se essa si possa esprimere nella forma (6.13), per un'opportuna funzione f_X . A tal fine, risulta molto utile il seguente criterio (si ricordi la Definizione 6.8).

Proposizione 6.16. *Sia X una variabile aleatoria reale. Se la funzione di ripartizione F_X è \mathcal{C}^1 a tratti, allora X è assolutamente continua, con densità $f_X(x) := F'_X(x)$ (definita arbitrariamente dove F_X non è derivabile).*

Dimostrazione. È un corollario immediato del Teorema 6.9 (ii), dal momento che $\lim_{a \rightarrow -\infty} F_X(a) = 0$, come mostrato nella Proposizione 3.89. \square

Esempio 6.17. Scegliamo un numero casuale X uniformemente nell'intervallo $(0, \frac{1}{2})$. Detta Y l'area del quadrato di lato X , mostriamo che Y è una variabile aleatoria assolutamente continua e determiniamone la densità.

Ricordando l'Esempio 6.13, supponiamo che X sia una variabile aleatoria assolutamente continua con densità

$$f_X(x) = 2 \mathbb{1}_{(0, \frac{1}{2})}(x).$$

Calcoliamo la funzione di ripartizione di X . Si ha $F_X(x) = 0$ per $x \leq 0$ e $F_X(x) = 1$ per $x \geq \frac{1}{2}$ (perché?), mentre nel regime “interessante” $x \in (0, \frac{1}{2})$

$$F_X(x) = \int_{-\infty}^x f_X(t) dt = \int_0^x 2 dt = 2x.$$

In definitiva, abbiamo mostrato che

$$F_X(x) = \begin{cases} 0 & \text{se } x \leq 0 \\ 2x & \text{se } 0 < x < \frac{1}{2} \\ 1 & \text{se } x \geq \frac{1}{2} \end{cases}. \quad (6.18)$$

L'area del quadrato di lato X è data da $Y := X^2$. Determiniamo quindi la funzione di ripartizione $F_Y(y) = P(Y \leq y)$ di Y . Dato che Y è una variabile aleatoria positiva, per $y < 0$ si ha $\{Y \leq y\} = \emptyset$ e dunque $F_Y(y) = 0$. Per $y \geq 0$, notiamo che l'evento $\{Y \leq y\} = \{X^2 \leq y\}$ coincide con l'evento $\{-\sqrt{y} \leq X \leq \sqrt{y}\}$: infatti

$$\begin{aligned} \omega \in \{Y \leq y\} &\iff Y(\omega) \leq y \iff X(\omega)^2 \leq y \\ &\iff -\sqrt{y} \leq X(\omega) \leq \sqrt{y} \iff \omega \in \{-\sqrt{y} \leq X \leq \sqrt{y}\}. \end{aligned}$$

Quindi, grazie alla relazione (6.17), per $y \geq 0$ si ha

$$F_Y(y) = P(Y \leq y) = P(X \in [-\sqrt{y}, \sqrt{y}]) = F_X(\sqrt{y}) - F_X(-\sqrt{y}) = F_X(\sqrt{y}),$$

avendo usato il fatto che $F_X(x) = 0$ per $x \leq 0$. Dato che $0 \leq \sqrt{y} < \frac{1}{2}$ se e solo se $0 \leq y < \frac{1}{4}$, dalla relazione (6.18) otteniamo

$$F_Y(y) = \begin{cases} 0 & \text{se } y < 0 \\ 2\sqrt{y} & \text{se } 0 \leq y < \frac{1}{4} \\ 1 & \text{se } y \geq \frac{1}{4} \end{cases}.$$

La funzione F_Y è continua su tutto \mathbb{R} . Inoltre è derivabile, con derivata continua, negli intervalli $(-\infty, 0)$, $(0, \frac{1}{4})$ e $(\frac{1}{4}, +\infty)$. Quindi F_Y è \mathcal{C}^1 a tratti, secondo la Definizione 6.8, scegliendo come insieme $N = \{0, \frac{1}{4}\}$. Applicando la Proposizione 6.16, concludiamo che la variabile aleatoria Y è assolutamente continua, con densità

$$f_Y(y) = F'_Y(y) = \frac{1}{\sqrt{y}} \mathbb{1}_{(0, \frac{1}{4})}(y).$$

(Si può verificare che, in questo caso, la funzione F_Y non è derivabile nei punti di N , ossia per $y = 0$ e $y = \frac{1}{4}$, ma non è necessario farlo!)

Concludiamo osservando che $f_Y(y)$ diverge a $+\infty$ per $y \downarrow 0$, e, per l'Esempio 6.2, la divergenza è integrabile (ovviamente, dal momento che f_Y è una densità.) \square

Nell'esempio precedente, data una variabile aleatoria reale X assolutamente continua e una funzione $\varphi : \mathbb{R} \rightarrow \mathbb{R}$, abbiamo mostrato che la variabile aleatoria $Y = \varphi(X)$ è assolutamente continua, calcolandone la funzione di ripartizione F_Y e poi derivandola, grazie alla Proposizione 6.16. Questo metodo è generale e funziona con molte funzioni φ , ad esempio ogniqualvolta i valori di x per cui $\varphi(x) \leq y$ formano un'unione di intervalli i cui estremi dipendono in modo "regolare" da y . Questo è il caso della funzione $\varphi(x) = x^2$, ad esempio, come abbiamo visto (e come rivedremo nel Paragrafo 6.5.1). Il lettore potrà trovare altri esempi fra gli esercizi.

Per la sua importanza, formuliamo esplicitamente il caso di trasformazioni φ lineari-affini e invertibili. Il caso in cui φ è una funzione strettamente monotona e regolare è del tutto analogo ed è lasciato al lettore (si veda l'Esercizio 6.1).

Proposizione 6.18 (Trasformazioni lineari-affini). *Sia X una variabile aleatoria reale assolutamente continua, con densità f_X . Per ogni $a, b \in \mathbb{R}$ con $a \neq 0$, la variabile aleatoria $Y := aX + b$ è assolutamente continua e ha densità*

$$f_Y(y) = \frac{1}{|a|} f_X\left(\frac{y-b}{a}\right). \quad (6.19)$$

Dimostrazione. Supponiamo che $a > 0$ (il caso $a < 0$ è analogo). Per $y \in \mathbb{R}$ si ha

$$F_Y(y) = P(Y \leq y) = P(aX + b \leq y) = P\left(X \leq \frac{y-b}{a}\right) = F_X\left(\frac{y-b}{a}\right).$$

Intuitivamente, derivando questa espressione si ottiene la densità di Y :

$$f_Y(y) = F'_Y(y) = \frac{1}{a} F'_X\left(\frac{y-b}{a}\right) = \frac{1}{a} f_X\left(\frac{y-b}{a}\right).$$

Questa derivazione è giustificata se F_X è \mathcal{C}^1 a tratti, grazie alla Proposizione 6.16.

Lo stesso risultato si può ottenere in generale, applicando la formula di cambio di variabili (6.12): infatti, ponendo $x = \psi(t) := (t - b)/a$ si ottiene

$$F_Y(y) = F_X\left(\frac{y-b}{a}\right) = \int_{-\infty}^{\frac{y-b}{a}} f_X(x) dx = \int_{-\infty}^y f_X\left(\frac{t-b}{a}\right) \frac{1}{a} dt.$$

Ricordando la Definizione 6.11, questo mostra che Y è assolutamente continua e la funzione $y \mapsto f_X\left(\frac{y-b}{a}\right) \frac{1}{a}$ ne è una densità. \square

Osservazione 6.19. Anche se X è una variabile aleatoria assolutamente continua, la distribuzione di $Y := \varphi(X)$ può essere *del tutto arbitraria* se non si fanno opportune ipotesi sulla funzione φ (come mostreremo nella Proposizione 6.27). In particolare, non è detto che Y sia assolutamente continua. Ad esempio, se $\varphi(x) \equiv c$ è una funzione costante, $Y = \varphi(X) = c$ è una variabile aleatoria costante, dunque discreta, qualunque sia la variabile aleatoria X ! \square

6.2.3 Il calcolo del valor medio

Sia $g : \mathbb{R} \rightarrow \mathbb{R}$ una funzione misurabile. La densità f_X di una variabile aleatoria reale assolutamente continua X permette di determinare se la variabile aleatoria $g(X)$ ammette valor medio e, in caso affermativo, di calcolarlo, in analogia con il caso di variabili aleatorie discrete (si ricordi la formula (3.30)).

È necessario fare alcune ipotesi sulla funzione g . Per semplicità, enunceremo i risultati nel caso in cui g è una funzione continua tranne al più in un insieme $N \subseteq \mathbb{R}$ di punti isolati (ma tutto vale nel caso generale in cui g soddisfa la proprietà (6.8)). Si può mostrare che sotto queste ipotesi g è misurabile, dunque $g(X)$ è una variabile aleatoria. Inoltre, se g è positiva, l'integrale su \mathbb{R} della funzione $g(x)f_X(x)$ è sempre ben definito (eventualmente vale $+\infty$), per quanto visto nel Paragrafo 6.1.4.

Proposizione 6.20 (Calcolo del valor medio). *Sia X una variabile aleatoria reale assolutamente continua, con densità f_X , e sia $g : \mathbb{R} \rightarrow [0, +\infty)$ una funzione positiva e continua, tranne al più in un insieme di punti isolati. Il valor medio della variabile aleatoria positiva $g(X)$ è sempre ben definito in $[0, +\infty]$ ed è dato da*

$$\mathbb{E}[g(X)] = \int_{-\infty}^{+\infty} g(x)f_X(x) dx. \quad (6.20)$$

Nel caso di funzioni $g : \mathbb{R} \rightarrow \mathbb{R}$ non necessariamente positive, si può applicare la Proposizione 6.20 alle funzioni positive g^+ , g^- e $|g|$. Ricordando la Definizione 5.17, otteniamo i seguenti risultati.

Proposizione 6.21 (Calcolo del valor medio). *Sia X una variabile aleatoria reale assolutamente continua, con densità f_X , e sia $g : \mathbb{R} \rightarrow \mathbb{R}$ una funzione continua, tranne al più in un insieme di punti isolati.*

La variabile aleatoria $g(X)$ ammette valor medio finito se e solo se la funzione $|g(x)| f_X(x)$ è integrabile su \mathbb{R} :

$$E[|g(X)|] < +\infty \quad \Longleftrightarrow \quad \int_{-\infty}^{+\infty} |g(x)| f_X(x) dx < +\infty,$$

nel qual caso il valor medio $E[g(X)]$ è dato dalla formula (6.20).

Corollario 6.22. *Una variabile aleatoria reale X assolutamente continua, con densità f_X , ammette momento di ordine k finito se e solo se*

$$E(|X|^k) = \int_{-\infty}^{+\infty} |x|^k f_X(x) dx < +\infty,$$

nel qual caso il momento di ordine k di X è dato dall'integrale

$$E(X^k) = \int_{-\infty}^{+\infty} x^k f_X(x) dx. \quad (6.21)$$

La funzione generatrice dei momenti di X è data dall'integrale

$$M_X(t) = E[e^{tX}] = \int_{-\infty}^{+\infty} e^{tx} f_X(x) dx.$$

Vediamo ora qualche esempio di calcolo di valori medi. Ne vedremo altri nel Paragrafo 6.3, studiando le variabili aleatorie assolutamente continue notevoli.

Esempio 6.23. Sia Y una variabile aleatoria assolutamente continua, con densità

$$f_Y(y) = \frac{1}{\sqrt{y}} \mathbb{1}_{(0, \frac{1}{4})}(y),$$

come nell'Esempio 6.17. La variabile Y assume valori in $(0, \frac{1}{4})$, come si deduce dal fatto che la densità f_Y è nulla al di fuori di tale intervallo. Dato che Y è positiva, il valor medio $E(Y)$ è sempre ben definito e, grazie alla relazione (6.21),

$$E(Y) = \int_{-\infty}^{+\infty} y f_Y(y) dy = \int_0^{\frac{1}{4}} y \frac{1}{\sqrt{y}} dy = \frac{y^{3/2}}{3/2} \Big|_0^{\frac{1}{4}} = \frac{1}{12}.$$

perché $G(y) := \frac{y^{3/2}}{3/2}$ è una primitiva di \sqrt{y} .

In alternativa, ricordando dall'Esempio 6.17 che $Y = X^2$, dove X ha densità $f_X(x) = 2\mathbb{1}_{(0, \frac{1}{2})}(x)$, si può applicare la formula (6.20) alla funzione $g(x) := x^2$:

$$E(Y) = E(X^2) = \int_{-\infty}^{+\infty} x^2 f_X(x) dx = \int_0^{\frac{1}{2}} 2x^2 dx = \frac{2}{3} x^3 \Big|_0^{\frac{1}{2}} = \frac{1}{12},$$

in accordo col calcolo precedente. \square

Esempio 6.24. Sia X una variabile aleatoria reale con densità

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

Una tale variabile aleatoria è detta *di Cauchy*. Dato che $\arctan'(x) = \frac{1}{1+x^2}$, si ha

$$\int_{-\infty}^{+\infty} f_X(x) dx = \frac{1}{\pi} (\arctan(+\infty) - \arctan(-\infty)) = \frac{1}{\pi} \left(\frac{\pi}{2} + \frac{\pi}{2} \right) = 1,$$

dunque f_X è effettivamente una densità.

Mostriamo che X non ammette valor medio. Per la relazione (6.20),

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

dove l'ultima uguaglianza segue dall'Esempio 6.3, perché $\frac{x}{1+x^2} \sim \frac{1}{x}$ per $x \rightarrow +\infty$. Analogamente, per simmetria, $E(X^-) = +\infty$, dunque X non ammette valor medio.

Vale la pena di osservare che per ogni $a > 0$ si ha

$$\int_{-a}^{+a} x f_X(x) dx = \frac{1}{\pi} \int_{-a}^{+a} \frac{x}{1+x^2} dx = 0, \quad (6.22)$$

per simmetria: infatti la funzione $\frac{x}{1+x^2}$ è integrabile su $[-a, a]$ (è continua) e *dispari*. Prendendo il limite $a \uparrow +\infty$, si potrebbe essere tentati di dedurre da (6.22) che $E(X) = \int_{-\infty}^{+\infty} x f_X(x) dx = 0$, *commettendo un errore* (in analogia con l'Esempio 6.4). Infatti, come abbiamo mostrato sopra, il valor medio $E(X)$ non è definito.

Sottolineiamo che ci sono ragioni sostanziali per non attribuire a X valor medio nullo (o un altro valor medio). Per esempio, nel Capitolo 7 dimostreremo la legge dei grandi numeri, un importante teorema che si applica a variabili aleatorie reali con valor medio finito, e che non vale per variabili aleatorie di Cauchy. \square

6.2.4 Calcoli con variabili aleatorie indipendenti

Uno studio approfondito dell'indipendenza di variabili aleatorie reali assolutamente continue richiede l'uso di vettori aleatori assolutamente continui, che introdurremo più avanti. Vedremo in particolare, nella Proposizione 6.41, criteri espliciti per l'indipendenza analoghi al caso discreto.

Per il momento, ci limitiamo a presentare un risultato molto utile, analogo a quanto mostrato per variabili aleatorie discrete. La sua dimostrazione è rinviata alla Proposizione 6.46 (si veda anche l'Esempio 6.51).

Proposizione 6.25 (Convoluzione tra densità). *Siano X e Y variabili aleatorie reali assolutamente continue indipendenti, con densità f_X e f_Y . La variabile aleatoria $X + Y$ è assolutamente continua, con densità data dalla convoluzione di f_X e f_Y , definita da:*

$$f_{X+Y}(z) = (f_X * f_Y)(z) := \int_{-\infty}^{+\infty} f_X(x) f_Y(z-x) dx = \int_{-\infty}^{+\infty} f_X(z-x) f_Y(x) dx.$$

Concludiamo ricordando che la Proposizione 3.93, sul massimo e minimo di variabili aleatorie reali *indipendenti* X_1, \dots, X_n , è valida qualunque sia la distribuzione delle X_i (non necessariamente discreta o assolutamente continua): definendo

$$Z := \max(X_1, \dots, X_n), \quad W := \min(X_1, \dots, X_n), \quad (6.23)$$

si ha

$$F_Z(x) = \prod_{k=1}^n F_{X_k}(x), \quad F_W(x) = 1 - \prod_{k=1}^n [1 - F_{X_k}(x)]. \quad (6.24)$$

Nel caso in cui X_1, \dots, X_n siano assolutamente continue, queste espressioni permettono di mostrare che Z e W sono anch'esse assolutamente continue e di ricavarne la densità (si veda ad esempio l'Esercizio 6.2).

Esercizi

Esercizio 6.1. Sia X una variabile aleatoria reale assolutamente continua e sia (a, b) un intervallo aperto (limitato o illimitato) di \mathbb{R} , tale che $P(X \in (a, b)) = 1$. Sia $\varphi : (a, b) \rightarrow \mathbb{R}$ una funzione di classe \mathcal{C}^1 tale che $\varphi'(x) > 0$ (oppure < 0) per ogni $x \in (a, b)$. Allora l'immagine di φ è un intervallo aperto (c, d) . L'obiettivo di questo esercizio è di mostrare che la variabile aleatoria $Y := \varphi(X)$ è assolutamente continua, con densità

$$f_Y(y) = \frac{1}{|\varphi'(\varphi^{-1}(y))|} f_X(\varphi^{-1}(y)) \mathbb{1}_{(c,d)}(y), \quad (6.25)$$

dove $\varphi^{-1} : (c, d) \rightarrow (a, b)$ indica la funzione inversa di φ .

Si consideri solo il caso $\varphi'(x) > 0$ per ogni $x \in (a, b)$ (il caso $\varphi'(x) < 0$ è analogo).

(i) Si mostri che la funzione di ripartizione di Y è data da

$$F_Y(y) = \begin{cases} 0 & \text{se } y \leq c \\ F_X(\varphi^{-1}(y)) & \text{se } c < y < d \\ 1 & \text{se } y \geq d \end{cases}$$

- (ii) Assumendo che F_X sia \mathcal{C}^1 a tratti, si concluda che Y è assolutamente continua, con densità f_Y data da (6.25).
- (iii) Si giunga alla stessa conclusione senza assumere che F_X sia \mathcal{C}^1 a tratti, usando la formula di cambio di variabili (6.12).

Esercizio 6.2. Siano X_1, \dots, X_n variabili aleatorie reali indipendenti, con la stessa distribuzione. Indicando con $F = F_{X_i}$ la comune funzione di ripartizione, facciamo l'ipotesi che F sia \mathcal{C}^1 a tratti, così che $f(x) = F'(x)$ è la densità delle X_i .

Si mostri che $Z := \max(X_1, \dots, X_n)$ e $W := \min(X_1, \dots, X_n)$ sono variabili aleatorie assolutamente continue, con densità

$$f_Z(x) = n(F(x))^{n-1}f(x), \quad f_W(x) = n(1 - F(x))^{n-1}f(x). \quad (6.26)$$

6.3 Classi notevoli di variabili aleatorie reali assolutamente continue

Come abbiamo fatto per le variabili aleatorie discrete, in questo paragrafo presentiamo alcune delle distribuzioni di variabili aleatorie assolutamente continue che sono più rilevanti per le applicazioni.

6.3.1 Uniforme continua

Siano $a, b \in \mathbb{R}$ con $a < b$. Una variabile aleatoria reale assolutamente continua X si dice *uniforme su (a, b)* , e scriviamo $X \sim U(a, b)$, se ha densità

$$f_X(x) = f_{U(a,b)}(x) := \frac{1}{b-a} \mathbb{1}_{(a,b)}(x) \quad (6.27)$$

(si veda la Figura 6.2). Si verifica facilmente che questa funzione è effettivamente una densità, ossia si ha $\int_{-\infty}^{+\infty} f_{U(a,b)}(x) dx = 1$.

È naturale interpretare una variabile aleatoria $X \sim U(a, b)$ come “un punto scelto uniformemente nell'intervallo (a, b) ”, perché il valore della “densità di probabilità” $f_X(x)$ è lo stesso, strettamente positivo, per ogni $x \in (a, b)$, e nullo al di fuori di tale intervallo (si ricordi l'Osservazione 6.15). Più concretamente, per ogni $x \in (a, b)$ fissato, se $\varepsilon > 0$ è sufficientemente piccolo, la probabilità che X assuma valori nell'intorno $(x - \varepsilon, x + \varepsilon)$ vale $\frac{2\varepsilon}{b-a}$ e dunque non dipende da x .

Osservazione 6.26. Rimpiazzare l'intervallo aperto (a, b) con l'intervallo chiuso $[a, b]$ nella densità (6.27) significa modificare il valore di f_X in due punti. Come abbiamo visto, una tale modifica alla densità non cambia la funzione di ripartizione e dunque la distribuzione. Di conseguenza, una variabile aleatoria $X \sim U(a, b)$ è anche detta uniforme su $[a, b]$ (o, analogamente, su $[a, b)$ o su $(a, b]$). Questa “ambiguità” è consistente col fatto che $P(X = a) = P(X = b) = 0$ (si ricordi (6.16)). \square

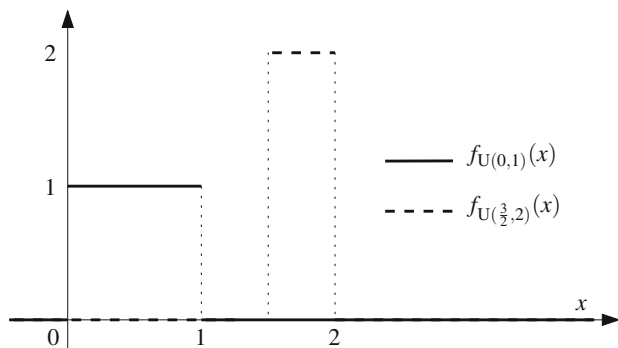


Figura 6.2 Densità di una variabile aleatoria $U(a, b)$, nei due casi $(a, b) = (0, 1)$ e $(a, b) = (\frac{3}{2}, 2)$

Calcoliamo ora media e varianza di $X \sim U(a, b)$. Essendo la densità f_X nulla al di fuori dell'intervallo (a, b) , si ha $P(X \in (a, b)) = 1$, ossia la variabile aleatoria X è quasi certamente limitata e dunque ammette momenti finiti di ogni ordine. Applicando la formula (6.21) si ottiene

$$E(X) = \frac{1}{b-a} \int_a^b x dx = \frac{a+b}{2}, \quad E(X^2) = \frac{1}{b-a} \int_a^b x^2 dx = \frac{a^2 + ab + b^2}{3},$$

da cui

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

La funzione generatrice dei momenti vale

$$M_X(t) = E(e^{tX}) = \frac{1}{b-a} \int_a^b e^{tx} dx = \frac{e^{tb} - e^{ta}}{t(b-a)}.$$

Anche la funzione di ripartizione è esplicita:

$$F_X(x) = \int_{-\infty}^x f_X(t) dt = \begin{cases} 0 & \text{se } x \leq a \\ \frac{x-a}{b-a} & \text{se } x \in (a, b) \\ 1 & \text{se } x \geq b \end{cases}. \quad (6.28)$$

Le variabili uniformi possono essere “simulate” al calcolatore da speciali programmi, chiamati *generatori di numeri casuali*, che sono alla base di numerosi algoritmi, detti *algoritmi stocastici*. Il fatto interessante è che una variabile aleatoria uniforme può essere usata come ingrediente per “costruire” esplicitamente variabili aleatorie reali con distribuzione arbitraria, come mostra il seguente risultato.

Proposizione 6.27. Sia $F : \mathbb{R} \rightarrow [0, 1]$ una funzione che soddisfa le proprietà (5.4). Sia $h : (0, 1) \rightarrow \mathbb{R}$ la funzione crescente definita da:

$$h(y) := \inf\{z \in \mathbb{R} : F(z) \geq y\}.$$

(Si noti che si ha $h = F^{-1}$ se F è biunivoca.) Se $Y \sim U(0, 1)$, la variabile aleatoria $h(Y)$ ha funzione di ripartizione data da F .

In particolare, se X è una generica variabile aleatoria reale, con funzione di ripartizione $F = F_X$, la variabile aleatoria $h(Y)$ ha la stessa distribuzione di X .

Dimostrazione. Fissiamo $y \in (0, 1)$ e $x \in \mathbb{R}$. Se $F(x) \geq y$, segue immediatamente dalla definizione di h che $h(y) \leq x$. Viceversa, se $F(x) < y$, essendo F una funzione continua da destra, esiste $\varepsilon > 0$ tale che $F(x + \varepsilon) < y$; per monotonia, $F(z) < y$ per ogni $z \leq x + \varepsilon$ e dunque $h(y) \geq x + \varepsilon$, in particolare $h(y) > x$, sempre per definizione di h . In definitiva, abbiamo mostrato che

$$h(y) \leq x \iff F(x) \geq y.$$

Sia dunque $Z := h(Y)$ con $Y \in U(0, 1)$. Si noti che Z è una variabile aleatoria dato che h , essendo crescente, è misurabile. Poiché $P(Y \in (0, 1)) = 1$, abbiamo che

$$F_Z(x) = P(h(Y) \leq x) = P(F(x) \geq Y) = F_Y(F(x)) = F(x),$$

avendo applicato nell'ultima uguaglianza la relazione (6.28) con $a = 0$, $b = 1$ e il fatto che $F(x) \in [0, 1]$. Dunque Z ha funzione di ripartizione F , come richiesto.

Infine, se X è una variabile aleatoria con funzione di ripartizione $F = F_X$, i passi precedenti mostrano che $F_Z = F_X$. Dal fatto che X e Z hanno la stessa funzione di ripartizione segue che esse hanno la stessa distribuzione. \square

6.3.2 Gamma

Cominciamo con alcuni richiami sulla funzione “Gamma di Eulero”. Si tratta della funzione $\Gamma : (0, \infty) \rightarrow \mathbb{R}$ definita da

$$\Gamma(\alpha) = \int_0^{+\infty} x^{\alpha-1} e^{-x} dx. \quad (6.29)$$

Si osservi che l'integrale è finito (ossia la funzione $x^{\alpha-1} e^{-x}$ è integrabile su $[0, +\infty)$) proprio per $\alpha > 0$, come segue dagli Esempi 6.2 e 6.5. Una proprietà fondamentale della funzione Γ è la seguente:

$$\Gamma(\alpha + 1) = \alpha \Gamma(\alpha), \quad \forall \alpha > 0. \quad (6.30)$$

Infatti, integrando per parti:

$$\Gamma(\alpha + 1) = \int_0^{+\infty} x^\alpha e^{-x} dx = \left(-x^\alpha e^{-x} \right) \Big|_0^{+\infty} + \alpha \int_0^{+\infty} x^{\alpha-1} e^{-x} dx = \alpha \Gamma(\alpha),$$

perché $x^\alpha e^{-x} \rightarrow 0$ per $x \rightarrow +\infty$.

Osservazione 6.28. La relazione (6.30) permette di determinare $\Gamma(\alpha + 1)$ se è noto $\Gamma(\alpha)$. Dato che

$$\Gamma(1) = \int_0^{+\infty} e^{-x} dx = (-e^{-x})|_0^{\infty} = 1,$$

si ottengono i valori della funzione Γ sui numeri naturali. Dato che il fattoriale soddisfa la relazione $n! = n(n-1)!$, e inoltre $0! = 1$, segue che

$$\Gamma(n) = (n-1)! \quad \forall n \in \mathbb{N}. \quad (6.31)$$

La funzione Γ fornisce dunque un'estensione del fattoriale ai numeri reali positivi.

Un altro valore notevole si ha per $\alpha = \frac{1}{2}$: con il cambio di variabile $x = y^2$,

$$\Gamma\left(\frac{1}{2}\right) = \int_0^{+\infty} x^{-1/2} e^{-x} dx = 2 \int_0^{+\infty} e^{-y^2} dy = \int_{-\infty}^{+\infty} e^{-y^2} dy = \sqrt{\pi}, \quad (6.32)$$

dove si è usato il valore dell'integrale di Gauss (6.7). Questo permette di ottenere i valori di $\Gamma(\alpha)$ quando α è semi-intero:

$$\Gamma\left(n + \frac{1}{2}\right) = \sqrt{\pi} \prod_{k=0}^{n-1} \left(k + \frac{1}{2}\right), \quad \forall n \in \mathbb{N}, \quad (6.33)$$

avendo usato ancora la relazione (6.30). \square

Se nella definizione (6.29) operiamo il cambio di variabili $x = \lambda t$, con $\lambda > 0$, otteniamo

$$\Gamma(\alpha) = \int_0^{+\infty} \lambda^\alpha t^{\alpha-1} e^{-\lambda t} dt.$$

Ciò garantisce la bontà della seguente definizione: una variabile aleatoria reale X è detta *Gamma di parametri* $\alpha, \lambda > 0$, e scriveremo $X \sim \text{Gamma}(\alpha, \lambda)$, se è assolutamente continua con densità

$$f_X(x) = f_{\text{Gamma}(\alpha, \lambda)}(x) := \frac{1}{\Gamma(\alpha)} \lambda^\alpha x^{\alpha-1} e^{-\lambda x} \mathbf{1}_{(0, +\infty)}(x), \quad (6.34)$$

si veda la Figura 6.3. Sottolineiamo che il secondo parametro di una $\text{Gamma}(\alpha, \lambda)$ è un semplice fattore di scala: più precisamente

$$X \sim \text{Gamma}(\alpha, \lambda) \quad \Longleftrightarrow \quad Y := \lambda X \sim \text{Gamma}(\alpha, 1), \quad (6.35)$$

come si verifica facilmente grazie alla Proposizione 6.18.

Media e varianza si calcolano facilmente: se $Y \sim \text{Gamma}(\alpha, 1)$, applicando la formula (6.21) e ricordando la relazione (6.30) si ottiene

$$E(Y) = \int_{-\infty}^{+\infty} y f_Y(y) dy = \frac{1}{\Gamma(\alpha)} \int_0^{+\infty} y^\alpha e^{-y} dy = \frac{\Gamma(\alpha+1)}{\Gamma(\alpha)} = \alpha,$$

e analogamente

$$E(Y^2) = \int_{-\infty}^{+\infty} y^2 f_Y(y) dy = \frac{1}{\Gamma(\alpha)} \int_0^{+\infty} y^{\alpha+1} e^{-y} dy = \frac{\Gamma(\alpha+2)}{\Gamma(\alpha)} = \alpha(\alpha+1),$$

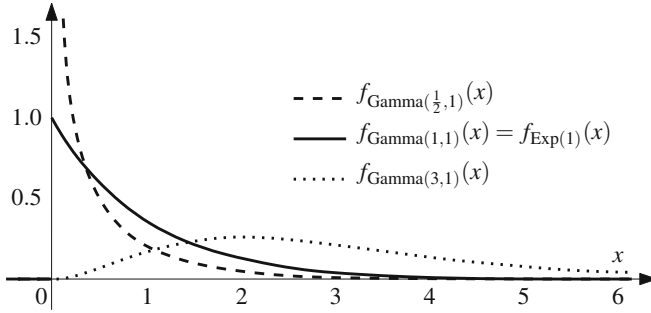


Figura 6.3 Densità di una variabile aleatoria $\text{Gamma}(\alpha, \lambda)$ con $\lambda = 1$, per $\alpha = \frac{1}{2}$, $\alpha = 1$ e $\alpha = 3$. Per $\alpha = 1$ si ha $\text{Gamma}(1, 1) = \text{Exp}(1)$ (si veda il Paragrafo 6.3.3)

da cui, facilmente,

$$\text{Var}(Y) = E(Y^2) - E(Y)^2 = \alpha.$$

Per la relazione (6.35), $X := \frac{1}{\lambda}Y \sim \text{Gamma}(\alpha, \lambda)$. Per la linearità del valor medio e le proprietà della varianza (Proposizione 3.69), otteniamo dunque media e varianza di una variabile aleatoria $X \sim \text{Gamma}(\alpha, \lambda)$:

$$E(X) = \frac{\alpha}{\lambda}, \quad \text{Var}(X) = \frac{\alpha}{\lambda^2}. \quad (6.36)$$

In alternativa, i risultati precedenti si sarebbero potuti ottenere facilmente calcolando la funzione generatrice dei momenti. Infatti

$$M_X(t) = \int_{-\infty}^{+\infty} e^{tx} f_X(x) dx = \frac{\lambda^\alpha}{\Gamma(\alpha)} \int_0^{+\infty} x^{\alpha-1} e^{-(\lambda-t)x} dx.$$

Quest'ultimo integrale è finito solo per $t < \lambda$, e in tal caso, con il cambio di variabile $x = z/(\lambda - t)$ si ottiene

$$\int_0^{+\infty} x^{\alpha-1} e^{-(\lambda-t)x} dx = \frac{1}{(\lambda-t)^\alpha} \int_0^{+\infty} z^{\alpha-1} e^{-z} dz = \frac{\Gamma(\alpha)}{(\lambda-t)^\alpha}.$$

In definitiva

$$M_X(t) = \begin{cases} \left(\frac{\lambda}{\lambda-t} \right)^\alpha & \text{per } t < \lambda \\ +\infty & \text{per } t \geq \lambda \end{cases}.$$

Concludiamo con una proprietà importante delle variabili aleatorie Gamma.

Proposizione 6.29 (Somma di Gamma indipendenti). *Siano X e Y variabili aleatorie reali indipendenti, con $X \sim \text{Gamma}(\alpha, \lambda)$ e $Y \sim \text{Gamma}(\beta, \lambda)$. Allora $X + Y \sim \text{Gamma}(\alpha + \beta, \lambda)$.*

Dimostrazione. Usando il Teorema 6.25, per $z > 0$ abbiamo:

$$\begin{aligned} f_{X+Y}(z) &= \int f_X(x) f_Y(z-x) dx \\ &= \frac{\lambda^\alpha \lambda^\beta}{\Gamma(\alpha) \Gamma(\beta)} \int \mathbb{1}_{(0,+\infty)}(x) x^{\alpha-1} e^{-\lambda x} \mathbb{1}_{(0,+\infty)}(z-x) (z-x)^{\beta-1} e^{-\lambda(z-x)} dx \\ &= \frac{\lambda^{\alpha+\beta}}{\Gamma(\alpha) \Gamma(\beta)} e^{-\lambda z} \int_0^z x^{\alpha-1} (z-x)^{\beta-1} dx. \end{aligned}$$

Con il cambio di variabile $x = zt$ si ottiene dunque

$$f_{X+Y}(z) = C z^{\alpha+\beta-1} e^{-\lambda z}, \quad \text{con} \quad C := \frac{\lambda^{\alpha+\beta}}{\Gamma(\alpha) \Gamma(\beta)} \int_0^1 t^{\alpha-1} (1-t)^{\beta-1} dt.$$

Essendo proporzionale a $z^{\alpha+\beta-1} e^{-\lambda z}$, $f_{X+Y}(z)$ è necessariamente la densità di una Gamma($\alpha + \beta, \lambda$) (e la costante C coincide con $\lambda^{\alpha+\beta} / \Gamma(\alpha + \beta)$). \square

Osservazione 6.30. Definiamo la funzione $\beta : (0, \infty) \times (0, \infty) \rightarrow (0, \infty)$, che ritroveremo nel Paragrafo 6.5.2, ponendo

$$\beta(a, b) := \int_0^1 t^{a-1} (1-t)^{b-1} dt. \quad (6.37)$$

Vale allora la non ovvia identità

$$\Gamma(a+b) = \frac{\Gamma(a) \Gamma(b)}{\beta(a, b)}, \quad \forall a, b > 0, \quad (6.38)$$

come segue dalla dimostrazione precedente. \square

6.3.3 Esponenziale

Di grande interesse applicativo sono alcuni casi particolari di variabili aleatorie Gamma. In questo paragrafo introduciamo le variabili aleatorie dette esponenziali, mentre nel Paragrafo 6.5.1 vedremo le chi-quadro.

Una variabile aleatoria reale X si dice *esponenziale di parametro λ* , e scriveremo $X \sim \text{Exp}(\lambda)$, se $X \sim \text{Gamma}(1, \lambda)$, ossia se

$$f_X(x) = f_{\text{Exp}(\lambda)}(x) := \lambda e^{-\lambda x} \mathbb{1}_{[0,+\infty)}(x)$$

(si veda la Figura 6.3). La funzione di ripartizione è data da $F_X(x) = 0$ per $x < 0$, mentre per $x \geq 0$

$$F_X(x) = \int_{-\infty}^x f_X(t) dt = \int_0^x \lambda e^{-\lambda t} dt = (-e^{-\lambda t}) \Big|_0^x = 1 - e^{-\lambda x}. \quad (6.39)$$

Ricordando la relazione (6.36), media e varianza di $X \sim \text{Exp}(\lambda)$ valgono

$$E(X) = \frac{1}{\lambda}, \quad \text{Var}(X) = \frac{1}{\lambda^2}. \quad (6.40)$$

Le variabili aleatorie esponenziali possono essere viste come l'analogo continuo delle variabili aleatorie geometriche. Esse, infatti, soddisfano ad una proprietà di assenza di memoria del tutto analoga.

Proposizione 6.31 (Assenza di memoria). Se $X \sim \text{Exp}(\lambda)$,

$$P(X > s+t | X > s) = P(X > t), \quad \forall s, t \geq 0.$$

Dimostrazione. Osservando che

$$P(X > s+t | X > s) = \frac{P(X > s+t)}{P(X \geq s)},$$

ed essendo $P(X > s) = 1 - P(X \leq s) = 1 - F_X(s) = e^{-\lambda s}$, il risultato desiderato segue immediatamente da (6.39). \square

Come per le variabili aleatorie geometriche, si può dimostrare (non lo faremo) che le variabili aleatorie esponenziali sono le uniche variabili aleatorie a valori reali positivi per cui vale la precedente proprietà di assenza di memoria.

Le variabili aleatorie esponenziali sono spesso usate come modelli per “tempi di attesa”, quali ad esempio il tempo di decadimento di atomi radioattivi, l'intervallo temporale tra due terremoti successivi, il tempo intercorrente tra l'arrivo di due clienti ad uno sportello, ecc. Questa interpretazione verrà approfondita nei Paragrafi 6.5.3 e 6.5.4, in cui presentiamo il *processo di Poisson*, un modello probabilistico di grande importanza, costruito a partire da variabili aleatorie esponenziali.

In analogia con le variabili aleatorie geometriche (si veda l'Esempio 3.118), mostriamo che il minimo di variabili aleatorie esponenziali indipendenti ha ancora distribuzione esponenziale.

Esempio 6.32. Siano X_1, X_2, \dots, X_n variabili aleatorie indipendenti, con $X_i \sim \text{Exp}(\lambda_i)$. Allora

$$W := \min(X_1, X_2, \dots, X_n) \sim \text{Exp}(\lambda_1 + \lambda_2 + \dots + \lambda_n).$$

Usando la Proposizione 3.93, segue da (6.39) che per ogni $x > 0$

$$F_W(x) = 1 - \prod_{k=1}^n [1 - F_{X_k}(x)] = 1 - \exp[-(\lambda_1 + \lambda_2 + \dots + \lambda_n)x],$$

Derivando quest'ultima identità la conclusione segue immediatamente. \square

6.3.4 Normale

Una variabile aleatoria reale assolutamente continua Z si dice *normale* (o *gaussiana standard*), e si scrive $Z \sim N(0, 1)$, se è assolutamente continua con densità

$$f_Z(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}. \quad (6.41)$$

Notiamo che si tratta di una buona definizione, come segue dall'integrale di Gauss (6.7) con un semplice cambio di variabili. Per la formula (6.21)

$$E(Z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} x e^{-\frac{x^2}{2}} dx = 0,$$

perché la funzione $x e^{-x^2/2}$ è integrabile su \mathbb{R} e dispari. Inoltre, integrando per parti,

$$\begin{aligned} \text{Var}(Z) &= E(Z^2) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} x^2 e^{-\frac{x^2}{2}} dx \\ &= \frac{1}{\sqrt{2\pi}} \left[-x e^{-\frac{x^2}{2}} \Big|_{-\infty}^{+\infty} + \int_{-\infty}^{+\infty} e^{-\frac{x^2}{2}} dx \right] = 1. \end{aligned}$$

Data $Z \sim N(0, 1)$ e i numeri reali $\mu \in \mathbb{R}$, $\sigma \geq 0$, definiamo la variabile aleatoria

$$X := \sigma Z + \mu. \quad (6.42)$$

Dalle proprietà elementari di valor medio e varianza segue che

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

Se $\sigma = 0$, X non è altro che la costante μ . Se invece $\sigma > 0$, per la Proposizione 6.18, la variabile aleatoria X è assolutamente continua con densità

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

Una variabile aleatoria con la distribuzione di X si dice *normale* (o *gaussiana*) di media $\mu \in \mathbb{R}$ e varianza $\sigma^2 \geq 0$, e si scrive $X \sim N(\mu, \sigma^2)$. Più esplicitamente:

- $X \sim N(\mu, \sigma^2)$, con $\sigma > 0$, se e solo se X è assolutamente continua con densità

$$f_X(x) = f_{N(\mu, \sigma^2)}(x) := \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (6.43)$$

(si veda la Figura 6.4).

- $X \sim N(\mu, 0)$ se e solo se $X \equiv \mu$ è (quasi certamente) costante.

Sottolineiamo che $X \sim N(\mu, \sigma^2)$ è una variabile aleatoria assolutamente continua solo se $\sigma^2 > 0$, mentre, se $\sigma^2 = 0$, X è una variabile aleatoria discreta.

Invertendo la costruzione precedente, segue che

$$\text{per } \sigma > 0: \quad X \sim N(\mu, \sigma^2) \quad \Longleftrightarrow \quad Z := \frac{X - \mu}{\sigma} \sim N(0, 1),$$

una relazione utile quando si lavora con variabili aleatorie normali. Come semplice corollario (esercizio), otteniamo il seguente risultato.

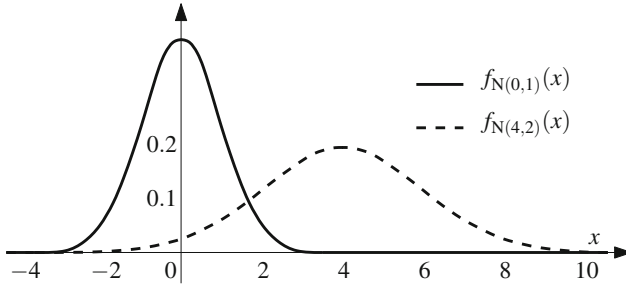


Figura 6.4 Densità di una variabile aleatoria $N(\mu, \sigma^2)$, per $(\mu = 0, \sigma = 1)$ e $(\mu = 4, \sigma = 2)$

Proposizione 6.33 (Trasformazioni lineari-affini di normali). *Una trasformazione lineare-affine di una variabile aleatoria normale è ancora una variabile aleatoria normale: $\forall a, b \in \mathbb{R}$*

$$X \sim N(\mu, \sigma^2) \implies aX + b \sim N(\tilde{\mu}, \tilde{\sigma}^2),$$

dove (come segue dalle trasformazioni di valor medio e varianza)

$$\tilde{\mu} = a\mu + b, \quad \tilde{\sigma}^2 = a^2\sigma^2.$$

Le variabili aleatorie normali sono tra le più usate nelle applicazioni. Detto in modo grossolano, ogni qual volta una quantità aleatoria è la somma di molte quantità aleatorie indipendenti e “non troppo grandi”, la sua distribuzione è approssimativamente normale. Una versione parziale, ma rigorosa, di tale affermazione, verrà data nel Capitolo 7, con il teorema limite centrale.

Anche per le variabili aleatorie normali è facile calcolare la funzione generatrice dei momenti. Se $Z \sim N(0, 1)$, si ha

$$M_Z(t) = \int_{-\infty}^{+\infty} e^{tz} f_Z(z) dz = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{tx - \frac{1}{2}x^2} dx.$$

Questo integrale si può calcolare usando il metodo del *completamento dei quadrati*, molto utile per integrali gaussiani. Si tratta di osservare che

$$tx - \frac{1}{2}x^2 = -\frac{1}{2}(x-t)^2 + \frac{t^2}{2},$$

pertanto

$$M_Z(t) = e^{\frac{t^2}{2}} \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x-t)^2} dx.$$

Si osservi che la funzione integranda è la densità di una variabile aleatoria normale

di media t e varianza 1, quindi l'integrale vale 1. Ne segue che

$$M_Z(t) = e^{\frac{t^2}{2}}. \quad (6.44)$$

Passando al caso generale, una variabile aleatoria $X \sim N(\mu, \sigma^2)$ si può scrivere nella forma $X = \sigma Z + \mu$, con $Z := (X - \mu)/\sigma \sim N(0, 1)$, pertanto

$$M_X(t) = E(e^{tX}) = E(e^{t(\sigma Z + \mu)}) = e^{\mu t} E(e^{t\sigma Z}) = e^{\mu t} M_Z(t\sigma) = e^{\mu t + \frac{\sigma^2}{2} t^2}.$$

Concludiamo con una proprietà di fondamentale importanza: la somma di due variabili aleatorie normali *indipendenti* è normale.

Proposizione 6.34 (Somma di normali indipendenti). *Siano X e Y variabili aleatorie normali indipendenti: $X \sim N(\mu_1, \sigma_1^2)$ e $Y \sim N(\mu_2, \sigma_2^2)$. Allora anche la loro somma è normale: $X + Y \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$.*

Dimostrazione. Se $\sigma_1^2 = 0$ (risp. $\sigma_2^2 = 0$) la conclusione segue immediatamente dalla Proposizione 6.33, perché in tal caso $X_1 = \mu_1$ (risp. $X_2 = \mu_2$) quasi certamente. Quindi assumiamo per il seguito che $\sigma_1^2 > 0$ e $\sigma_2^2 > 0$.

Se $X \sim N(\mu_1, \sigma_1^2)$ e $Y \sim N(\mu_2, \sigma_2^2)$, per la Proposizione 6.33 si ha $(X - \mu_1) \sim N(0, \sigma_1^2)$ e $(Y - \mu_2) \sim N(0, \sigma_2^2)$. Supponendo di aver dimostrato la conclusione quando entrambe le medie sono nulle, si ha

$$(X - \mu_1) + (Y - \mu_2) \sim N(0, \sigma_1^2 + \sigma_2^2),$$

e applicando ancora la Proposizione 6.33 otteniamo la tesi:

$$X + Y = [(X - \mu_1) + (Y - \mu_2)] + (\mu_1 + \mu_2) \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2).$$

Resta da considerare il caso $\mu_1 = \mu_2 = 0$. Dal Teorema 6.25, si ha

$$f_{X+Y}(z) = \int_{-\infty}^{+\infty} f_X(x) f_Y(z-x) dx = \frac{1}{2\pi\sigma_1\sigma_2} \int_{-\infty}^{+\infty} \exp\left[-\frac{x^2}{2\sigma_1^2} - \frac{(z-x)^2}{2\sigma_2^2}\right] dx.$$

Introducendo le costanti

$$\xi := \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2}, \quad \hat{\sigma} := \left(\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}\right)^{-1/2} = \frac{\sigma_1\sigma_2}{\sqrt{\sigma_1^2 + \sigma_2^2}}, \quad (6.45)$$

si verifica con un calcolo (completamento dei quadrati) che vale la seguente identità:

$$\frac{x^2}{2\sigma_1^2} + \frac{(z-x)^2}{2\sigma_2^2} = \frac{z^2}{2(\sigma_1^2 + \sigma_2^2)} + \frac{(x - \xi z)^2}{2\hat{\sigma}^2}.$$

Di conseguenza

$$f_{X+Y}(z) = \frac{1}{2\pi\sigma_1\sigma_2} \exp\left[-\frac{z^2}{2(\sigma_1^2 + \sigma_2^2)}\right] \int_{-\infty}^{+\infty} \exp\left[-\frac{(x - \xi z)^2}{2\hat{\sigma}^2}\right] dx.$$

La funzione che compare nell'integrale coincide con la densità di una $N(\xi z, \hat{\sigma}^2)$ a meno del fattore moltiplicativo $(2\pi)^{-1/2} \hat{\sigma}^{-1}$, cf. (6.43), pertanto

$$\int_{-\infty}^{+\infty} \exp \left[-\frac{(x - \xi z)^2}{2\hat{\sigma}^2} \right] dx = \hat{\sigma} \sqrt{2\pi} \int_{-\infty}^{+\infty} f_{N(\xi z, \hat{\sigma}^2)}(x) dx = \hat{\sigma} \sqrt{2\pi},$$

da cui si ottiene, ricordando (6.45),

$$f_{X+Y}(z) = \frac{1}{\sqrt{2\pi} \sqrt{\sigma_1^2 + \sigma_2^2}} \exp \left[-\frac{z^2}{2(\sigma_1^2 + \sigma_2^2)} \right],$$

cioè $X + Y \sim N(0, \sigma_1^2 + \sigma_2^2)$. □

Corollario 6.35. *Siano X_1, \dots, X_n variabili aleatorie normali indipendenti. Ogni combinazione lineare-affine $X := a_1 X_1 + \dots + a_n X_n + b$, con $a_1, \dots, a_n, b \in \mathbb{R}$, è una variabile aleatoria normale.*

Dimostrazione. Il caso $n = 1$ non è altro che la Proposizione 6.33. Procedendo per induzione, definiamo

$$Y := a_1 X_1 + \dots + a_{n-1} X_{n-1}, \quad W := a_n X_n + b,$$

così che, per il passo induttivo, sia Y che W sono variabili aleatorie normali. Esse sono inoltre indipendenti, perché costruite a partire da insiemi disgiunti di variabili aleatorie indipendenti (si ricordi il Corollario 3.41). Di conseguenza $X = Y + W$ è normale, per la Proposizione 6.34. □

Esercizi

Esercizio 6.3. Sia X una variabile aleatoria con distribuzione $\text{Exp}(\lambda)$, definita su uno spazio di probabilità (Ω, \mathcal{A}, P) . Per $s \in [0, +\infty)$, definiamo

$$P^{(s)} := P(\cdot | X > s), \quad X^{(s)} := X - s.$$

Si mostri che la variabile aleatoria $X^{(s)}$, rispetto alla probabilità (condizionale) $P^{(s)}$, ha distribuzione $\text{Exp}(\lambda)$, per ogni $s \in [0, \infty)$.

[Sugg. Si calcoli la funzione di ripartizione $P^{(s)}(X^{(s)} \leq x)$, sfruttando la Proposizione 6.31.]

6.4 Vettori aleatori assolutamente continui *

Il contenuto di questo paragrafo richiede la conoscenza di una nozione di integrale per funzioni di due o più variabili reali, ed è pertanto da considerarsi più avanzato

di quanto finora trattato. Per questa ragione, assumeremo una buona conoscenza dell'integrale di Riemann multidimensionale. Diremo che una funzione $f : \mathbb{R}^n \rightarrow \mathbb{R}$ è integrabile se il suo integrale sul dominio illimitato \mathbb{R}^n esiste finito.

Per ogni sottoinsieme $C \subseteq \mathbb{R}^n$ la cui funzione indicatrice $\mathbb{1}_C : \mathbb{R}^n \rightarrow \mathbb{R}$ è integrabile, si definisce *misura n -dimensionale di C* (secondo Peano-Jordan) la quantità

$$\text{mis}(C) := \int_{\mathbb{R}^n} \mathbb{1}_C(x) dx \in [0, +\infty), \quad (6.46)$$

I sottoinsiemi C per cui $\text{mis}(C) = 0$ sono detti di misura nulla. Questi comprendono gli insiemi dati dall'unione di un numero finito di sottospazi lineari-affini *propri* (ossia con dimensione strettamente minore di n); in particolare, ogni sottoinsieme finito di punti ha misura nulla (e, più in generale, ogni sottoinsieme di punti isolati). Se $f, g : \mathbb{R}^n \rightarrow \mathbb{R}$ sono funzioni che differiscono solo su un insieme di misura nulla, allora f è integrabile se e solo se lo è g , nel qual caso gli integrali coincidono.

Ricordiamo l'importante teorema di Fubini-Tonelli, limitandoci per semplicità al caso bidimensionale: per ogni funzione $f : \mathbb{R}^2 \rightarrow [0, \infty)$ *positiva* e integrabile si ha

$$\int_{\mathbb{R}^2} f(x, y) dx dy = \int_{-\infty}^{+\infty} \left(\int_{-\infty}^{+\infty} f(x, y) dx \right) dy = \int_{-\infty}^{+\infty} \left(\int_{-\infty}^{+\infty} f(x, y) dy \right) dx,$$

purché la funzione $x \mapsto f(x, y)$ sia integrabile su \mathbb{R} per ogni $y \in \mathbb{R}$, e analogamente lo sia $y \mapsto f(x, y)$ per ogni $x \in \mathbb{R}$.[†] Lo stesso vale per funzioni $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ non necessariamente positive, purché le funzioni positive f^+ e f^- siano entrambe integrabili (equivalentemente, purché sia f che $|f|$ siano integrabili).

Dato $V \subseteq \mathbb{R}^n$ un insieme aperto, una funzione $\psi : V \rightarrow \mathbb{R}^n$ si dice *diffeomorfismo sull'immagine* se

- ψ è iniettiva;
- ψ è differenziabile in ogni punto $z \in V$;
- la matrice jacobiana $D\psi(z)$ è non singolare ($\det D\psi(z) \neq 0$) per ogni $z \in V$.

Si noti che, per il teorema di differenziazione della funzione inversa, la terza condizione è equivalente a richiedere che l'immagine $U := \psi(V)$ sia un aperto e che la funzione inversa $\psi^{-1} : U \rightarrow V$ sia differenziabile in ogni punto $x \in U$. Indicando esplicitamente l'immagine, diremo che $\psi : V \rightarrow U$ è un diffeomorfismo.

Richiamiamo infine la formula di “cambio di variabili”. Siano $U, V \subseteq \mathbb{R}^n$ aperti e sia $\psi : V \rightarrow U$ un diffeomorfismo. Per ogni funzione $g : U \rightarrow \mathbb{R}$ integrabile, la funzione $g(\psi(z)) |\det D\psi(z)| : V \rightarrow \mathbb{R}$ è integrabile e si ha

$$\int_U g(x) dx = \int_V g(\psi(z)) |\det D\psi(z)| dz, \quad (6.47)$$

che costituisce l'analogo multidimensionale della relazione (6.12).

[†] Queste ipotesi tecniche possono essere rimosse definendo opportunamente gli integrali interni per i valori “eccezionali” di x e y , ma non lo faremo.

Esempio 6.36 (Integrale di Gauss). Come applicazione della formula di cambio di variabili, calcoliamo il valore dell'integrale di Gauss, dimostrando la formula (6.7). In altri termini, ponendo

$$c := \int_{-\infty}^{+\infty} e^{-x^2} dx, \quad (6.48)$$

mostriamo che $c = \sqrt{\pi}$. Consideriamo la funzione $g : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ definita da $g(x, y) := e^{-x^2 - y^2}$. Per il teorema di Fubini-Tonelli, si ha

$$\int_{\mathbb{R}^2} g(x, y) dx dy = \int_{-\infty}^{+\infty} e^{-y^2} \left(\int_{-\infty}^{+\infty} e^{-x^2} dx \right) dy = c^2. \quad (6.49)$$

Passiamo a *coordinate polari*: indicando con $C := \{(x, y) \in \mathbb{R}^2 : x \geq 0, y = 0\}$ il semiasse positivo delle x , definiamo gli aperti di \mathbb{R}^2

$$U := \mathbb{R}^2 \setminus C, \quad V := (0, +\infty) \times (0, 2\pi),$$

e consideriamo la funzione $\psi : V \rightarrow U$ definita da $\psi(r, \theta) := (r \cos \theta, r \sin \theta)$, che si verifica facilmente essere un diffeomorfismo, con matrice jacobiana

$$D\psi(r, \theta) = \begin{pmatrix} \cos \theta & r \sin \theta \\ \sin \theta & -r \cos \theta \end{pmatrix},$$

da cui segue che $|\det D\psi(r, \theta)| = r$. L'insieme C ha misura 2-dimensionale nulla, quindi possiamo restringere il dominio del primo integrale in (6.49) a $\mathbb{R}^2 \setminus C = U$. Applicando la formula (6.47) e il teorema di Fubini-Tonelli, otteniamo quindi

$$\begin{aligned} c^2 &= \int_U g(x, y) dx dy = \int_V g(\psi(r, \theta)) r dr d\theta \\ &= \int_{(0, +\infty) \times (0, 2\pi)} e^{-r^2(\cos \theta)^2 - r^2(\sin \theta)^2} r dr d\theta \\ &= \int_0^{2\pi} \left(\int_0^{+\infty} r e^{-r^2} dr \right) d\theta = \int_0^{2\pi} \left(-\frac{1}{2} e^{-r^2} \Big|_0^{\infty} \right) d\theta = \frac{2\pi}{2} = \pi, \end{aligned}$$

e dunque $c = \sqrt{\pi}$, come annunciato. \square

6.4.1 Definizione e prime proprietà *

L'estensione al caso vettoriale della nozione di variabile aleatoria assolutamente continua è lo scopo della seguente definizione.

Definizione 6.37 (Vettore aleatorio assolutamente continuo). Un vettore aleatorio n -dimensionale X , ossia una variabile aleatoria X a valori in \mathbb{R}^n , si dice *assolutamente continuo* se esiste una funzione $f_X : \mathbb{R}^n \rightarrow [0, +\infty)$ integrabile su \mathbb{R}^n tale che, per ogni scelta di I_1, \dots, I_n intervalli (non necessariamente limitati) di \mathbb{R} , si abbia

$$P(X \in I_1 \times \dots \times I_n) = \int_{I_1 \times \dots \times I_n} f_X(x) dx. \quad (6.50)$$

La funzione f_X si dice *densità* del vettore aleatorio X .

In analogia con il caso unidimensionale, valgono le seguenti proprietà.

- Scegliendo $I_1 = \dots = I_n = \mathbb{R}$ in (6.50), segue che

$$\int_{\mathbb{R}^n} f_X(x) dx = 1.$$

Viceversa, per ogni funzione positiva $f : \mathbb{R}^n \rightarrow [0, +\infty)$, integrabile su \mathbb{R}^n con integrabile 1, esiste un vettore aleatorio X , definito su un opportuno spazio di probabilità (Ω, \mathcal{A}, P) , assolutamente continuo con densità $f_X = f$.

- La densità f_X di un vettore aleatorio assolutamente continuo X non è univocamente identificata. Infatti, se $g : \mathbb{R}^n \rightarrow [0, +\infty)$ è una funzione che coincide con f_X al di fuori di un sottoinsieme $C \subseteq \mathbb{R}^n$ di misura nulla, la relazione (6.50) continua a valere con g al posto di f_X ; pertanto g è una densità di X .

Osservazione 6.38. Si può dimostrare che da (6.50) segue che

$$P(X \in C) = \int_C f_X(x) dx = \int_{\mathbb{R}^n} f_X(x) \mathbb{1}_C(x) dx, \quad (6.51)$$

per ogni sottoinsieme $C \subseteq \mathbb{R}^n$ tale che $\mathbb{1}_C(x)f(x)$ è integrabile su \mathbb{R}^n . Questo permette di calcolare le probabilità degli eventi di interesse che coinvolgono X .

Se si conosce la teoria dell'integrazione secondo Lebesgue su \mathbb{R}^n , che permette di estendere l'integrale di Riemann a una classe più vasta di funzioni, la relazione (6.51) vale per ogni sottoinsieme $C \subseteq \mathbb{R}^n$ boreliano, ossia per ogni $C \in \mathcal{B}(\mathbb{R}^n)$. Questo mostra che la densità di un vettore aleatorio assolutamente continuo ne caratterizza la distribuzione: se due vettori aleatori assolutamente continui hanno la stessa densità, allora hanno anche la stessa distribuzione. \square

Le variabili aleatorie reali con distribuzione uniforme su un intervallo $(a, b) \subseteq \mathbb{R}$, introdotte nel Paragrafo 6.3.1, ammettono una generalizzazione al caso vettoriale, che presentiamo nell'esempio seguente. Un'altra classe importante è costituita dai vettori aleatori normali, generalizzazione multidimensionale delle variabili aleatorie reali normali, che studieremo nel Paragrafo 6.6.

Esempio 6.39 (Vettori aleatori uniformi continui). Sia $C \subseteq \mathbb{R}^n$ un sottoinsieme con $0 < \text{mis}(C) < +\infty$. Un vettore aleatorio n -dimensionale X si dice *uniforme su C* , e si

scrive $X \sim U(C)$, se X è assolutamente continuo con densità

$$f_X(x) = \frac{1}{\text{mis}(C)} \mathbb{1}_C(x).$$

Un vettore aleatorio $X \sim U(C)$ traduce in termini matematici l'idea di "scegliere a caso uniformemente" un punto nell'insieme C . Pur avendo un'apparenza innocua, tale interpretazione va presa con attenzione, come mostra un celebre paradosso, dovuto al matematico francese J. Bertrand, che descriviamo nel Paragrafo 6.5.5. \square

6.4.2 Densità congiunta e marginali *

I risultati di questo paragrafo sono strettamente analoghi a quelli dimostrati nel Capitolo 3 per variabili aleatorie discrete.

Dato un vettore aleatorio assolutamente continuo $X = (X_1, \dots, X_n)$ a valori in \mathbb{R}^n , la sua densità f_X viene anche indicata con f_{X_1, \dots, X_n} , quando si voglia sottolineare la dipendenza dalle n componenti del vettore. Mostriamo ora che ogni sottoinsieme di componenti $X_J := (X_i)_{i \in J}$, con $J \subseteq \{1, \dots, n\}$, forma un vettore aleatorio assolutamente continuo a valori in $\mathbb{R}^{|J|}$, con densità esplicita. Per semplicità di formulazione, enunciamo il risultato nel caso in cui $J = \{1, \dots, k\}$.

Proposizione 6.40 (Densità congiunte e marginali). *Sia $X = (X_1, \dots, X_n)$ un vettore aleatorio assolutamente continuo, a valori in \mathbb{R}^n , con densità $f_X = f_{X_1, \dots, X_n}$. Per ogni $k < n$, il vettore aleatorio (X_1, \dots, X_k) , a valori in \mathbb{R}^k , è assolutamente continuo, con densità*

$$f_{X_1, \dots, X_k}(x_1, \dots, x_k) = \int f_X(x_1, \dots, x_k, x_{k+1}, \dots, x_n) dx_{k+1} \cdots dx_n. \quad (6.52)$$

Dimostrazione. Siano I_1, \dots, I_k intervalli di \mathbb{R} e poniamo $I_m = \mathbb{R}$ per $k < m \leq n$. Applicando il teorema di Fubini-Tonelli alla relazione (6.50), con tale scelta di intervalli, si ha

$$\begin{aligned} P(X_1 \in I_1, \dots, X_k \in I_k) &= P(X_1 \in I_1, \dots, X_k \in I_k, X_{k+1} \in I_{k+1}, \dots, X_n \in I_n) \\ &= \int_{I_1 \times \dots \times I_k} \left[\int_{\mathbb{R}^{n-k}} f_X(x_1, \dots, x_k, x_{k+1}, \dots, x_n) dx_{k+1} \cdots dx_n \right] dx_1 \cdots dx_k, \end{aligned}$$

da cui la conclusione segue. \square

In particolare, se (X, Y) è un vettore aleatorio assolutamente continuo bidimensionale, ossia a valori in \mathbb{R}^2 , le sue componenti X e Y sono variabili aleatorie reali assolutamente continue, con densità

$$f_X(x) = \int_{-\infty}^{+\infty} f_{X,Y}(x, y) dy, \quad f_Y(y) = \int_{-\infty}^{+\infty} f_{X,Y}(x, y) dx. \quad (6.53)$$

La densità f_{X_1, \dots, X_n} di un vettore aleatorio (X_1, \dots, X_n) è anche detta densità congiunta delle variabili aleatorie X_1, \dots, X_n , mentre le densità f_{X_i} delle singole componenti sono dette densità marginali. Anche per variabili aleatorie assolutamente continue l'indipendenza può essere caratterizzata in funzione delle densità congiunte e marginali, come mostra il seguente risultato, analogo alla Proposizione 3.32.

Proposizione 6.41 (Indipendenza e densità). *Sia $X = (X_1, \dots, X_n)$ un vettore aleatorio assolutamente continuo n -dimensionale, per cui vale la seguente relazione per ogni $x = (x_1, \dots, x_n) \in \mathbb{R}^n$ (a meno di un insieme di misura n -dimensionale nulla):*

$$f_{X_1, \dots, X_n}(x_1, \dots, x_n) = f_{X_1}(x_1) \cdots f_{X_n}(x_n). \quad (6.54)$$

Allora le componenti X_1, \dots, X_n sono variabili aleatorie reali indipendenti.

Viceversa, se X_1, \dots, X_n sono variabili aleatorie reali indipendenti e assolutamente continue, con densità f_{X_1}, \dots, f_{X_n} , il vettore aleatorio $X = (X_1, \dots, X_n)$ è assolutamente continuo, con densità data da (6.54).

Nel caso di due variabili aleatorie reali X e Y , la relazione (6.54) diventa

$$f_{X,Y}(x, y) = f_X(x) f_Y(y). \quad (6.55)$$

In effetti, per dimostrare l'indipendenza di due variabili aleatorie X e Y , è sufficiente mostrare che la densità congiunta $f_{X,Y}(x, y)$ si fattorizza come prodotto di una funzione di x per una funzione di y , analogamente al caso discreto (si ricordi l'Osservazione 3.34). La dimostrazione è lasciata al lettore, nell'Esercizio 6.4.

Osservazione 6.42. In generale, per dedurre che X e Y non sono indipendenti, non basta mostrare che $f_{X,Y}(x_0, y_0) \neq f_X(x_0) f_Y(y_0)$ in un punto (x_0, y_0) , perché il valore di una densità in un punto può essere modificato arbitrariamente; tuttavia, questo basta se le densità sono funzioni continue (si veda l'Esercizio 6.5). \square

Osservazione 6.43. La Proposizione 6.41 continua a valere se le X_i sono esse stesse vettori aleatori. In questo caso, se X_i ha dimensione m_i , X è un vettore aleatorio di dimensione $m_1 + m_2 + \cdots + m_n$. \square

Vediamo ora un esempio concreto.

Esempio 6.44. Sia (X, Y) un vettore aleatorio bidimensionale, con distribuzione uniforme nel cerchio unitario C_1 , dove poniamo $C_r := \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \leq r^2\}$. Determiniamo le distribuzioni delle componenti X e Y e di $R := \sqrt{X^2 + Y^2}$.

Dato che $\text{mis}(C_1) = \pi$, la densità di (X, Y) è data da

$$f_{X,Y}(x, y) = \frac{1}{\pi} \mathbb{1}_{C_1}(x, y). \quad (6.56)$$

Per la relazione (6.53), X è assolutamente continua con densità

$$f_X(x) = \int_{-\infty}^{+\infty} \frac{1}{\pi} \mathbb{1}_C(x, y) dy = \int_{-\infty}^{+\infty} \frac{1}{\pi} \mathbb{1}_{\{y \in \mathbb{R}: y^2 \leq 1-x^2\}} dy.$$

L'insieme degli $y \in \mathbb{R}$ che soddisfano la relazione $y^2 \leq 1-x^2$ è vuoto se $|x| > 1$, mentre è l'intervallo $[-\sqrt{1-x^2}, +\sqrt{1-x^2}]$ se $|x| \leq 1$. Pertanto

$$f_X(x) = \mathbb{1}_{[-1,1]}(x) \left(\int_{-\sqrt{1-x^2}}^{+\sqrt{1-x^2}} \frac{1}{\pi} dy \right) = \frac{2}{\pi} \sqrt{1-x^2} \mathbb{1}_{[-1,1]}(x). \quad (6.57)$$

Con un calcolo analogo, o più semplicemente con un argomento di simmetria, si mostra che Y ha la stessa densità di X .

Determiniamo ora la distribuzione della variabile aleatoria reale $R := \sqrt{X^2 + Y^2}$. Non sapendo a priori se R sia assolutamente continua, cominciamo a calcolarne la funzione di ripartizione. Chiaramente $F_R(r) = 0$ se $r < 0$ e $F_R(r) = 1$ se $r \geq 1$, perché $0 \leq R \leq 1$, mentre per $r \in [0, 1]$

$$\begin{aligned} F_R(r) &= P(R \leq r) = P(X^2 + Y^2 \leq r^2) = P((X, Y) \in C_r) \\ &= \int_{C_r} f_{X,Y}(x, y) dx dy = \frac{1}{\pi} \text{mis}(C_r) = \frac{1}{\pi} (\pi r^2) = r^2, \end{aligned}$$

avendo usato le relazioni (6.51), (6.56), (6.46) e il fatto che $C_r \subseteq C_1$. Questo mostra che la funzione F_R è continua su \mathbb{R} e di classe \mathcal{C}^1 su $\mathbb{R} \setminus \{0, 1\}$, dunque è \mathcal{C}^1 a tratti: per la Proposizione 6.16, R è assolutamente continua con densità

$$f_R(r) = F'_R(r) = 2r \mathbb{1}_{[0,1]}(r). \quad (6.58)$$

Concludiamo notando che le variabili aleatorie X e Y *non* sono indipendenti, perché la relazione (6.55) non è verificata a meno di un insieme di punti $(x, y) \in \mathbb{R}^2$ di misura nulla. Infatti la densità congiunta $f_{X,Y}(x, y)$, data da (6.56), è nulla al di fuori del cerchio unitario C_1 , e in particolare per $(x, y) \in Q := (\frac{3}{4}, 1) \times (\frac{3}{4}, 1)$, un insieme di misura positiva; invece $f_X(x) > 0$ e $f_Y(y) > 0$ se $(x, y) \in Q$, come segue dalla relazione (6.57), essendo $f_Y = f_X$. \square

Sottolineiamo che le variabili aleatorie X e R dell'esempio precedente, che rappresentano rispettivamente l'ascissa e la distanza dall'origine di un punto scelto uniformemente nel cerchio unitario, *non* hanno distribuzione uniforme sugli intervalli in cui assumono valori, ossia rispettivamente $[-1, +1]$ e $[0, 1]$, contrariamente a quanto un'intuizione errata potrebbe suggerire: si vedano le relazioni (6.57) e (6.58). Ad esempio, per quanto riguarda R , i valori in prossimità di $r = 1$, in cui la densità $f_R(r)$ è massima, sono più probabili di quelli vicino a $r = 0$, in cui la densità è minima. Una spiegazione inuitiva è che ci sono “molti più punti” vicino alla circonferenza di quanti ce ne siano vicino al centro di un cerchio, pertanto è più probabile che un punto scelto uniformemente nel cerchio sia vicino alla circonferenza.

Osservazione 6.45. Segue dalla Proposizione 6.40 che, se un vettore aleatorio è assolutamente continuo, allora le sue componenti sono variabili aleatorie reali assolutamente continue. Il viceversa è vero se le componenti sono indipendenti, per la Proposizione 6.41, ma è *falso* in generale. Ad esempio, se X è una variabile aleatoria reale qualsiasi, anche assolutamente continua, il vettore (X, X) non può essere assolutamente continuo. Procedendo per assurdo, supponiamo che (X, X) abbia una densità $f_{X,X}$, e sia

$$C := \{(x, y) \in \mathbb{R}^2 : y = x\},$$

la bisettrice del primo e terzo quadrante in \mathbb{R}^2 . Dato che $\{(X, X) \in C\}$ è l'evento certo, si ha

$$1 = P((X, X) \in C) = \int_C f_{X,X}(x, y) dx dy,$$

avendo applicato la relazione (6.51). D'altro canto, C è un sottospazio affine di dimensione 1, quindi ha misura 2-dimensionale nulla e di conseguenza

$$\int_C f_{X,X}(x, y) dx dy = \int_{\mathbb{R}^2} f_{X,X}(x, y) \mathbb{1}_C(x, y) dx dy = 0,$$

ottenendo una contraddizione. □

6.4.3 Calcoli con densità*

Determiniamo la distribuzione della somma di variabili aleatorie reali di cui sia nota la densità congiunta. Ricordando la relazione (6.55), si ottiene come corollario immediato la Proposizione 6.25, enunciata a suo tempo senza dimostrazione.

Proposizione 6.46. *Sia (X, Y) un vettore aleatorio bidimensionale assolutamente continuo, con densità $f_{X,Y}$. Allora la somma delle componenti $X + Y$ è una variabile aleatoria reale assolutamente continua, con densità*

$$f_{X+Y}(z) = \int_{-\infty}^{+\infty} f_{X,Y}(x, z-x) dx = \int_{-\infty}^{+\infty} f_{X,Y}(z-x, x) dx, \quad (6.59)$$

purché le funzioni $x \mapsto f_{X,Y}(x, z-x)$ e $x \mapsto f_{X,Y}(z-x, x)$ siano integrabili su \mathbb{R} .

Dimostrazione. Calcoliamo la funzione di ripartizione di $X + Y$: introducendo per $z \in \mathbb{R}$ il sottoinsieme $D_z := \{(x, y) \in \mathbb{R}^2 : x + y \leq z\}$, possiamo scrivere

$$\begin{aligned} F_{X+Y}(z) &= P(X + Y \leq z) = P((X, Y) \in D_z) = \int_{D_z} f_{X,Y}(x, y) dx dy \\ &= \int_{-\infty}^{+\infty} \left(\int_{-\infty}^{+\infty} f_{X,Y}(x, y) \mathbb{1}_{D_z}(x, y) dy \right) dx = \int_{-\infty}^{+\infty} \left(\int_{-\infty}^{z-x} f_{X,Y}(x, y) dy \right) dx, \end{aligned}$$

avendo applicato la relazione (6.51) e il teorema di Fubini-Tonelli. Operando il cambio di variabile $y = \psi(t) := t - x$ nell'integrale interno e applicando ancora il teorema

di Fubini-Tonelli, si ottiene

$$F_{X+Y}(z) = \int_{-\infty}^{+\infty} \left(\int_{-\infty}^z f_{X,Y}(x, t-x) dt \right) dx = \int_{-\infty}^z \left(\int_{-\infty}^{+\infty} f_{X,Y}(x, t-x) dx \right) dt.$$

Abbiamo mostrato che $F_{X+Y}(z) = \int_{-\infty}^z g(t) dt$, dove $g(t)$ indica l'integrale interno nell'ultima espressione, quindi $X+Y$ è una variabile assolutamente continua, con densità $g(z)$. Con argomenti analoghi si ottiene la seconda espressione in (6.59). \square

Enunciamo ora la generalizzazione multidimensionale della Proposizione 6.21.

Proposizione 6.47 (Calcolo del valor medio). *Sia X un vettore aleatorio assolutamente continuo n -dimensionale, con densità f_X , e sia $g: \mathbb{R}^n \rightarrow \mathbb{R}$ una funzione continua, tranne al più in un insieme di misura n -dimensionale nulla.*

La variabile aleatoria reale $g(X)$ ammette valor medio finito se e solo se la funzione $|g(x)|f_X(x)$ è integrabile su \mathbb{R}^n , e in tal caso

$$E[g(X)] = \int_{\mathbb{R}^n} g(x) f_X(x) dx. \quad (6.60)$$

La formula (6.60) è alla base di numerosi calcoli con vettori aleatori.

Esempio 6.48. Riprendendo l'Esempio 6.44, sia (X, Y) un vettore aleatorio bidimensionale, con distribuzione uniforme nel cerchio unitario. Le componenti X e Y sono variabili aleatorie limitate, dunque in L^2 e pertanto è ben definita la covarianza $\text{Cov}(X, Y) = E(XY) - E(X)E(Y)$. Abbiamo già calcolato la densità di X , cf. (6.57), che è una funzione pari e pertanto (perché?) $E(X) = 0$. Analogamente $E(Y) = 0$, perché Y ha la stessa distribuzione di X . Calcoliamo ora $E(XY)$ applicando la formula (6.60) alla funzione $g(x, y) := xy$:

$$E(XY) = \frac{1}{\pi} \int_{\{x^2+y^2 \leq 1\}} (xy) dx dy = \frac{1}{\pi} \int_{-1}^1 x \left(\int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} y dy \right) dx = 0,$$

perché l'integrale interno si annulla, per simmetria. Quindi $\text{Cov}(X, Y) = 0$.

Per fare un altro esempio, calcoliamo $E(|X|R)$ dove $R := \sqrt{X^2 + Y^2}$. Dalla formula (6.60) si ottiene

$$E(|X|R) = \frac{1}{\pi} \int_{\{x^2+y^2 \leq 1\}} (|x| \sqrt{x^2+y^2}) dx dy.$$

Conviene passare a coordinate polari, ossia usare il cambio di variabili $(x, y) = \psi(r, \theta) := (r \cos \theta, r \sin \theta)$, già incontrato nell'Esempio 6.36. Dato che $|\det D\psi| = r$,

applicando la formula (6.47) e il teorema di Fubini-Tonelli si ottiene

$$\begin{aligned} E(|X|R) &= \frac{1}{\pi} \int_{\{0 \leq r \leq 1, 0 < \theta < 2\pi\}} (|r \cos \theta| r) r dr d\theta \\ &= \left(\int_0^1 r^3 dr \right) \left(\frac{1}{\pi} \int_0^{2\pi} |\cos \theta| d\theta \right) = \frac{1}{4} \frac{4}{\pi} = \frac{1}{\pi}. \end{aligned}$$

Con calcoli analoghi (esercizio) si ottiene $E(|X|) = \frac{4}{3\pi}$ e $E(R) = \frac{2}{3}$, da cui

$$\text{Cov}(|X|, R) = E(|X|R) - E(|X|)E(R) = \frac{1}{\pi} - \frac{8}{9\pi} = \frac{1}{9\pi} \neq 0.$$

In particolare, le variabili aleatorie $|X|$ e R non sono indipendenti. \square

Mostriamo quindi che l'assoluta continuità di un vettore aleatorio si preserva per diffeomorfismi, generalizzando l'Esercizio 6.1.

Proposizione 6.49 (Trasformazione di un vettore assolutamente continuo).

Sia X un vettore aleatorio n -dimensionale assolutamente continuo, sia $U \subseteq \mathbb{R}^n$ un aperto tale che $P(X \in U) = 1$ e sia $\varphi : U \rightarrow V$ un diffeomorfismo. Il vettore aleatorio n -dimensionale $Y := \varphi(X)$ è assolutamente continuo, con densità

$$f_Y(y) = \begin{cases} f_X(\varphi^{-1}(y)) |\det D\varphi^{-1}(y)| & \text{se } y \in V \\ 0 & \text{altrimenti} \end{cases}. \quad (6.61)$$

Dimostrazione. Ricordando la Definizione 6.37, dobbiamo mostrare che la probabilità $P(Y \in I)$, con $I = I_1 \times \cdots \times I_n$ prodotto di intervalli, coincide con l'integrale su I della funzione f_Y definita in (6.61). Per la formula (6.51), si ha

$$P(Y \in I) = P(X \in \varphi^{-1}(I)) = \int_U \mathbb{1}_{\varphi^{-1}(I)}(x) f_X(x) dx,$$

dove abbiamo ristretto l'integrale a U perché $P(X \in U) = 1$, e dunque possiamo supporre che $f_X(x) = 0$ per $x \notin U$. Dato che $\varphi^{-1} : V \rightarrow U$ è un diffeomorfismo, possiamo applicare la formula (6.47) con $g(x) = f_X(x) \mathbb{1}_{\varphi^{-1}(I)}(x)$ e $\psi := \varphi^{-1}$, ottenendo

$$\begin{aligned} P(Y \in I) &= \int_V f_X(\varphi^{-1}(y)) \mathbb{1}_{\varphi^{-1}(I)}(\varphi^{-1}(y)) |\det D\varphi^{-1}(y)| dy \\ &= \int_I f_X(\varphi^{-1}(y)) |\det D\varphi^{-1}(y)| \mathbb{1}_V(y) dy, \end{aligned}$$

che è quanto dovevamo dimostrare. \square

Osserviamo che la matrice jacobiana di φ^{-1} , per il teorema di differenziazione della funzione inversa, è data da

$$(D\varphi^{-1})(x) = (D\varphi(\varphi^{-1}(x)))^{-1},$$

quindi

$$\det(D\varphi^{-1})(x) = \frac{1}{\det(D\varphi(\varphi^{-1}(x)))}.$$

Consideriamo il caso speciale, ma particolarmente importante, di una trasformazione lineare-affine invertibile: $\varphi(x) = Ax + b$, dove A è una matrice quadrata $n \times n$ non singolare, ossia $\det A \neq 0$, $b \in \mathbb{R}^n$ e Ax indica l'ordinario prodotto matrice per vettore. Chiaramente $\varphi : \mathbb{R}^n \rightarrow \mathbb{R}^n$ è un diffeomorfismo, con $\varphi^{-1}(y) = A^{-1}(y - b)$ e dunque $\det D\varphi^{-1}(y) = \det(A^{-1}) = \frac{1}{\det A}$. Otteniamo dunque il seguente corollario, l'analogo multidimensionale della Proposizione 6.18.

Proposizione 6.50 (Trasformazioni lineari-affini). *Sia X un vettore aleatorio di dimensione n , assolutamente continuo con densità f_X . Sia inoltre A una matrice invertibile $n \times n$ e $b \in \mathbb{R}^n$. Il vettore aleatorio $Y := AX + b$ è assolutamente continuo, con densità*

$$f_Y(y) = \frac{1}{|\det(A)|} f_X(A^{-1}(y - b)).$$

Come applicazione, diamo una dimostrazione alternativa della formula per la densità della somma di due variabili reali assolutamente continue indipendenti.

Esempio 6.51. Siano X, Y variabili aleatorie reali indipendenti assolutamente continue, con densità f_X e f_Y . Per la Proposizione 6.41, il vettore aleatorio bidimensionale (X, Y) è assolutamente continuo, con densità

$$f_{X,Y}(x, y) = f_X(x) f_Y(y).$$

Sia ora $\varphi : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ la trasformazione lineare data da

$$\varphi\left(\begin{pmatrix} x \\ y \end{pmatrix}\right) := \begin{pmatrix} x+y \\ y \end{pmatrix} = A \begin{pmatrix} x \\ y \end{pmatrix}, \quad \text{con } A := \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}.$$

Si noti che $\det(A) = 1$ e $\varphi^{-1}(z, w) = (z - w, w)$. Perciò, posto $Z := X + Y$ e osservato che $\varphi(X, Y) = (Z, Y)$, per la Proposizione 6.50 si ha che (Z, Y) è un vettore aleatorio assolutamente continuo con densità

$$f_{Z,Y}(z, y) = \frac{1}{|\det A|} f_{X,Y}(\varphi^{-1}(z, y)) = f_X(z - y) f_Y(y).$$

La densità marginale di $Z = X + Y$ si ottiene allora applicando la Proposizione 6.41:

$$f_Z(z) = \int_{-\infty}^{+\infty} f_{Z,Y}(z, y) dy = \int_{-\infty}^{+\infty} f_X(z - y) f_Y(y) dy = \int_{-\infty}^{+\infty} f_X(x) f_Y(z - x) dx,$$

dove l'ultima uguaglianza segue dal semplice cambio di variabili $y = z - x$. □

Esercizi *

Esercizio 6.4 (*). Sia (X, Y) un vettore aleatorio bidimensionale assolutamente continuo. Supponiamo che esistano due funzioni integrabili $a : \mathbb{R} \rightarrow \mathbb{R}^+$ e $b : \mathbb{R} \rightarrow \mathbb{R}^+$ tali che

$$f_{X,Y}(x, y) = a(x)b(y), \quad \forall (x, y) \in \mathbb{R}^2.$$

Si deduca che X e Y sono variabili aleatorie indipendenti.

Esercizio 6.5 (*). Sia (X, Y) un vettore aleatorio bidimensionale assolutamente continuo. Supponiamo che esista un punto $(x_0, y_0) \in \mathbb{R}^2$ tale che

$$f_{X,Y}(x_0, y_0) \neq f_X(x_0)f_Y(y_0).$$

Assumendo che le funzioni $f_{X,Y}$, f_X e f_Y siano continue rispettivamente nei punti (x_0, y_0) , x_0 e y_0 , si deduca che le variabili aleatorie X e Y *non* sono indipendenti.

6.5 Esempi e applicazioni

Presentiamo ora alcuni esempi e modelli che coinvolgono variabili aleatorie assolutamente continue. I Paragrafi 6.5.1, 6.5.2 e 6.5.3 trattano di variabili aleatorie reali, mentre per i Paragrafi successivi 6.5.4 e 6.5.5 è richiesta la conoscenza dei vettori aleatori multidimensionali.

6.5.1 Le variabili aleatorie chi-quadro

Data una variabile aleatoria $Z \sim N(0, 1)$, determiniamo la distribuzione della variabile aleatoria $X := Z^2$. Chiaramente $F_X(x) = 0$ se $x < 0$, mentre per $x \geq 0$

$$\begin{aligned} F_X(x) &= P(X \leq x) = P(Z^2 \leq x) = P(-\sqrt{x} \leq Z \leq \sqrt{x}) \\ &= P(Z \leq \sqrt{x}) - P(Z < -\sqrt{x}) = F_Z(\sqrt{x}) - F_Z(-\sqrt{x}), \end{aligned}$$

dove si è usato il fatto che $P(Z < a) = P(Z \leq a) = F_Z(a)$, perché $P(Z = a) = 0$. La densità f_Z di Z è continua (si ricordi la relazione (6.41)), quindi la funzione di ripartizione F_Z è di classe \mathcal{C}^1 su tutto \mathbb{R} , per il Teorema 6.9 (i). Di conseguenza, F_X è una funzione continua su \mathbb{R} e di classe \mathcal{C}^1 su $\mathbb{R} \setminus \{0\}$, dunque \mathcal{C}^1 a tratti. Per la Proposizione 6.16, la variabile aleatoria X è assolutamente continua, con densità

$$\begin{aligned} f_X(x) &= F'_X(x) = \frac{1}{2\sqrt{x}} [F'_Z(\sqrt{x}) + F'_Z(-\sqrt{x})] \\ &= \frac{1}{2\sqrt{x}} [f_Z(\sqrt{x}) + f_Z(-\sqrt{x})] \mathbb{1}_{(0, +\infty)}(x) = \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{x}} e^{-\frac{1}{2}x} \mathbb{1}_{(0, +\infty)}(x). \end{aligned}$$

Ricordando le relazioni (6.34) e (6.32), abbiamo mostrato che $X \sim \text{Gamma}(\frac{1}{2}, \frac{1}{2})$.

Consideriamo ora n variabili aleatorie indipendenti Z_1, Z_2, \dots, Z_n , ciascuna con distribuzione $N(0, 1)$. Le variabili aleatorie $Z_1^2, Z_2^2, \dots, Z_n^2$ sono indipendenti per la Proposizione 3.40, in quanto funzioni di variabili aleatorie indipendenti, e $Z_i^2 \sim \Gamma(\frac{1}{2}, \frac{1}{2})$ per quanto appena mostrato. Dalla Proposizione 6.29 segue dunque che

$$X := Z_1^2 + Z_2^2 + \dots + Z_n^2 \sim \text{Gamma}\left(\frac{n}{2}, \frac{1}{2}\right). \quad (6.62)$$

Le variabili aleatorie con distribuzione $\text{Gamma}(\frac{n}{2}, \frac{1}{2})$ emergono in diverse applicazioni della probabilità, in particolare in statistica matematica (come vedremo nel Paragrafo 6.6 e nel Capitolo 8). Esse sono dette χ^2 (“chi-quadro”, o “chi-quadrato”) a n gradi di libertà e sono indicate col simbolo $\chi^2(n)$. Più esplicitamente, si ha $X \sim \chi^2(n)$ se e solo se $X \sim \text{Gamma}(\frac{n}{2}, \frac{1}{2})$, ossia

$$f_X(x) = \frac{1}{2^{\frac{n}{2}} \Gamma(\frac{n}{2})} x^{\frac{n}{2}-1} e^{-\frac{1}{2}x} \mathbb{1}_{(0, \infty)}(x).$$

(Osserviamo che il valore di $\Gamma(\frac{n}{2})$ è noto esplicitamente per ogni $n \in \mathbb{N}$, grazie alle relazioni (6.31) e (6.33).) Nella Figura 6.3 a pagina 265 è rappresentata la densità di una $\text{Gamma}(\frac{1}{2}, 1)$, che coincide con quella di una $\chi^2(1) = \text{Gamma}(\frac{1}{2}, \frac{1}{2})$ a meno di un semplice fattore di scala (si ricordi la relazione (6.35)).

In definitiva, possiamo riformulare quanto mostrato in questo paragrafo con la seguente affermazione: *la somma dei quadrati di n variabili aleatorie $N(0, 1)$ indipendenti è una variabile aleatoria $\chi^2(n)$.*

6.5.2 Statistiche d'ordine e variabili aleatorie Beta

Siano X_1, X_2, \dots, X_n variabili aleatorie reali, definite tutte sullo stesso spazio di probabilità (Ω, \mathcal{A}, P) . Per ogni $\omega \in \Omega$, consideriamo l'ordinamento crescente dell'insieme di punti $\{X_1(\omega), X_2(\omega), \dots, X_n(\omega)\}$:

$$\{X_1(\omega), X_2(\omega), \dots, X_n(\omega)\} = \{X^{(1)}(\omega), X^{(2)}(\omega), \dots, X^{(n)}(\omega)\},$$

con $X^{(1)}(\omega) \leq X^{(2)}(\omega) \leq \dots \leq X^{(n)}(\omega)$. Si può mostrare che la funzione che “mette in ordine” una n -upla di numeri reali è misurabile, pertanto le $X^{(k)}$, come funzioni di ω , sono variabili aleatorie, e vengono chiamate *statistiche d'ordine*. Chiaramente,

$$X^{(1)} = \min(X_1, X_2, \dots, X_n), \quad X^{(n)} = \max(X_1, X_2, \dots, X_n).$$

Nel caso in cui le variabili aleatorie X_1, X_2, \dots, X_n siano indipendenti, la distribuzione di $X^{(1)}$ e $X^{(n)}$ si può determinare usando la Proposizione 3.93. La distribuzione delle altre statistiche d'ordine è più difficile, in generale, da ottenere. Una espressione esplicita può essere determinata nel caso di variabili aleatorie uniformi.

Proposizione 6.52. *Siano X_1, X_2, \dots, X_n variabili aleatorie indipendenti, con distribuzione $U(0, 1)$. Allora, per $k = 1, 2, \dots, n$, la statistica d'ordine $X^{(k)}$ è una variabile aleatoria assolutamente continua con densità*

$$f_k(x) := \binom{n}{k} k x^{k-1} (1-x)^{n-k} \mathbb{1}_{(0,1)}(x).$$

Dimostrazione. Dimosteremo la seguente affermazione equivalente: per ogni $x \in \mathbb{R}$, la funzione di ripartizione di $X^{(k)}$, indicata con F_k , è data da

$$F_k(x) = \int_{-\infty}^x f_k(t) dt. \quad (6.63)$$

Si noti che per $x < 0$ la precedente identità è ovvia. È sufficiente dimostrare (6.63) per $x \in [0, 1]$. Infatti, entrambi i membri di (6.63) sono costanti per $x \geq 1$. Dimostriamo dunque, per induzione su k , l'identità (6.63) per $x \in [0, 1]$.

Caso $k = 1$. Si noti che $F_1(x) = F_W(x)$, dove $W := \min(X_1, X_2, \dots, X_n)$. Usando la Proposizione 3.93, per $x \in [0, 1]$

$$F_1(x) = 1 - (1 - F_{X_1}(x))^n = 1 - (1 - x)^n.$$

Inoltre

$$\int_{-\infty}^x f_1(t) dt = n \int_0^x (1-t)^{n-1} dt = 1 - (1-x)^n,$$

e quindi in passo base dell'induzione è dimostrato.

Passo induttivo. Si ha, per $x \in [0, 1]$

$$\begin{aligned} 1 - F_{k+1}(x) &= P(X^{(k+1)} > x) = P(X^{(k+1)} > x, X^{(k)} > x) + P(X^{(k+1)} > x, X^{(k)} \leq x) \\ &= P(X^{(k)} > x) + P(X^{(k+1)} > x, X^{(k)} \leq x), \end{aligned}$$

avendo usato l'inclusione di eventi $\{X^{(k)} > x\} \subseteq \{X^{(k+1)} > x\}$, pertanto

$$F_{k+1}(x) = F_k(x) - P(X^{(k+1)} > x, X^{(k)} \leq x). \quad (6.64)$$

L'evento $\{X^{(k+1)} > x, X^{(k)} \leq x\}$ corrisponde all'affermazione che tra le variabili aleatorie X_1, X_2, \dots, X_n ve ne sono esattamente k che assumono valore minore o uguale a x , ossia che si verifichino esattamente k tra gli eventi

$$C_1 := \{X_1 \leq x\}, \quad \dots \quad C_n := \{X_n \leq x\}.$$

Tali eventi sono indipendenti e hanno la stessa probabilità $P(C_i) = P(X_i \leq x) = x$, ossia costituiscono uno schema di n prove ripetute e indipendenti con probabilità di successo $p = x$ (si ricordi la Definizione 1.66). La probabilità di avere k successi, grazie alla relazione (1.56), è dunque data da

$$P(X^{(k+1)} > x, X^{(k)} \leq x) = \binom{n}{k} x^k (1-x)^{n-k}.$$

Ricordando (6.64) e l'ipotesi induttiva, si ha

$$F_{k+1}(x) = \binom{n}{k} \int_0^x k t^{k-1} (1-t)^{n-k} dt - \binom{n}{k} x^k (1-x)^{n-k}.$$

Con una semplice integrazione per parti, i cui dettagli sono lasciati al lettore, possiamo riscrivere il membro destro di questa relazione come

$$\begin{aligned} \binom{n}{k} (n-k) \int_0^x t^k (1-t)^{n-k-1} dt &= \frac{n!}{k!(n-(k+1))!} \int_0^x t^k (1-t)^{n-(k+1)} dt \\ &= \int_{-\infty}^x f_{k+1}(t) dt, \end{aligned}$$

e la dimostrazione è completa. \square

Le densità che compaiono nella Proposizione 6.52 motivano l'introduzione di una nuova classe di variabili aleatorie notevoli. Si dice che una variabile aleatoria reale X ha distribuzione *Beta di parametri* $a, b \in (0, \infty)$, e si scrive $X \sim \text{Beta}(a, b)$, se X è assolutamente continua con densità

$$f_X(x) = \frac{1}{\beta(a, b)} x^{a-1} (1-x)^{b-1} \mathbb{1}_{(0,1)}(x), \quad (6.65)$$

dove la costante di normalizzazione

$$\beta(a, b) := \int_0^1 x^{a-1} (1-x)^{b-1} dx,$$

è già stata introdotta in (6.37). Ricordando la relazione (6.38), che possiamo riscrivere come

$$\beta(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}, \quad (6.66)$$

e il fatto che $\Gamma(m) = (m-1)!$ per $m \in \mathbb{N}$, per $a, b \in \mathbb{N}$ si ha

$$\frac{1}{\beta(a, b)} = \frac{(a+b-1)!}{(a-1)!(b-1)!} = a \binom{a+b-1}{a}.$$

In particolare, possiamo riformulare la Proposizione 6.52 affermando che

$$X^{(k)} \sim \text{Beta}(k, n-k+1).$$

Dalla relazione (6.66) e dal fatto che $\Gamma(\alpha+1) = \alpha\Gamma(\alpha)$ per ogni $\alpha > 0$, cf. (6.30), segue facilmente che

$$\beta(a+1, b) = \frac{a}{a+b} \beta(a, b).$$

Usando tale identità, è facile verificare che, se $X \sim \text{Beta}(a, b)$,

$$E(X) = \frac{a}{a+b}, \quad \text{Var}(X) = \frac{ab}{(a+b)^2(a+b+1)}. \quad (6.67)$$

6.5.3 Il processo di Poisson (parte I)

In numerose applicazioni è utile avere a disposizione modelli probabilistici per i tempi in cui accadono vicende di particolare rilevanza, ma che l'evidenza empirica fa ritenere "imprevedibili". Esempi di questo genere sono gli istanti in cui:

- un dato vulcano erutta;
- una certa zona è colpita da un terremoto;
- si verifica un "crack" nel mercato finanziario;
- una data compagnia di assicurazioni ha un cliente che reclama un rimborso.

Per descrivere fenomeni di questo tipo, consideriamo una famiglia $(T_k)_{k \in \mathbb{N}}$ di variabili aleatorie reali positive, in cui T_k rappresenta l'istante in cui il fenomeno in esame avviene per la k -esima volta. Supporremo quindi che $0 < T_1 < \dots < T_k < +\infty$. Il sottoinsieme aleatorio $\{T_1, T_2, \dots\} \subseteq [0, +\infty)$ è detto *processo di punto*.

Il quadro descritto finora è molto generale. Per caratterizzare un modello preciso, occorre specificare la distribuzione dei tempi aleatori $(T_k)_{k \in \mathbb{N}}$. Questo dipende da cosa si intenda per "imprevedibilità". Noi assumeremo quanto segue:

- la conoscenza dei primi k tempi T_1, T_2, \dots, T_k non fornisce alcuna informazione sull'intervallo di tempo $(T_{k+1} - T_k)$ che intercorre tra la k -esima e la $(k+1)$ -esima occorrenza del fenomeno;
- per ogni $h > 0$, il verificarsi dell'evento $\{T_{k+1} > T_k + h\}$ (ossia il fatto che nell'intervallo $(T_k, T_k + h]$ non vi sia stata alcuna occorrenza del fenomeno) non fornisce alcuna informazione su $T_{k+1} - T_k - h$, cioè sul tempo che ancora si deve attendere per la prossima occorrenza.

Queste specifiche, assieme alla Proposizione 6.31 sulla *assenza di memoria* delle variabili aleatorie esponenziali (si veda anche l'Esercizio 6.3), giustificano l'ipotesi che gli *intertempi* $(T_{k+1} - T_k)_{k \in \mathbb{N}}$ siano variabili aleatorie indipendenti e con distribuzione esponenziale. Un'ulteriore ipotesi di *omogeneità temporale* induce a supporre che tali variabili aleatorie esponenziali abbiano tutte lo stesso parametro. Arriviamo dunque alla seguente definizione.

Definizione 6.53 (Processo di Poisson). Sia $(X_n)_{n \in \mathbb{N}}$ una successione di variabili aleatorie indipendenti e con la stessa distribuzione $\text{Exp}(\lambda)$, dove $\lambda > 0$ è un parametro fissato. Definiamo quindi:

- per ogni $k \in \mathbb{N}$, la variabile aleatoria $T_k := X_1 + \dots + X_k$;
- per ogni $t \in [0, +\infty)$, la variabile aleatoria N_t che conta il numero di punti dell'insieme $\{T_1, T_2, \dots\}$ che cadono nell'intervallo $[0, t]$, ossia

$$N_t := |\{k \in \mathbb{N} : T_k \leq t\}|. \quad (6.68)$$

La famiglia $(N_t)_{t \geq 0}$ è detta *processo di Poisson con intensità λ* .

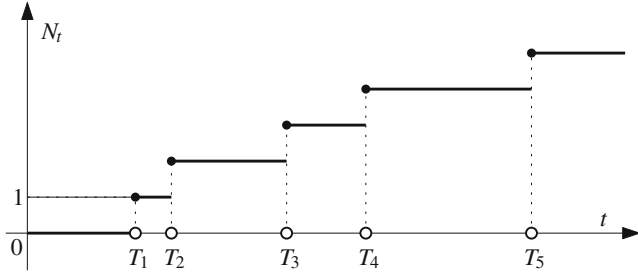


Figura 6.5 Rappresentazione grafica di una traiettoria $t \mapsto N_t$ del processo di Poisson

Si noti che N_t è una variabile aleatoria, definita sullo stesso spazio di probabilità (Ω, \mathcal{A}, P) in cui sono definite le $(T_k)_{k \in \mathbb{N}}$, a valori in $\mathbb{N}_0 \cup \{+\infty\}$: per ogni $\omega \in \Omega$

$$N_t(\omega) := |\{k \in \mathbb{N} : T_k(\omega) \leq t\}| = \max \{k \in \mathbb{N} : T_k(\omega) \leq t\}, \quad (6.69)$$

dove poniamo $\max \emptyset := 0$ e $\max \mathbb{N} := +\infty$. Si veda la Figura 6.5.

Pur nella sua relativa semplicità, il processo di Poisson è uno dei modelli più importanti del calcolo delle probabilità. Il primo risultato che presentiamo mostra che la variabile aleatoria N_t ha una distribuzione di Poisson, per ogni $t > 0$; segue in particolare che $P(N_t < +\infty) = 1$, un fatto non ovvio a priori.

Proposizione 6.54. *Se $(N_t)_{t \in [0, \infty)}$ è un processo di Poisson di intensità λ , si ha*

$$N_t \sim \text{Pois}(\lambda t), \quad \forall t \in (0, +\infty).$$

Dimostrazione. Dato che $\text{Exp}(\lambda) = \text{Gamma}(\lambda, 1)$, si mostra facilmente per induzione che $T_n \sim \text{Gamma}(n, \lambda)$, per ogni $n \geq 1$, grazie alla Proposizione 6.29. Notando che

$$\{N_t = n\} = \{T_n \leq t\} \setminus \{T_{n+1} \leq t\},$$

e che $\{T_{n+1} \leq t\} \subseteq \{T_n \leq t\}$, si ha

$$\begin{aligned} P(N_t = n) &= P(T_n \leq t) - P(T_{n+1} \leq t) \\ &= \frac{\lambda^n}{(n-1)!} \int_0^t x^{n-1} e^{-\lambda x} dx - \frac{\lambda^{n+1}}{n!} \int_0^t x^n e^{-\lambda x} dx. \end{aligned} \quad (6.70)$$

Integrando per parti

$$\int_0^t x^n e^{-\lambda x} dx = -\frac{t^n}{\lambda} e^{-\lambda t} + \frac{n}{\lambda} \int_0^t x^{n-1} e^{-\lambda x} dx,$$

e inserendo questa identità in (6.70) si trova subito

$$P(N_t = n) = e^{-\lambda t} \frac{(\lambda t)^n}{n!} =: p_{\text{Pois}(\lambda t)}(n), \quad \forall n \in \mathbb{N}.$$

Resta il caso $n = 0$, che è facile: infatti $\{N_t = 0\} = \{T_1 > t\}$ e pertanto

$$P(N_t = 0) = P(T_1 > t) = \int_t^{+\infty} \lambda e^{-\lambda x} dx = e^{-\lambda t} = p_{\text{Pois}(\lambda t)}(0).$$

Questo mostra che $N_t \sim \text{Pois}(\lambda t)$, come cercato. \square

Vogliamo ora rafforzare la Proposizione 6.54, determinando la distribuzione congiunta di un numero finito di variabili aleatorie $N_{t_0}, N_{t_1}, \dots, N_{t_k}$. Risulta conveniente considerarne gli *incrementi* $N_{t_1} - N_{t_0}, \dots, N_{t_k} - N_{t_{k-1}}$, che hanno un'interpretazione importante. Infatti, segue dalla relazione (6.68) che per ogni $0 \leq s < t < \infty$

$$N_t - N_s = |\{n \in \mathbb{N} : T_n \in (s, t]\}|, \quad (6.71)$$

cioè la variabile aleatoria $N_t - N_s$ conta il numero di tempi aleatori T_n che cadono nell'intervallo $(s, t]$. Il risultato seguente è di importanza fondamentale.

Teorema 6.55. *Sia $(N_t)_{t \in [0, \infty)}$ un processo di Poisson di intensità λ e siano $0 = t_0 < t_1 < \dots < t_k < \infty$ istanti temporali fissati. Le variabili aleatorie $(N_{t_i} - N_{t_{i-1}})_{i=1}^k$ sono indipendenti, e hanno distribuzioni marginali*

$$N_{t_i} - N_{t_{i-1}} \sim \text{Pois}(\lambda(t_i - t_{i-1})).$$

La dimostrazione richiede la conoscenza dei vettori aleatori assolutamente continui ed è posposta al Paragrafo 6.5.4.

Il Teorema 6.55 mostra, in particolare, che un processo di Poisson $(N_t)_{t \in [0, \infty)}$ gode delle seguenti proprietà:

- (i) ha *incrementi indipendenti*, cioè le variabili aleatorie $(N_{t_i} - N_{t_{i-1}})_{i=1}^k$ sono indipendenti, per ogni scelta degli istanti $0 = t_0 < t_1 < \dots < t_k < \infty$;
- (ii) ha *incrementi stazionari*, cioè la distribuzione della variabile aleatoria $N_t - N_s$ dipende solo dalla differenza degli istanti $(t - s)$.

In generale, le famiglie di variabili aleatorie reali $(N_t)_{t \in [0, \infty)}$ che soddisfano tali proprietà sono dette *processi di Lévy*. Si tratta di una classe molto ricca e studiata di modelli probabilistici, di cui il processo di Poisson è l'esempio più semplice.

Altre proprietà, più “specifiche”, del processo di Poisson seguono dalla definizione (6.69) e dal fatto che $P(N_t < +\infty) = 1$ per ogni $t > 0$: per ogni $\omega \in \Omega$, a meno di un insieme di probabilità nulla, si ha che $(N_t)_{t \in [0, \infty)}$ (si veda la Figura 6.5)

- (iii) è un *processo contatore*, cioè $N_0(\omega) = 0$ e la funzione $t \mapsto N_t(\omega)$ è crescente, continua da destra, assume valori in $\mathbb{N}_0 = \{0, 1, 2, \dots\}$ e fa salti di 1.

Ebbene, è possibile mostrare che le proprietà (i), (ii) e (iii) *caratterizzano il processo di Poisson*: ogni processo contatore $(N_t)_{t \in [0, \infty)}$ con incrementi indipendenti e stazionari è necessariamente un processo di Poisson!

Questo risultato (che non dimostriamo) mostra che la distribuzione di Poisson emerge in modo “inevitabile”, quale distribuzione del numero di occorrenze di un fenomeno “imprevedibile” in un dato intervallo di tempo.

Osservazione 6.56. Dal Teorema 6.55 segue che, per intervalli disgiunti $I = (a, b]$ e $J = (c, d]$, il numero di tempi aleatori T_n che cadono rispettivamente in I e J , ossia

$$N_I := N_b - N_a, \quad N_J := N_d - N_c,$$

hanno la proprietà di essere variabili aleatorie *indipendenti* con distribuzione di Poisson, con parametri proporzionali alle lunghezze di I e J . Non è difficile mostrare che questa proprietà vale, più in generale, nel caso in cui gli insiemi disgiunti I e J sono unioni finite di intervalli (si veda l'Esercizio 6.6). Con tecniche più avanzate, al di là degli scopi di questo libro, si può estendere la proprietà a ogni coppia di insiemi limitati e disgiunti $I, J \subseteq [0, +\infty)$ nella σ -algebra di Borel. \square

6.5.4 Il processo di Poisson (parte II) *

Completiamo l'analisi del processo di Poisson, introdotto nel Paragrafo 6.5.3, dimostrando il Teorema 6.55. Usiamo le stesse notazioni della Definizione 6.53.

Dimostrazione (del Teorema 6.55). Il caso $k = 1$ è esattamente il contenuto della Proposizione 6.54. Nella dimostrazione ci limitiamo per semplicità di notazioni al caso $k = 2$, ma il caso generale è del tutto analogo. Procediamo in quattro passi.

Passo 1. Siano $0 < s < t < \infty$ e $h, \ell \in \mathbb{N}_0$ arbitrari, ma fissati. Dobbiamo mostrare che le variabili aleatorie N_s e $N_t - N_s$ sono indipendenti e hanno distribuzioni marginali $\text{Pois}(\lambda s)$ e $\text{Pois}(\lambda(t-s))$ rispettivamente, ossia

$$P(N_s = h, N_t - N_s = \ell) = p_{\text{Pois}(\lambda s)}(h) p_{\text{Pois}(\lambda(t-s))}(\ell).$$

Già sappiamo che $P(N_s = h) = p_{\text{Pois}(\lambda s)}(h)$, per la Proposizione 6.54, quindi ci basta mostrare che

$$P(N_t - N_s = \ell | N_s = h) = p_{\text{Pois}(\lambda(t-s))}(\ell). \quad (6.72)$$

L'euristica è la seguente: condizionalmente al verificarsi dell'evento $\{N_s = h\}$, l'insieme aleatorio $\{T_1, T_2, \dots\} \cap (s, +\infty) = \{T_{h+1}, T_{h+2}, \dots\}$, dopo aver traslato l'origine dei tempi in s , ha la stessa distribuzione dell'insieme originale $\{T_1, T_2, \dots\}$. Di conseguenza, il numero di tempi T_n che cadono in $(s, t]$ ha la stessa distribuzione del numero di tempi T_n che cadono in $(0, t-s]$, ossia di $N_{t-s} \sim \text{Pois}(\lambda(t-s))$, per la Proposizione 6.54. In questo modo si ottiene (6.72).

Passo 2. Intuitivamente, se si verifica l'evento $\{N_s = h\}$, si ha $T_h \leq s$ e $T_{h+1} > s$, quindi nella relazione (6.71) possiamo limitarci a considerare i valori di $n \geq h+1$. Più formalmente, la variabile aleatoria

$$M_{s,t} := |\{n \in \mathbb{N}, n \geq h+1 : T_n \in (s, t]\}| \quad (6.73)$$

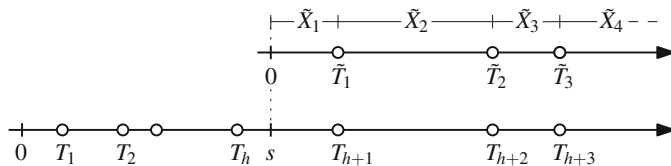


Figura 6.6 Rappresentazione grafica delle variabili aleatorie $(\tilde{T}_k)_{k \in \mathbb{N}}$ e $(\tilde{X}_k)_{k \in \mathbb{N}}$, definite rispettivamente in (6.75) e (6.78), in funzione delle variabili aleatorie $(T_k)_{k \in \mathbb{N}}$

coincide con la variabile aleatoria $(N_t - N_s)$ sull'evento $\{N_s = h\}$,[†] quindi la relazione (6.72) si può riscrivere come

$$P(M_{s,t} = \ell | N_s = h) = p_{\text{Pois}(\lambda(t-s))}(\ell). \quad (6.74)$$

Definiamo ora le variabili aleatorie

$$\tilde{T}_k := T_{h+k} - s, \quad \forall k \in \mathbb{N}, \quad (6.75)$$

che rappresentano i punti dell'insieme $\{T_1, T_2, \dots\} \cap (s, +\infty) = \{T_{h+1}, T_{h+2}, \dots\}$, tralata l'origine in s (si veda la Figura 6.6). La relazione (6.73) si riscrive come

$$M_{s,t} = |\{k \in \mathbb{N} : T_{h+k} \in (s, t]\}| = |\{k \in \mathbb{N} : \tilde{T}_k \leq t - s\}|. \quad (6.76)$$

L'osservazione cruciale (vedi sotto) è che, *rispetto alla probabilità condizionale*[‡]

$$\tilde{P}(\cdot) := P(\cdot | N_s = h), \quad (6.77)$$

le variabili aleatorie $(\tilde{T}_k)_{k \in \mathbb{N}}$ hanno la distribuzione “giusta” dei tempi aleatori di un processo di Poisson, cioè si può scrivere $\tilde{T}_k = \tilde{X}_1 + \dots + \tilde{X}_k$, dove $(\tilde{X}_n)_{n \in \mathbb{N}}$ sono variabili aleatorie indipendenti e con la stessa distribuzione $\text{Exp}(\lambda)$. Possiamo allora applicare la Definizione 6.53: per la relazione (6.68), il membro destro in (6.76) coincide con \tilde{N}_{t-s} , dove $(\tilde{N}_u)_{u \in [0, \infty)}$ è un processo di Poisson di intensità λ (rispetto alla probabilità condizionale \tilde{P}). Applicando la Proposizione 6.54, otteniamo dunque

$$\tilde{P}(M_{s,t} = \ell) = \tilde{P}(\tilde{N}_{t-s} = \ell) = p_{\text{Pois}(\lambda(t-s))}(\ell),$$

che coincide proprio con la relazione (6.74) da dimostrare.

[†] Cioè vale l'uguaglianza di eventi $\{N_t - N_s = \ell, N_s = h\} = \{M_{s,t} = \ell, N_s = h\}$, per ogni $\ell \in \mathbb{N}_0$.

[‡] Se (Ω, \mathcal{A}, P) è uno spazio di probabilità, non è difficile mostrare che, per ogni evento fissato C , la probabilità condizionale $P(\cdot | C)$ è una probabilità sullo spazio misurabile (Ω, \mathcal{A}) , come per gli spazi di probabilità discreti (si ricordi la Proposizione 1.44).

Passo 3. Definiamo infine

$$\tilde{X}_1 := \tilde{T}_1, \quad \tilde{X}_k := \tilde{T}_k - \tilde{T}_{k-1}, \quad \forall k \geq 2, \quad (6.78)$$

così che, per costruzione, si ha $\tilde{T}_k = \tilde{X}_1 + \dots + \tilde{X}_k$ per ogni $n \in \mathbb{N}$. Resta solo da mostrare che le variabili aleatorie $(\tilde{X}_n)_{n \in \mathbb{N}}$, rispetto alla probabilità condizionale $\tilde{\mathbb{P}}$, sono indipendenti e hanno la stessa distribuzione $\text{Exp}(\lambda)$. Questo significa che, per ogni $k \in \mathbb{N}$ e per ogni scelta degli insiemi misurabili $I_1, \dots, I_k \subseteq (0, \infty)$, si ha

$$\tilde{\mathbb{P}}(\tilde{X}_1 \in I_1, \dots, \tilde{X}_k \in I_k) = \left(\int_{I_1} \lambda e^{-\lambda x_1} dx_1 \right) \dots \left(\int_{I_k} \lambda e^{-\lambda x_k} dx_k \right) \quad (6.79)$$

Mettendo insieme le relazioni (6.75) e (6.78), e ricordando la Definizione 6.53 di processo di Poisson, otteniamo la seguente rappresentazione:

$$\tilde{X}_1 := T_{h+1} - s, \quad \tilde{X}_2 := X_{h+2}, \quad \tilde{X}_3 = X_{h+3}, \quad \dots$$

Ricordando (6.77) e il fatto che $\{N_s = h\} = \{T_h \leq s, T_{h+1} > s\}$ (dove definiamo $T_0 := 0$, nel caso in cui $h = 0$), possiamo riscrivere (6.79) come

$$\begin{aligned} & \mathbb{P}(T_h \leq s, (T_{h+1} - s) \in I_1, X_{h+2} \in I_2, \dots, X_{h+k} \in I_k) \\ &= \mathbb{P}(N_s = h) \cdot \left(\int_{I_1} \lambda e^{-\lambda x_1} dx_1 \right) \dots \left(\int_{I_k} \lambda e^{-\lambda x_k} dx_k \right). \end{aligned} \quad (6.80)$$

Ricordiamo che, per definizione di processo di Poisson, $(X_n)_{n \in \mathbb{N}}$ sono variabili aleatorie indipendenti con distribuzione $\text{Exp}(\lambda)$. Dato che

$$T_h = X_1 + \dots + X_h, \quad T_{h+1} = X_1 + \dots + X_{h+1}, \quad (6.81)$$

le variabili aleatorie X_{h+2}, \dots, X_{h+k} sono indipendenti da T_h, T_{h+1} , quindi la relazione (6.80) si fattorizza e si riduce a

$$\mathbb{P}(T_h \leq s, (T_{h+1} - s) \in I_1) = \mathbb{P}(N_s = h) \int_{I_1} \lambda e^{-\lambda x_1} dx_1, \quad (6.82)$$

per ogni sottoinsieme misurabile $I_1 \subseteq (0, \infty)$.

Passo 4. Calcoliamo infine il primo membro della relazione (6.82). Il caso $h = 0$ va considerato a parte, ma è più facile, pertanto supporremo che $h \geq 1$. Segue dalla Proposizione 6.29 che $T_h = X_1 + \dots + X_h \sim \text{Gamma}(h, \lambda)$. Dato che $X_{h+1} \sim \text{Exp}(\lambda)$ è indipendente da T_h , il vettore aleatorio bidimensionale (T_h, X_{h+1}) è assolutamente continuo, per la Proposizione 6.41, con densità

$$\begin{aligned} f_{(T_h, X_{h+1})}(z, x) &= f_{\text{Gamma}(h, \lambda)}(z) f_{\text{Exp}(\lambda)}(x) \\ &= \frac{\lambda^{h+1}}{(h-1)!} z^{h-1} e^{-\lambda(z+x)} \mathbb{1}_{(0, \infty)}(z) \mathbb{1}_{(0, \infty)}(x). \end{aligned}$$

Determiniamo ora la distribuzione del vettore aleatorio $(T_h, T_{h+1} - s)$, che compare in (6.82). Dato che $T_{h+1} = T_h + X_{h+1}$, possiamo scrivere

$$\begin{pmatrix} T_h \\ T_{h+1} - s \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} T_h \\ X_{h+1} \end{pmatrix} + \begin{pmatrix} 0 \\ -s \end{pmatrix} =: A \begin{pmatrix} T_h \\ X_{h+1} \end{pmatrix} + b.$$

Segue dalla Proposizione 6.50 che (T_h, \tilde{X}_1) è assolutamente continuo, con densità

$$\begin{aligned} f_{(T_h, T_{h+1} - s)}(z, x) &= f_{(T_h, X_{h+1})}(A^{-1}\left(\begin{pmatrix} z \\ x \end{pmatrix} - b\right)) = f_{(T_h, X_{h+1})}(z, x + s - z) \\ &= \frac{\lambda^{h+1}}{(h-1)!} z^{h-1} e^{-\lambda(x+s)} \mathbb{1}_{(0, \infty)}(z) \mathbb{1}_{(0, \infty)}(x + s - z). \end{aligned}$$

Per calcolare il membro sinistro in (6.82), dobbiamo integrare questa densità sull'insieme dei punti $(z, x) \in \mathbb{R}^2$ per cui $z \leq s$ e $x \in I_1 \subseteq (0, \infty)$. Per tali valori si ha $x + s - z > 0$, pertanto la funzione indicatrice $\mathbb{1}_{(0, \infty)}(x + s - z)$ “scompare” e si ha

$$\begin{aligned} P(T_h \leq s, (T_{h+1} - s) \in I_1) &= \frac{\lambda^h}{(h-1)!} e^{-\lambda s} \left(\int_0^s z^{h-1} dz \right) \left(\int_{I_1} \lambda e^{-\lambda x} dx \right) \\ &= \frac{(\lambda s)^h}{h!} e^{-\lambda s} \int_{I_1} \lambda e^{-\lambda x} dx = P(N_s = h) \int_{I_1} \lambda e^{-\lambda x} dx, \end{aligned}$$

perché $N_s \sim \text{Pois}(\lambda s)$, per la Proposizione 6.54. Questo dimostra la relazione (6.82) e conclude la dimostrazione. \square

6.5.5 I vettori aleatori uniformi e il paradosso di Bertrand *

In questo paragrafo discutiamo un “paradosso”, proposto nel 1888 dal matematico francese Joseph Bertrand, che mostra come l'interpretazione intuitiva di un vettore aleatorio uniforme su un sottoinsieme $C \subseteq \mathbb{R}^n$ (si ricordi l'Esempio 6.39) possa nascondere delle insidie.

Nel piano cartesiano Oxy , si consideri la circonferenza di centro O e di raggio r , in cui è inscritto un triangolo equilatero. Scegliendo “a caso” una corda di tale circonferenza, qual è la probabilità che la corda abbia lunghezza maggiore del lato del triangolo? Bertrand propose le seguenti tre “soluzioni”.

Soluzione 1. Scegliamo i due estremi A e B della corda, indipendentemente l'uno dall'altro, con distribuzione uniforme sulla circonferenza. Più precisamente, identificando un punto A della circonferenza con l'angolo che la semiretta OA forma con il semiasse positivo delle x , assumiamo $A, B \sim U(0, 2\pi)$, e che siano indipendenti. Per la Proposizione 6.41, il vettore aleatorio (A, B) ha densità

$$f(x, y) = \frac{1}{4\pi^2} \mathbb{1}_{[0, 2\pi)}(x) \mathbb{1}_{[0, 2\pi)}(y).$$

Tenendo presente che A e B sono rappresentati da angoli, e dunque la loro somma o differenza va interpretata modulo 2π (scriveremo mod. 2π), la corda AB è più lunga

del lato del triangolo equilatero se e solo se $A - B \in [\frac{2}{3}\pi, \frac{4}{3}\pi] \bmod. 2\pi$. La probabilità che ciò avvenga è dunque, usando il teorema di Fubini-Tonelli,

$$\begin{aligned} P(A - B \in [\tfrac{2}{3}\pi, \tfrac{4}{3}\pi] \bmod. 2\pi) &= \frac{1}{4\pi^2} \int_{\{(x,y): x-y \in [\tfrac{2}{3}\pi, \tfrac{4}{3}\pi] \bmod. 2\pi\}} \mathbb{1}_{[0,2\pi)}(x) \mathbb{1}_{[0,2\pi)}(y) dx dy \\ &= \frac{1}{2\pi} \int_0^{2\pi} \left[\frac{1}{2\pi} \int_{\{x: x-y \in [\tfrac{2}{3}\pi, \tfrac{4}{3}\pi] \bmod. 2\pi\}} \mathbb{1}_{[0,2\pi)}(x) dx \right] dy. \end{aligned}$$

Valutiamo l'integrale interno: per ogni $y \in [0, 2\pi)$ fissato, l'insieme degli $x \in [0, 2\pi)$ per cui $x - y \in [\frac{2}{3}\pi, \frac{4}{3}\pi] \bmod. 2\pi$ forma un intervallo di ampiezza $\frac{2}{3}\pi$, pertanto

$$P(A - B \in [\tfrac{2}{3}\pi, \tfrac{4}{3}\pi] \bmod. 2\pi) = \frac{1}{2\pi} \int_0^{2\pi} \left[\frac{1}{2\pi} \frac{2}{3}\pi \right] dy = \frac{1}{3}.$$

Soluzione 2. Notiamo che se P è un punto del cerchio $C_r := \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \leq r\}$ diverso dal centro O , esso è il punto medio di un'unica corda. Dunque la scelta a caso di una corda può essere identificata con la scelta, con distribuzione uniforme, di un punto (X, Y) in C_r , cioè $(X, Y) \sim U(C_r)$. L'eventualità che $(X, Y) = (0, 0)$, nel qual caso il punto non identifica univocamente una corda, può essere trascurata, in quanto avviene con probabilità 0, essendo $P((X, Y) = (x, y)) = 0$ per ogni $(x, y) \in \mathbb{R}^2$.

Con un elementare argomento geometrico, si vede che la corda avente (X, Y) come punto medio è più lunga del lato del triangolo equilatero inscritto se e solo se (X, Y) dista dal centro meno di $r/2$. Quindi la probabilità da calcolare è

$$P(X^2 + Y^2 \leq r^2/4) = \frac{1}{\text{mis}(C_r)} \int_{\{(x,y): x^2+y^2 \leq r^2/4\}} dx dy = \frac{\text{mis}(C_{r/2})}{\text{mis}(C_r)} = \frac{\pi \frac{r^2}{4}}{\pi r^2} = \frac{1}{4}.$$

Soluzione 3. Procediamo come nella soluzione precedente, identificando una corda con il suo punto medio P . Tale punto può essere “scelto a caso” assegnando distribuzione $U(0, 2\pi)$ all'angolo Θ che la semiretta OP forma con il semiasse positivo delle x , e distribuzione $U(0, r)$ alla sua distanza da O , che indichiamo con R . La distribuzione congiunta di (Θ, R) , che si potrebbe ottenere assumendo l'indipendenza di θ e R , non è rilevante, poiché la probabilità richiesta è

$$P(R < r/2) = \frac{1}{2}.$$

Come si vede, abbiamo ottenuto tre risposte diverse! Ma questo perché abbiamo formalizzato in tre modi *non equivalenti* l'idea di “scelta casuale di una corda”. Tale idea è risultata troppo vaga per identificare un solo modello corrispondente, e ciò ha causato l'apparente paradosso. Il passaggio da una formulazione intuitiva ad una matematica di un modello richiede, spesso, grande attenzione.

Esercizi

Esercizio 6.6. Sia $(N_t)_{t \in [0, \infty)}$ un processo di Poisson di intensità λ . Per ogni intervallo limitato $(a, b] \subseteq [0, \infty)$, definiamo la variabile aleatoria

$$N_{(a, b]} := N_b - N_a.$$

Se $I = \bigcup_{i=1}^n I_i \subseteq [0, \infty)$ è un'unione finita di intervalli $I_i = (a_i, b_i]$ disgiunti, definiamo

$$N_I := \sum_{i=1}^n N_{I_i} = \sum_{i=1}^n (N_{b_i} - N_{a_i}).$$

Indichiamo con $|(a, b]| = b - a$ la lunghezza di un intervallo e, analogamente, definiamo $|I| = \sum_{i=1}^n |I_i|$ come la lunghezza totale del sottoinsieme $I = \bigcup_{i=1}^n I_i$.

Fissati arbitrariamente $k \geq 2$ e $0 = t_0 < t_1 < \dots < t_k < \infty$, definiamo gli intervalli I_1, \dots, I_k ponendo $I_j = (t_{j-1}, t_j]$. Per il Teorema 6.55, le variabili aleatorie N_{I_1}, \dots, N_{I_k} sono indipendenti, con distribuzione $N_{I_j} \sim \text{Pois}(\lambda |I_j|)$.

Siano infine A e B due sottoinsiemi non vuoti e disgiunti di $\{1, \dots, k\}$. Definendo $I := \bigcup_{i \in A} I_i$ e $J := \bigcup_{j \in B} I_j$, si mostri che le variabili aleatorie N_I e N_J sono indipendenti, con distribuzioni $N_I \sim \text{Pois}(\lambda |I|)$ e $N_J \sim \text{Pois}(\lambda |J|)$.

6.6 Vettori aleatori normali *

6.6.1 Matrice di covarianza *

Cominciamo con qualche richiamo di algebra lineare. Nel seguito indicheremo con A^T la matrice trasposta di una matrice A e con I la matrice identità (di qualunque dimensione), le cui componenti sono $I_{i,j} = 1$ se $i = j$ e $I_{i,j} = 0$ altrimenti. Un elemento $x \in \mathbb{R}^n$, che nel testo scriveremo $x = (x_1, x_2, \dots, x_n)$, nelle operazioni con matrici verrà identificato con il *vettore colonna*

$$\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}.$$

In particolare, se A è una matrice di dimensione $m \times n$, il vettore $Ax \in \mathbb{R}^m$ ha componenti $(Ax)_i = \sum_{j=1}^n A_{ij}x_j$.

Per il teorema spettrale, ogni matrice V reale *simmetrica* è diagonalizzabile mediante matrici ortogonali, ossia può essere scritta nella forma

$$V = O^T D O, \tag{6.83}$$

dove la matrice O è ortogonale, cioè $O^T O = I$, e D è diagonale:

$$D = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \lambda_n \end{pmatrix}. \quad (6.84)$$

I numeri $\lambda_1, \dots, \lambda_n \in \mathbb{R}$ sono gli autovalori di V . Supponiamo ora che V , oltre che simmetrica, sia anche *semidefinita positiva*, ossia

$$x^T V x = \sum_{i,j=1}^n V_{ij} x_i x_j \geq 0, \quad \forall x \in \mathbb{R}^n.$$

In questo caso, la matrice diagonale D è anch'essa semidefinita positiva e dunque $\lambda_i \geq 0$ per $i = 1, 2, \dots, n$. Possiamo allora porre

$$A := O^T \begin{pmatrix} \sqrt{\lambda_1} & 0 & \cdots & 0 \\ 0 & \sqrt{\lambda_2} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \sqrt{\lambda_n} \end{pmatrix} O. \quad (6.85)$$

Si verifica facilmente che la matrice A è simmetrica e semidefinita positiva e si ha $V = A^2 := AA$. Inoltre, si può mostrare che A è univocamente determinata da tali proprietà. La indicheremo pertanto con la notazione $A := \sqrt{V}$.

Sia ora $X = (X_1, X_2, \dots, X_n)$ un vettore aleatorio di dimensione n . Se tutte le componenti X_i ammettono momento secondo finito, ossia $X_i \in L^2$, possiamo associare a X una matrice reale V di dimensione $n \times n$, detta *matrice di covarianza*, definita da:

$$V_{ij} := \text{Cov}(X_i, X_j).$$

Tale matrice, ovviamente simmetrica, è anche semidefinita positiva, dal momento che per ogni $x \in \mathbb{R}^n$

$$x^T V x = \sum_{i,j=1}^n x_i x_j \text{Cov}(X_i, X_j) = \text{Cov} \left(\sum_{i=1}^n x_i X_i, \sum_{j=1}^n x_j X_j \right) = \text{Var} \left(\sum_{i=1}^n x_i X_i \right) \geq 0,$$

dove abbiamo usato la bilinearità della covarianza.

Viceversa, mostriamo che *ogni matrice V simmetrica e semidefinita positiva* è matrice di covarianza di un vettore aleatorio. A tal fine, sia $Z = (Z_1, Z_2, \dots, Z_n)$ qualsiasi vettore aleatorio la cui matrice di covarianza è l'identità I , ossia le cui componenti sono variabili aleatorie reali scorrelate e con varianza uguale a 1. Ponendo $A := \sqrt{V}$,

il vettore aleatorio AZ ha matrice di covarianza V :

$$\begin{aligned} \text{Cov}((AZ)_i, (AZ)_j) &= \text{Cov}\left(\sum_{m=1}^n A_{im} Z_m, \sum_{k=1}^n A_{jk} Z_k\right) = \sum_{m,k=1}^n A_{im} A_{jk} \text{Cov}(Z_m, Z_k) \\ &= \sum_{m,k=1}^n A_{im} A_{jk} I_{m,k} = \sum_{k=1}^n A_{ik} A_{jk} = (AA^T)_{ij} = (A^2)_{ij} = V_{ij}. \end{aligned}$$

6.6.2 Definizione e proprietà principali *

I vettori aleatori normali costituiscono la generalizzazione multidimensionale delle variabili aleatorie reali normali, che abbiamo studiato nel Paragrafo 6.3.4. Per il loro studio risulta conveniente estendere la nozione di funzione generatrice dei momenti al caso di vettori aleatori.

Definizione 6.57 (Funzione generatrice dei momenti multidimensionale).

Sia $X = (X_1, X_2, \dots, X_n)$ un vettore aleatorio di dimensione n . La sua *funzione generatrice dei momenti* $M_X : \mathbb{R}^n \rightarrow [0, +\infty]$ è definita da

$$M_X(u) = M_X(u_1, u_2, \dots, u_n) := E[e^{u^T X}] = E\left(\exp\left[\sum_{i=1}^n u_i X_i\right]\right).$$

Un'importante proprietà, che enunciamo senza dimostrare, è che la funzione generatrice dei momenti caratterizza la distribuzione di un vettore aleatorio, purché essa sia finita in un intorno dell'origine.[†]

Teorema 6.58. *Siano X, Y vettori aleatori di dimensione n , le cui funzioni generatrici dei momenti M_X, M_Y sono finite per $u \in (-a, a)^n$, con $a > 0$. Se $M_X(u) = M_Y(u)$ per ogni $u \in \mathbb{R}^n$, allora X e Y hanno la stessa distribuzione.*

Come corollario, otteniamo un'utile caratterizzazione dell'indipendenza di vettori aleatori (la formuliamo per semplicità nel caso di due vettori, ma si generalizza facilmente al caso di un numero finito).

Proposizione 6.59. *Siano X e Y vettori aleatori, definiti sullo stesso spazio di probabilità, di dimensione n e m rispettivamente. Supponiamo che la funzione generatrice dei momenti del vettore aleatorio (X, Y) (di dimensione $n + m$) sia finita in un intorno dell'origine. Allora X e Y sono indipendenti se e solo se*

$$M_{(X,Y)}(u, v) = M_X(u) M_Y(v), \quad \forall u \in \mathbb{R}^n, v \in \mathbb{R}^m. \quad (6.86)$$

[†] Si noti che, già per variabili aleatorie reali, qualche ipotesi di finitezza è necessaria affinché la funzione generatrice dei momenti caratterizzi la distribuzione, come mostra l'Esercizio 3.15.

Dimostrazione. Se X e Y sono indipendenti, le variabili aleatorie reali $e^{u^T X}$ e $e^{v^T Y}$ sono indipendenti, per ogni $u \in \mathbb{R}^n$ e $v \in \mathbb{R}^m$ (perché?), pertanto la relazione (6.86) segue facilmente.

Viceversa, supponiamo che valga (6.86). Siano \tilde{X} e \tilde{Y} vettori aleatori *indipendenti*, con le stesse distribuzioni marginali (e dunque con le stesse funzioni generatrici dei momenti) di X e Y rispettivamente. Per la prima parte della dimostrazione, la funzione generatrice dei momenti del vettore aleatorio (\tilde{X}, \tilde{Y}) coincide con quella (6.86) di (X, Y) , quindi per il Teorema 6.58 i vettori aleatori (\tilde{X}, \tilde{Y}) e (X, Y) hanno la stessa distribuzione. Dato che \tilde{X} e \tilde{Y} sono per costruzione indipendenti, segue che anche X e Y lo sono. \square

Introduciamo dunque i vettori aleatori normali. Siano Z_1, Z_2, \dots, Z_n variabili aleatorie reali *indipendenti*, con distribuzione normale standard, e indichiamo con $Z := (Z_1, \dots, Z_n)$ il vettore che le ha per componenti. Ricordando la formula (6.44), la funzione generatrice dei momenti M_Z vale

$$M_Z(u) = E(e^{u^T Z}) = E\left(\prod_{i=1}^n e^{u_i Z_i}\right) = \prod_{i=1}^n E(e^{u_i Z_i}) = \prod_{i=1}^n e^{\frac{u_i^2}{2}} = \exp\left[\frac{1}{2}u^T u\right]. \quad (6.87)$$

Sia ora V un'arbitraria matrice simmetrica e semidefinita positiva di dimensione n . Posto $X = AZ + \mu$, con $A = \sqrt{V}$, dato che $AA^T = V$, si ha

$$\begin{aligned} M_X(u) &= E(\exp[u^T AZ + u^T \mu]) = e^{u^T \mu} E(\exp[(A^T u)^T Z]) \\ &= e^{u^T \mu} \exp\left[\frac{1}{2}(A^T u)^T A^T u\right] = e^{u^T \mu} \exp\left[\frac{1}{2}u^T V u\right]. \end{aligned} \quad (6.88)$$

Dato che la funzione generatrice dei momenti di un vettore aleatorio ne identifica la distribuzione, per il Teorema 6.58, possiamo fornire la seguente definizione.

Definizione 6.60 (Vettori aleatori normali). Sia $\mu \in \mathbb{R}^n$ e sia V una matrice di dimensione n simmetrica e semidefinita positiva. Un vettore aleatorio $X = (X_1, X_2, \dots, X_n)$ si dice *normale* con media μ e matrice di covarianza V , e si scrive $X \sim N(\mu, V)$, se la sua funzione generatrice dei momenti è data da

$$M_X(u) = e^{u^T \mu} \exp\left[\frac{1}{2}u^T V u\right]. \quad (6.89)$$

Dalle relazioni (6.87) e (6.88), ricordando ancora il Teorema 6.58, si ottiene facilmente il seguente risultato, in cui 0 indica il vettore nullo in \mathbb{R}^n . Si noti l'analogia con il caso reale, in cui la matrice di covarianza V si riduce al numero positivo σ^2 .

Proposizione 6.61. Si ha $Z \sim N(0, I)$ se e solo se le componenti di Z sono variabili aleatorie normali standard indipendenti. Inoltre, un vettore aleatorio $X \sim N(\mu, V)$ ha la stessa distribuzione di $\sqrt{V}Z + \mu$, con $Z \sim N(0, I)$.

Ricordiamo che una variabile aleatoria reale $X \sim N(\mu, \sigma^2)$ è assolutamente continua se e solo se $\sigma \neq 0$. Per vettori aleatori normali, la condizione che garantisce l'assoluta continuità è l'invertibilità della matrice di covarianza.

Teorema 6.62 (Densità di un vettore normale). *Sia $X \sim N(\mu, V)$. Se V è invertibile, X è un vettore aleatorio assolutamente continuo con densità*

$$f_X(x) = \frac{1}{(2\pi)^{d/2} \sqrt{\det(V)}} \exp \left[-\frac{1}{2} (x - \mu)^T V^{-1} (x - \mu) \right]. \quad (6.90)$$

Dimostrazione. Se $Z \sim N(0, I)$, le componenti Z_1, \dots, Z_n di Z sono variabili aleatorie reali $N(0, 1)$ indipendenti, grazie alla Proposizione 6.61. Pertanto, per la Proposizione 6.41, Z è un vettore assolutamente continuo con densità data da

$$f_Z(z_1, \dots, z_n) = f_{Z_1}(z_1) f_{Z_2}(z_2) \cdots f_{Z_n}(z_n) = \frac{e^{-\frac{1}{2}(z_1^2 + z_2^2 + \dots + z_n^2)}}{(2\pi)^{d/2}} = \frac{e^{-\frac{1}{2}z^T z}}{(2\pi)^{d/2}},$$

dove $z = (z_1, \dots, z_n)$. Se $X \sim N(\mu, V)$, allora X ha la stessa distribuzione di $AZ + \mu$, dove $A = \sqrt{V}$. Se V è invertibile, tutti i suoi autovalori sono strettamente positivi: $\lambda_1 > 0, \dots, \lambda_n > 0$, pertanto anche A è invertibile e $\det(A) = \sqrt{\det(V)}$ (si ricordi la relazione (6.85)). Applicando la Proposizione 6.50, si ottiene che $X = AZ + \mu$ è un vettore aleatorio assolutamente continuo, con densità

$$f_X(x) = \frac{1}{|\det(A)|} f_Z(A^{-1}(x - \mu)) = \frac{e^{-\frac{1}{2}(A^{-1}(x - \mu))^T A^{-1}(x - \mu)}}{(2\pi)^{d/2} \sqrt{\det(V)}},$$

che coincide con (6.90), perché $(A^{-1})^T A^{-1} = (A^{-1})^2 = (A^2)^{-1} = V^{-1}$. \square

Osservazione 6.63. Se $X \sim N(\mu, V)$ con V matrice non invertibile ($\det V = 0$), il vettore aleatorio X non è assolutamente continuo. Ricordiamo infatti che X ha la stessa distribuzione di $\sqrt{V}Z + \mu$, con $Z \sim N(0, I)$. Posto

$$H := \{\sqrt{V}x + \mu : x \in \mathbb{R}^n\},$$

H è un sottospazio affine di \mathbb{R}^n , la cui dimensione è uguale al rango della matrice \sqrt{V} , che è lo stesso della matrice V . Se X avesse densità f_X , si avrebbe

$$1 = P(X \in H) = \int_{\mathbb{R}^n} \mathbb{1}_H(x) f_X(x) dx, \quad (6.91)$$

per la relazione (6.51). Ma se V non è invertibile, H è un sottospazio affine di dimensione strettamente minore di n , che ha misura n -dimensionale nulla e pertanto l'integrale in (6.91) vale zero, ottenendo una contraddizione. \square

Una delle principali proprietà dei vettori aleatori normali è il fatto che la normalità sia preservata per trasformazioni lineari-affini. Ciò implica, in particolare, che le componenti di un vettore normale sono variabili aleatorie normali.

Proposizione 6.64 (Trasformazioni lineari-affini di vettori normali). Sia $X \sim N(\mu, V)$ un vettore aleatorio normale n -dimensionale e siano A una matrice $m \times n$ e $b \in \mathbb{R}^m$ (con m arbitrario). Il vettore aleatorio m -dimensionale $Y := AX + b$ è normale: più precisamente,

$$Y \sim N(A\mu + b, AVA^T).$$

Dimostrazione. Per $u \in \mathbb{R}^m$, basta osservare che

$$\begin{aligned} M_Y(u) &= E(\exp[u^T AX + u^T b]) = e^{u^T b} E(\exp[(A^T u)^T X]) = e^{u^T b} M_X(A^T u) \\ &= e^{u^T b} e^{(A^T u)^T \mu} \exp\left[\frac{1}{2}(A^T u)^T V (A^T u)\right] = \exp[u^T (A\mu + b)] \exp\left[\frac{1}{2}u^T AVA^T u\right], \end{aligned}$$

che dimostra la tesi. \square

Mostriamo che, se le componenti di un vettore aleatorio sono variabili aleatorie normali, non è detto che il vettore sia normale.

Esempio 6.65. Siano X e Y variabili aleatorie reali indipendenti tali che $X \sim N(0, 1)$ e $P(Y = 1) = P(Y = -1) = 1/2$. Definiamo $Z = XY$ e osserviamo che

$$\begin{aligned} F_Z(x) &= P(Z \leq x) = P(XY \leq x, Y = 1) + P(XY \leq x, Y = -1) \\ &= P(X \leq x, Y = 1) + P(X \geq -x, Y = -1) \\ &= \frac{1}{2} P(X \leq x) + \frac{1}{2} P(X \geq -x) = P(X \leq x) = F_X(x), \end{aligned}$$

dove abbiamo usato l'indipendenza di X e Y e il fatto che, essendo la densità di X una funzione pari, $P(X \geq -x) = P(X \leq x)$. Dato che $X \sim N(0, 1)$, segue che $Z \sim N(0, 1)$. Si osservi inoltre che

$$\text{Cov}(X, Z) = E(XZ) - E(X)E(Z) = E(XZ) = E(X^2 Y) = E(X^2)E(Y) = 0.$$

Consideriamo infine il vettore aleatorio (X, Z) , la cui matrice di covarianza è l'identità, per quanto appena mostrato. Se (X, Z) fosse normale, le variabili aleatorie X e Z sarebbero indipendenti, per la Proposizione 6.61, e si dovrebbe avere

$$E(|X||Z|) = E(|X|)E(|Z|) = (E(|X|))^2.$$

Ma così non è, dal momento che

$$E(|X||Z|) = E(|X||XY|) = E(|X|^2) = (E(|X|))^2 + \text{Var}(|X|) > (E(|X|))^2,$$

avendo usato il fatto che $|Y| = 1$ e $\text{Var}(|X|) > 0$ (perché?). \square

Abbiamo visto nel Corollario 3.74 che due variabili aleatorie reali indipendenti sono scorrelate. Il fatto interessante è che, per le componenti di un vettore normale, è vera anche l'implicazione inversa.

Proposizione 6.66 (Indipendenza e scorrelazione per vettori normali). *Sia $(X, Y) = (X_1, X_2, \dots, X_n, Y_1, Y_2, \dots, Y_m)$ un vettore aleatorio normale di dimensione $n + m$, tale che $\text{Cov}(X_i, Y_j) = 0$ per ogni $i \in \{1, 2, \dots, n\}$ e $j \in \{1, 2, \dots, m\}$. Allora X e Y sono vettori aleatori (normali) indipendenti.*

Dimostrazione. Per la Proposizione 6.59, è sufficiente mostrare che

$$M_{(X,Y)}(u, v) = M_X(u)M_Y(v). \quad (6.92)$$

Per ipotesi, la matrice di covarianza $V_{(X,Y)}$ del vettore (X, Y) è “a blocchi”:

$$V_{(X,Y)} = \begin{pmatrix} V_X & 0 \\ 0 & V_Y \end{pmatrix},$$

dove V_X e V_Y sono le matrici di covarianza di X e Y , mentre “0” indica una matrice identicamente nulla (di dimensione qualunque, in questo caso $n \times m$ e $m \times n$). Se μ_X (risp. μ_Y) è il vettore delle medie di X (risp. di Y), si ha allora

$$\begin{aligned} M_{(X,Y)}(u, v) &= \exp[(u, v)^T(\mu_X, \mu_Y)] \exp\left[\frac{1}{2}(u^T, v^T)V_{(X,Y)}\begin{pmatrix} u \\ v \end{pmatrix}\right] \\ &= \exp[u^T \mu_X] \exp\left[\frac{1}{2}u^T V_X u\right] \exp[v^T \mu_Y] \exp\left[\frac{1}{2}v^T V_Y v\right] = M_X(u)M_Y(v), \end{aligned}$$

che dimostra (6.92). □

6.6.3 Proiezioni ortogonali di vettori normali *

Concludiamo questo paragrafo con un risultato relativo ad una particolare classe di trasformazioni lineari applicate ad un vettore aleatorio con distribuzione $N(0, I)$. Sia H un sottospazio vettoriale di \mathbb{R}^n e indichiamo con H^\perp il sottospazio ortogonale:

$$H^\perp := \{z \in \mathbb{R}^n : z^T w = 0 \text{ per ogni } w \in H\}.$$

Ogni $x \in \mathbb{R}^n$ ammette un'unica decomposizione

$$x = y + z, \quad \text{con } y \in H, \quad z \in H^\perp.$$

L'applicazione $x \mapsto y := Px$ è lineare e viene detta *proiezione ortogonale* su H . Essa sarà identificata con la corrispondente matrice P , che è simmetrica e soddisfa $P^2 = P$. Si noti che $I - P$ è la proiezione ortogonale sul sottospazio H^\perp . Nel seguito, usiamo la notazione $\|x\|^2 := x^T x = x_1^2 + x_2^2 + \dots + x_n^2$, per $x \in \mathbb{R}^n$.

Ricordando che le variabili aleatorie $\chi^2(n)$ sono state definite nel Paragrafo 6.5.1, dimostriamo il seguente risultato.

Teorema 6.67. Sia $Z \sim N(0, I)$ un vettore aleatorio di dimensione n e sia P la proiezione ortogonale su un sottospazio H di dimensione $m < n$. Allora i vettori aleatori PZ e $(I - P)Z$ sono indipendenti e hanno le seguenti proprietà:

$$\|PZ\|^2 \sim \chi^2(m), \quad \|(I - P)Z\|^2 \sim \chi^2(d - m).$$

Dimostrazione. Sia X il vettore aleatorio $2n$ -dimensionale $(PZ, (I - P)Z)$. Per la Proposizione 6.64, X è un vettore normale. Se S è la matrice $n \times n$ i cui elementi sono dati da

$$S_{ij} = \text{Cov}((PZ)_i, ((I - P)Z)_j),$$

ricordando che $\text{Cov}(Z_k, Z_l) = I_{kl}$ vale 0 se $k \neq l$ e vale 1 se $k = l$, si vede che

$$S_{ij} = \sum_{k,l=1}^n P_{ik}(I - P)_{jl} \text{Cov}(Z_k, Z_l) = \sum_{k=1}^n P_{ik}(I - P)_{jk} = (P(I - P^T))_{ij}.$$

Dato che P è simmetrica, $P^T = P$ e pertanto

$$S = P(I - P) = P - P^2 = 0.$$

Quindi, per la Proposizione 6.66, PZ e $(I - P)Z$ sono vettori aleatori indipendenti.

Sia ora v_1, v_2, \dots, v_m una base ortonormale per il sottospazio H , e v_{m+1}, \dots, v_n un'altra base per H^\perp . Pertanto $v_1, v_2, \dots, v_m, v_{m+1}, \dots, v_n$ formano una base ortonormale di \mathbb{R}^n . La proiezione ortogonale P è caratterizzata dal fatto che

$$Pv_i = \begin{cases} v_i & \text{se } i \leq m \\ 0 & \text{altrimenti.} \end{cases}$$

Indichiamo con O la (matrice della) trasformazione lineare tale che $Ov_i = e_i$ per $i = 1, 2, \dots, n$, dove e_1, e_2, \dots, e_n sono i vettori della base canonica di \mathbb{R}^n . Poiché O trasforma una base ortonormale in un'altra base ortonormale, segue che O è una matrice ortogonale. Posto $Q := OPO^T$, si verifica facilmente che Q è la proiezione ortogonale sullo spazio generato da e_1, e_2, \dots, e_m . Se definiamo il vettore aleatorio $W := OZ$, abbiamo che, essendo $O^T O = I$,

$$\begin{aligned} \|QW\|^2 &= (QW)^T QW = (OPO^T OZ)^T (OPO^T OZ) \\ &= Z^T PO^T OPZ = (PZ)^T PZ = \|PZ\|^2. \end{aligned} \quad (6.93)$$

Inoltre

$$\|QW\|^2 = W_1^2 + W_2^2 + \dots + W_m^2. \quad (6.94)$$

Per la Proposizione 6.64 si ha $W := OZ \sim N(0, I)$. Ricordando il Paragrafo 6.5.1, cf. in particolare (6.62), le relazioni (6.93) e (6.94) implicano che

$$\|PZ\|^2 = \|QW\|^2 \sim \chi^2(m).$$

Poiché $I - P$ è anch'essa una proiezione ortogonale, il corrispondente risultato vale anche per $\|(I - P)Z\|^2$. \square

Come caso speciale, consideriamo il vettore $u \in \mathbb{R}^n$ con componenti $u_i = 1$ per ogni $i = 1, 2, \dots, n$, e sia H il sottospazio generato da u . Se P è la proiezione ortogonale su H si ha che

$$Px = \bar{x}u, \quad \text{dove} \quad \bar{x} := \frac{1}{n} \sum_{i=1}^n x_i.$$

Per verificarlo, bisogna mostrare che $x - Px$ è ortogonale ad ogni elemento di H ; è ovviamente sufficiente mostrare che $x - Px$ è ortogonale a u :

$$(x - Px)^T u = x^T u - (Px)^T u = \sum_{i=1}^n x_i - \bar{x}u^T u = 0,$$

visto che $u^T u = n$. Applichiamo dunque il Teorema 6.67: se $Z \sim N(0, I)$, il vettore aleatorio $Z - PZ = (Z_1 - \bar{Z}, Z_2 - \bar{Z}, \dots, Z_n - \bar{Z})$ è indipendente da $PZ = \bar{Z}u$ o, equivalentemente, dalla variabile aleatoria $\bar{Z} := \frac{1}{n} \sum_{i=1}^n Z_i$. Inoltre

$$\|Z - PZ\|^2 = \sum_{i=1}^n (Z_i - \bar{Z})^2 \sim \chi^2(n-1).$$

Più in generale, se X_1, X_2, \dots, X_n sono variabili aleatorie reali indipendenti e con la stessa distribuzione $N(\mu, \sigma^2)$, con $\sigma > 0$, allora le variabili $Z_i := \frac{X_i - \mu}{\sigma}$ sono indipendenti e hanno distribuzione $N(0, 1)$. Osservando che $\bar{X} = \mu + \sigma \bar{Z}$, deduciamo immediatamente il seguente importante risultato, che verrà ripreso nel Capitolo 8 ed è alla base di numerosi risultati di statistica.

Proposizione 6.68. *Siano X_1, X_2, \dots, X_n variabili aleatorie reali indipendenti con la stessa distribuzione $N(\mu, \sigma^2)$, con $\sigma > 0$, e definiamo*

$$\bar{X} := \frac{1}{n} \sum_{i=1}^n X_i.$$

La variabile aleatoria reale

$$\frac{1}{\sigma^2} \sum_{i=1}^n (X_i - \bar{X})^2$$

ha distribuzione $\chi^2(n-1)$ ed è indipendente da \bar{X} .

6.7 Esercizi di riepilogo

Ricordiamo che “i.i.d.” significa “indipendenti e identicamente distribuite”. Gli esercizi segnalati con un asterisco richiedono l’uso dei vettori aleatori.

Esercizio 6.7. Data una variabile aleatoria $X \sim U(-\pi/2, \pi/2)$, si mostri che $Y := \cos(X)$ è una variabile assolutamente continua e se ne determini la densità.

Esercizio 6.8. Sia X un punto scelto uniformemente nell’intervallo $[0, 2]$. Qual è la probabilità che il triangolo equilatero di lato X abbia area maggiore di 1?

Esercizio 6.9. Sia $X \sim U(0, 1)$ e sia $Y := 4X(1 - X)$.

- (i) Si determini la funzione di ripartizione di Y , si deduca che la variabile Y è assolutamente continua e se ne calcoli la densità.
- (ii) Si calcoli $\text{Cov}(X, Y)$ (che è ben definita: perché?).

Esercizio 6.10. Sia X una variabile aleatoria reale assolutamente continua con densità

$$f_X(x) = -\log(x^c) \mathbb{1}_{(0,1)}(x).$$

- (i) Si determini il valore di $c \in \mathbb{R}$ affinché f_X sia effettivamente una densità, e si determini la funzione di ripartizione di X .
- (ii) Sia $Y = -\log X$. Si mostri che $Y \sim \text{Gamma}(\alpha, \lambda)$, determinando α e λ .

Esercizio 6.11. Sia $X \sim U(-1, 1)$. Si determini la funzione di ripartizione della variabile aleatoria $Y := X^+ = \max(X, 0)$. Si deduca che la distribuzione di Y non è né discreta né assolutamente continua.

Esercizio 6.12. Sia X una variabile aleatoria assolutamente continua con densità

$$f_X(x) := 3x^2 \mathbb{1}_{[0,1]}(x).$$

Si considerino le variabili aleatorie

$$A := -\log X, \quad B := \min\{1, A\}.$$

- (i) Si calcoli la funzione di ripartizione della variabile aleatoria A e se ne identifichi la distribuzione (notevole).
- (ii) Consideriamo la seguente equazione di secondo grado per l’incognita x , con coefficienti determinati dalla variabile aleatoria A :

$$x^2 + 3Ax + 2A^2 + 4 = 0.$$

Qual è la probabilità che l’equazione non ammetta soluzioni reali?

- (iii) Si calcoli la funzione di ripartizione di B e si deduca che B non è una variabile aleatoria assolutamente continua. B è una variabile aleatoria discreta?

Esercizio 6.13. Sia X una variabile aleatoria con distribuzione $U(0, 1)$ e sia

$$Z := \frac{1}{X^r},$$

con $r \in \mathbb{R}$ parametro fissato.

- (i) Per quali valori di $p \in (0, \infty)$ si ha $E(|Z|^p) < \infty$, ossia $Z \in L^p$?
- (ii) Si determini la distribuzione della variabile aleatoria Z .

[Sugg. Si calcoli la funzione di ripartizione di Z , separando i casi $r < 0$, $r = 0$ e $r > 0$.]

Esercizio 6.14. Siano X e Y variabili aleatorie indipendenti, uniformemente distribuite nell'intervallo $(-1, +1)$. Si mostri che $Z := X + Y$ è una variabile aleatoria assolutamente continua, con densità

$$f_Z(z) = \frac{1}{4}(2 - |z|) \mathbb{1}_{(-2, +2)}(z).$$

Esercizio 6.15. Siano X, Y variabili aleatorie indipendenti, entrambe con distribuzione $\text{Exp}(1)$, e si definisca $Z := X - Y$.

- (i) Si mostri che Z è una variabile aleatoria assolutamente continua, con densità

$$f_Z(z) = \frac{1}{2}e^{-|z|}.$$

[Sugg. Si ponga $Y' := -Y$, così che $Z = X + Y'$.]

- (ii) Si mostri che $|Z| \sim \text{Exp}(1)$.

Esercizio 6.16. Sia X un punto aleatorio dell'intervallo $(0, 1)$ (non necessariamente uniformemente distribuito). Esso divide l'intervallo $(0, 1)$ in due segmenti. Sia $Y \geq 1$ il rapporto tra il segmento più lungo e quello più corto.

- (i) Si esprima Y in funzione di X .

[Sugg. Potrebbe risultare utile usare gli eventi $\{X < \frac{1}{2}\}$ e $\{X \geq \frac{1}{2}\}$.]

- (ii) Supponiamo che $X \sim U(0, 1)$. Si determinino la funzione di ripartizione e la densità di Y e si mostri che Y ha valor medio infinito.
- (iii) Assumiamo ora che X sia una variabile aleatoria assolutamente continua, a valori in $(0, 1)$, la cui densità f_X soddisfi la relazione:

$$f_X(x) + f_X(1-x) = 2, \quad \text{per ogni } x \in (0, 1). \quad (6.95)$$

Si mostri che la distribuzione di Y è uguale a quella trovata al punto precedente.

[Sugg. Si deduca dalla relazione (6.95) che $F_X(z) - F_X(1-z) = 2z - 1$ per $z \in (0, 1)$.]

- (iv) Si determini una densità f_X , soddisfacente alla relazione (6.95), diversa dalla densità di una variabile aleatoria con distribuzione $U(0, 1)$.

Esercizio 6.17. Siano $(X_n)_{n \in \mathbb{N}}$ variabili aleatorie i.i.d. con $X_i \sim U(0, 1)$.

- (i) Poniamo $Y_n := -\log(X_n)$ per $n \in \mathbb{N}$. Si determini la distribuzione di Y_n e si spieghi perché le variabili aleatorie $(Y_n)_{n \in \mathbb{N}}$ sono indipendenti.
 - (ii) Si determini la distribuzione di $S_n := Y_1 + \dots + Y_n$.
 - (iii) Si calcoli la funzione di ripartizione di $Z_n := X_1 \cdot X_2 \cdots X_n$ e si deduca che Z_n è una variabile aleatoria assolutamente continua. Se ne calcoli dunque la densità.
- [Sugg. Si sfruttino i punti precedenti]

Esercizio 6.18. Siano $(X_n)_{n \in \mathbb{N}}$ variabili aleatorie i.i.d. $U(0, 2)$ e sia

$$Y_n := \min\{Y_1, \dots, Y_n\}, \quad \forall n \in \mathbb{N}.$$

Sia $T \sim \text{Geo}(p)$ una variabile aleatoria indipendente dalle $(X_n)_{n \in \mathbb{N}}$, e si ponga

$$Z := Y_T, \quad \text{cioè} \quad Z(\omega) := Y_{T(\omega)}(\omega).$$

Si determini la funzione di ripartizione di Z , mostrando che è una variabile aleatoria assolutamente continua.

[Sugg. Si determini innanzitutto $P(Z \leq x, T = n)$ per ogni $n \in \mathbb{N}$.]

Esercizio 6.19. Sia X_1, X_2, \dots una successione di variabili aleatorie reali i.i.d. con distribuzione $U(0, 1)$. Introduciamo la variabile aleatoria $T : \Omega \rightarrow \mathbb{N} \cup \{+\infty\}$ e, per $k \in \mathbb{N}$, l'evento A_k definiti da

$$T(\omega) := \inf \left\{ k \geq 1 : X_k(\omega) \leq \frac{1}{3} \right\}, \quad A_k = \left\{ X_k \leq \frac{1}{3} \right\}.$$

Definiamo quindi la variabile aleatoria

$$Y := X_T \mathbb{1}_{\{T < \infty\}},$$

cioè $Y(\omega) := X_{T(\omega)}(\omega)$ se $T(\omega) < \infty$, mentre $Y(\omega) := 0$ altrimenti.

- (i) Per ogni $n \in \mathbb{N}$, si esprima l'evento $\{T = n\}$ in termini degli eventi $(A_k)_{k \in \mathbb{N}}$. Si deduca la distribuzione di T .
- (ii) Si determini la distribuzione di Y .

[Sugg. Si calcoli $P(Y \leq x, T = n)$ per $n \in \mathbb{N}$.]

Esercizio 6.20. Siano X e Y variabili aleatorie real indipendenti. Supponiamo che X sia assolutamente continua, con densità f_X , mentre Y sia discreta, con densità discreta p_Y . Supponiamo inoltre che Y assuma solo un numero finito di valori, cioè l'insieme $Y(\Omega) = \{x \in \mathbb{R} : p_Y(x) > 0\} = \{x_1, \dots, x_n\}$ è finito.

- (i) Si esprima la funzione di ripartizione di $Z := X + Y$ in termini di f_X e p_Y .
- (ii) Si mostri che Z è una variabile aleatoria assolutamente continua, determinandone la densità in funzione di f_X e p_Y .

Esercizio 6.21. L'ufficio informazioni delle Ferrovie dello Stato ha due numeri verdi. I tempi di attesa T_1 e T_2 per parlare con l'operatore sono, per entrambi i numeri, variabili aleatorie esponenziali, con media $\mu = 15$ minuti. Inoltre T_1 e T_2 si possono considerare indipendenti. Avendo a disposizione due telefoni, decido di chiamare contemporaneamente i due numeri, in modo da parlare con l'operatore che risponderà per primo.

- (i) Quanto tempo, in media, dovrò aspettare per parlare con un operatore?
- (ii) Qual è la probabilità di attendere meno di 5 minuti?

Esercizio 6.22. Un congegno elettronico è costituito da n componenti collegate in serie: esso smette di funzionare non appena una qualsiasi delle sue componenti si rompe. I tempi di vita T_1, T_2, \dots, T_n delle n componenti sono variabili aleatorie reali indipendenti e con la stessa distribuzione assolutamente continua, di cui indichiamo con f la densità. Chiaramente $f(x) = 0$ per $x \leq 0$ (i tempi sono quantità positive!). Supponiamo che f sia una funzione continua su $[0, \infty)$, con $f(0) > 0$. Indicando con X_n il tempo di vita dell'intero dispositivo, si mostri che, per ogni $\varepsilon > 0$,

$$\lim_{n \rightarrow +\infty} P(X_n > \varepsilon) = 0$$

Esercizio 6.23. Siano X_1, X_2, \dots, X_n variabili aleatorie i.i.d., con distribuzione $U(0, 1)$, e definiamo

$$L_n := \min\{X_1, X_2, \dots, X_n\}, \quad Z_n := nL_n.$$

Si mostri che la funzione di ripartizione $F_{Z_n}(t)$ converge per $n \rightarrow \infty$ verso un limite $F(t)$, che è la funzione di ripartizione di una variabile aleatoria $\text{Exp}(1)$.

Esercizio 6.24. Ricordiamo che una variabile aleatoria reale X è detta di *Cauchy* se è assolutamente continua con densità

$$f_X(x) := \frac{1}{\pi} \frac{1}{1+x^2}, \quad x \in \mathbb{R}.$$

Può essere utile ricordare che $\arctan(z) + \arctan(\frac{1}{z}) = \frac{\pi}{2}$, per ogni $z \in \mathbb{R}$.

- (i) Si dimostri che la variabile aleatoria $Y := 1/X$ è di Cauchy.
- (ii) Si calcoli $P(X > z)$ per ogni $z \in \mathbb{R}$ e si mostri che per $z \rightarrow +\infty$

$$P(X > z) \sim \frac{1}{\pi} \frac{1}{z},$$

intendendo che il rapporto dei due membri tende a 1 per $z \rightarrow +\infty$.

- (iii) Sia ora $(X_n)_{n \in \mathbb{N}}$ una successione i.i.d. di variabili di Cauchy indipendenti e si definisca $M_n := \max\{X_1, \dots, X_n\}$ per $n \in \mathbb{N}$. Si mostri che per ogni $t > 0$

$$\lim_{n \rightarrow \infty} P\left(\frac{M_n}{n} \leq t\right) = \exp\left(-\frac{1}{\pi} \frac{1}{t}\right).$$

Esercizio 6.25. Sia $\varphi : [0, +\infty) \rightarrow [0, 1)$ una funzione continua, crescente, tale che $\varphi(0) = 0$, $\lim_{t \rightarrow +\infty} \varphi(t) = 1$, e il cui comportamento asintotico per $x \downarrow 0$ è dato da

$$\varphi(x) = \alpha x^k + o(x^k), \quad \text{con } \alpha, k > 0.$$

- (i) Si mostri che la funzione $F(t) := 1 - \varphi(1/t)$ è una funzione di ripartizione, ossia soddisfa le proprietà (5.4).
- (ii) Si consideri una successione $(X_n)_{n \geq 1}$ di variabili aleatorie i.i.d. con funzione di ripartizione F definita sopra. Ponendo

$$Y_n := \frac{\max(X_1, X_2, \dots, X_n)}{n^{1/k}},$$

si mostri che, per ogni $y > 0$, si ha

$$\lim_{n \rightarrow +\infty} P(Y_n \leq y) = e^{-\alpha/y^k}.$$

Esercizio 6.26. Un lanciatore di giavellotto esegue $n \in \mathbb{N}$ lanci. Detta X_i la distanza ottenuta nell' i -esimo lancio, supponiamo che X_1, \dots, X_n siano variabili aleatorie i.i.d. con $X_n \sim \text{Exp}(\lambda)$, dove $\lambda \in (0, \infty)$. Indichiamo con M_n la massima distanza a cui è stato lanciato il giavellotto.

- (i) Sia $W_n := \frac{M_n}{\log(n)}$ e F_{W_n} la relativa funzione di ripartizione. Si mostri che

$$\lim_{n \rightarrow +\infty} F_{W_n}(x) = \begin{cases} 0 & \text{se } x < \frac{1}{\lambda} \\ e^{-1} & \text{se } x = \frac{1}{\lambda} \\ 1 & \text{se } x > \frac{1}{\lambda} \end{cases}.$$

- (ii) Si deduca che per ogni $\varepsilon > 0$

$$\lim_{n \rightarrow +\infty} P\left(\left|\frac{M_n}{\log(n)} - \frac{1}{\lambda}\right| > \varepsilon\right) = 0.$$

- (iii) Definiamo ora

$$Z_n := M_n - \frac{1}{\lambda} \log n,$$

e sia $F_{Z_n}(t)$ la relativa funzione di ripartizione. Si mostri che il limite $F(t) := \lim_{n \rightarrow \infty} F_{Z_n}(t)$ esiste, e lo si determini, per ogni $t \in \mathbb{R}$. Si osservi che F è la funzione di ripartizione di una variabile aleatoria assolutamente continua.

Esercizio 6.27. Pietro è un lanciatore di giavellotto. Dopo un lancio iniziale, in cui manda il giavellotto a una distanza X_0 , si cimenta in una successione di lanci ripetuti: nel lancio n -esimo il giavellotto cade a una distanza X_n . Pietro si interroga su quanti lanci T debba fare per migliorare il risultato iniziale, ossia

$$T := \min\{n \in \mathbb{N} : X_n > X_0\}.$$

Assumiamo che $(X_n)_{n \in \mathbb{N}_0}$ siano variabili aleatorie indipendenti e (ignorando l'effetto della fatica) con la stessa distribuzione, che supponiamo assolutamente continua. Mostriamo che T ha una distribuzione "universale", che non dipende dalla distribuzione delle X_k , con densità discreta

$$p_T(k) = \frac{1}{k(k+1)} \mathbb{1}_{\mathbb{N}}(k), \quad (6.96)$$

da cui segue che il numero di lanci richiesti per migliorarsi ha media $E(T) = +\infty$! (Aver ignorato la fatica rende questo risultato ancora più sorprendente.)

(i) Definendo gli eventi

$$A_k^{(n)} := \left\{ X_k = \max_{0 \leq i \leq n} X_i \right\}, \quad \text{per } n \in \mathbb{N}, k \in \{0, \dots, n\},$$

si mostri, con un argomento di simmetria, che

$$P(A_0^{(n)}) = P(A_1^{(n)}) = \dots = P(A_n^{(n)}).$$

(ii) Si mostri che per ogni $i \neq j$ si ha $P(X_i \neq X_j) = 1$.

[Sugg. Si noti che $X_i - X_j$ è una variabile reale assolutamente continua (perché?).]

(iii) Si spieghi l'inclusione di eventi $A_i^{(n)} \cap A_j^{(n)} \subseteq \{X_i = X_j\}$ e si deduca che

$$P(A_i^{(n)} \cap A_j^{(n)}) = 0, \quad \forall i \neq j.$$

(iv) Si spieghi perché $\bigcup_{k=0}^n A_k^{(n)} = \Omega$ e si deduca che

$$P(A_k^{(n)}) = \frac{1}{n+1}, \quad \forall n \in \mathbb{N}, 0 \leq k \leq n.$$

[Sugg. Si ricordi l'Esercizio 1.4.]

(v) Si spieghi perché $\{T > n\} = A_0^{(n)}$ e, dunque, $P(T > n) = \frac{1}{n+1}$. Si deduca la formula (6.96) e il fatto che $E(T) = +\infty$.

Esercizio 6.28. Siano $(X_n)_{n \in \mathbb{N}}$ variabili aleatorie i.i.d. $U(-\frac{1}{2}, \frac{1}{2})$.

(i) Si mostri che le variabili aleatorie $(Y_n := X_n + \frac{1}{2})_{n \in \mathbb{N}}$ sono i.i.d. $U(0, 1)$.

(ii) Applicando l'Esercizio 6.23, si deduca che per ogni $r > 0$

$$\lim_{n \rightarrow \infty} P\left(\min\{X_1, \dots, X_n\} < -\frac{1}{2} + \frac{r}{n}\right) = 1 - e^{-r}.$$

Si deduca che esistono $\bar{r} > 0$ e $n_0 < \infty$ tale che per ogni $n > n_0$

$$P\left(\min\{X_1, \dots, X_n\} < -\frac{1}{2} + \frac{\bar{r}}{n}\right) \geq 0.99.$$

(iii) Introduciamo per $n \in \mathbb{N}$ e $\delta > 0$ il sottoinsieme $C_n(\delta) \subseteq \mathbb{R}^n$ definito da

$$C_n(\delta) := \left(-\frac{1}{2}, \frac{1}{2}\right)^n \setminus \left(-\frac{1}{2} + \delta, \frac{1}{2} - \delta\right)^n,$$

che rappresenta la “buccia interna” di spessore δ del cubo n -dimensionale $\left(-\frac{1}{2}, \frac{1}{2}\right)^n$. Si mostri che per $n > n_0$

$$P\left((X_1, \dots, X_n) \in C_n\left(\frac{\bar{\tau}}{n}\right)\right) \geq 0.99.$$

(iv) (*) Si mostri che, per ogni $n \in \mathbb{N}$, il vettore aleatorio (X_1, \dots, X_n) ha distribuzione uniforme continua nell'insieme $\left(-\frac{1}{2}, \frac{1}{2}\right)^n \subseteq \mathbb{R}^n$ (si ricordi l'Esempio 6.39). Ricordando la definizione (6.46) di misura n -dimensionale, si osservi che $\text{mis}\left(\left(-\frac{1}{2}, \frac{1}{2}\right)^n\right) = 1$ e si deduca che per $n > n_0$

$$\text{mis}\left(C_n\left(\frac{\bar{\tau}}{n}\right)\right) \geq 99\%.$$

In altre parole, per n grande, “quasi tutto il volume di un cubo n -dimensionale è contenuto in una buccia sottile”, il cui spessore $\bar{\tau}/n$ tende a zero per $n \rightarrow \infty$!

Esercizio 6.29 (*). Sia (X, Y) un vettore aleatorio bidimensionale assolutamente continuo, con densità

$$f_{X,Y}(x, y) = \frac{\alpha(\alpha + 1)}{(1 + x + y)^{2+\alpha}} \mathbb{1}_{[0, \infty)}(x) \mathbb{1}_{[0, \infty)}(y),$$

dove $\alpha \in (0, \infty)$ è una costante fissata.

- (i) Senza fare conti, si spieghi perché le componenti X e Y hanno la stessa densità.
- (ii) Si mostri che la funzione di ripartizione di X (e di Y) è data da

$$F_X(t) = \left(1 - \frac{1}{(1+t)^\alpha}\right) \mathbb{1}_{[0, \infty)}(t).$$

(iii) Per quali valori di $p \in [1, \infty)$ si ha che $X \in L^p$?

Esercizio 6.30 (*). Sia (X, Y) un vettore aleatorio a valori in \mathbb{R}^2 , con densità

$$f_{X,Y}(x, y) := \begin{cases} c e^{-x} & \text{se } 0 < x < y < x + 1 \\ 0 & \text{altrimenti} \end{cases},$$

dove $c \in \mathbb{R}$ è una opportuna costante.

- (i) Si mostri che X è una variabile aleatoria con distribuzione $\text{Exp}(1)$ e si calcoli il valore della costante c .
- (ii) Si mostri che $Z := \log(X)$ è una variabile aleatoria assolutamente continua e se ne determini la densità. Per quali valori di p si ha $Z \in L^p$?
- (iii) Si determini la densità di Y . Si calcoli $E(e^{X-Y})$.

Esercizio 6.31 (*). Sia (X, Y) un vettore aleatorio bidimensionale con densità f data da

$$f(x, y) = cye^{-xy} \mathbb{1}_{[0, \infty) \times [0, 2]}(x, y).$$

- (i) Si determini il valore di $c \in \mathbb{R}$ affinché f sia effettivamente una densità.
- (ii) Si determinino le densità marginali di X e Y e si riconosca la distribuzione di Y . Le variabili aleatorie X e Y sono indipendenti?
- (iii) Si mostri che $V := \max(X, Y)$ è una variabile aleatoria reale assolutamente continua e se ne determini la densità.
- (iv) Posto $U := X + Y$, si dica se U e V sono indipendenti.

[Sugg. Non è necessario calcolare la densità congiunta di (U, V) .]

Esercizio 6.32 (*). Sia $Z := (X, Y)$ un vettore aleatorio bidimensionale con distribuzione uniforme nel sottoinsieme $C := ([0, \frac{1}{2}] \times [0, \frac{1}{2}]) \cup ([\frac{1}{2}, 1] \times [\frac{1}{2}, 1])$. Si determinino le distribuzioni delle variabili aleatorie reali X e Y . Esse sono indipendenti?

Esercizio 6.33 (*). Sia (X, Y) un vettore aleatorio bidimensionale con distribuzione uniforme sul cerchio di raggio unitario centrato nell'origine. Si determinino, possibilmente senza fare calcoli, le seguenti probabilità condizionali:

$$P\left(\max(|X|, |Y|) \leq \frac{1}{2\sqrt{2}} \mid X^2 + Y^2 \leq \frac{1}{4}\right), \quad P\left(\max(|X|, |Y|) \leq \frac{1}{2} \mid |X| + |Y| \leq 1\right).$$

[Sugg. Fare un disegno.]

Esercizio 6.34 (*). Un segnale viene trasmesso in un istante aleatorio X . Il ricevitore viene acceso in un istante aleatorio Y e resta acceso per un intervallo di tempo aleatorio Z . Supponendo che X, Y, Z siano variabili aleatorie indipendenti con $X \sim U[0, 2]$ e $Y, Z \sim U[0, 1]$, qual è la probabilità che il segnale venga ricevuto?

Esercizio 6.35 (*). Siano $\{X_n\}_{n \in \mathbb{N}_0}$ variabili aleatorie reali i.i.d. $\text{Exp}(1)$. Definiamo per $n \in \mathbb{N}$

$$U_n := \frac{X_0}{X_1 + \dots + X_n}.$$

- (i) Si mostri che la funzione di ripartizione di U_n è data, per $t \geq 0$, da

$$F_{U_n}(t) = 1 - \frac{1}{(1+t)^n}.$$

[Sugg. Si osservi che $Y_n := X_1 + \dots + X_n$ ha distribuzione ... ed è indipendente da ...]

- (ii) Si deduca che la variabile aleatoria U_n è assolutamente continua e, per ogni n fissato, si determini per quali valori di $p > 0$ si ha $U_n \in L^p$.

Esercizio 6.36 (*). Siano $X \sim N(0, 1)$ e $T \sim \text{Exp}(1/2)$ variabili aleatorie indipendenti. Si definiscano $Y = X/\sqrt{T}$ e $S = X^2 + T$.

- (i) Determinare la densità congiunta di (Y, S) .
- (ii) Y e S sono variabili aleatorie indipendenti?

Esercizio 6.37 (*). Siano X e Y variabili aleatorie reali indipendenti, in cui Y ha distribuzione $U(-\pi/2, \pi/2)$ mentre X è assolutamente continua con densità

$$f_X(x) = xe^{-\frac{x^2}{2}} 1_{[0, +\infty)}(x).$$

Si determinino la distribuzione congiunta e le distribuzioni marginali delle variabili aleatorie Z e W , definite da

$$Z := X \cos Y, \quad W := X \sin Y.$$

[Sugg. Si osservi che $X = \sqrt{Z^2 + W^2}$ e $Y = \arctan(W/Z)$.]

6.8 Note bibliografiche

L'*integrale di Riemann*, sviluppato da B. Riemann in [48], è stata la prima nozione rigorosa di integrale di una funzione su un intervallo. Un'analisi più completa e rigorosa delle variabili aleatorie assolutamente continue, a cominciare dalla dimostrazione della Proposizione 6.20 e di altri risultati enunciati in questo capitolo, richiede una nozione di integrale più avanzata, l'*integrale di Lebesgue*, introdotto da H. Lebesgue in [36]. Per un'introduzione all'integrazione secondo Lebesgue suggeriamo il testo di G.B. Folland [25].

Il *processo di Poisson* (Paragrafi 6.5.3 e 6.5.4) è l'esempio più elementare, e fondamentale, di *processo di punto*, cioè di sottoinsieme aleatorio discreto di \mathbb{R} . Ottime monografie su questo argomento sono i testi di D. J. Daley e D. Vere-Jones [16, 17] e di P. Brémaud [11].

I vettori aleatori normali (Paragrafo 6.6) hanno una naturale generalizzazione in famiglie infinite di variabili aleatorie, dette *processi gaussiani*. Su tale argomento segnaliamo la monografia di M. Lifshits [39].

Teoremi limite

Sommario. In questo capitolo presentiamo i teoremi limite classici, la *legge dei grandi numeri* e il *teorema limite centrale*, che costituiscono il nucleo del calcolo delle probabilità, per la loro portata sia teorica che applicativa. Le dimostrazioni di questi risultati sono accompagnate da una discussione dettagliata del loro significato e da numerosi esempi e applicazioni.

7.1 La legge dei grandi numeri

Consideriamo una successione $(X_i)_{i \in \mathbb{N}}$ di variabili aleatorie reali, definite nello stesso spazio di probabilità (Ω, \mathcal{A}, P) . Supponiamo che tali variabili aleatorie abbiano la stessa distribuzione marginale, con valor medio finito: in particolare $E(X_i) = \mu \in \mathbb{R}$ per ogni $i \in \mathbb{N}$. Per avere più chiaro il senso di ciò che segue, si può immaginare che le X_i rappresentino misurazioni successive di una stessa grandezza fisica, la cui aleatorietà è dovuta all'imprecisione degli strumenti di misura e/o a fenomeni di "disturbo" esterni; in alternativa, le X_i possono rappresentare i risultati ottenuti nelle ripetizioni di uno stesso esperimento aleatorio, come ad esempio i lanci successivi di un dado oppure le vincite alla roulette. Spesso è naturale richiedere che le variabili $(X_i)_{i \in \mathbb{N}}$ siano indipendenti, ma per il momento non facciamo tale ipotesi.

Avendo in mente gli esempi sopra descritti, è assai naturale considerare la media aritmetica dei risultati ottenuti: per ogni $n \in \mathbb{N}$ si definisce quindi

$$\bar{X}_n := \frac{1}{n} \sum_{i=1}^n X_i. \quad (7.1)$$

Nel linguaggio del calcolo delle probabilità, \bar{X}_n è detta *media campionaria*. Sottolineiamo che \bar{X}_n è *essa stessa una variabile aleatoria*, definita sullo stesso spazio di probabilità (Ω, \mathcal{A}, P) su cui sono definite le X_i . Per capire intuitivamente in che senso la variabile \bar{X}_n è aleatoria, si consideri l'esempio dei lanci successivi di un dado: ripetendo gli n lanci, si ottiene in generale un valore diverso di \bar{X}_n .

È naturale chiedersi quali valori possa assumere la variabile aleatoria \bar{X}_n e con

quali probabilità, ossia studiarne la distribuzione. Per l'effetto di compensazione della media aritmetica in (7.1), è plausibile che \bar{X}_n “abbia meno fluttuazioni” delle singole X_i – per lo meno se le variabili $(X_i)_{i \in \mathbb{N}}$ sono indipendenti – e ci si può aspettare che, per n grande, i “valori tipici” assunti da \bar{X}_n tendano a concentrarsi. Vedremo che questo è proprio ciò che accade, in un senso preciso, e il “centro” dei valori tipici di \bar{X}_n risulta essere il valor medio μ delle singole variabili X_i .

7.1.1 Enunciato, dimostrazione e discussione

Cominciamo a dare la *definizione* della proprietà che vogliamo dimostrare.

Definizione 7.1 (Legge dei grandi numeri). Sia data una successione $(X_i)_{i \in \mathbb{N}}$ di variabili aleatorie reali, definite nello stesso spazio di probabilità (Ω, \mathcal{A}, P) , tutte con lo stesso valor medio finito: $E(X_i) = \mu \in \mathbb{R}$ per ogni $i \in \mathbb{N}$. Diremo che la successione $(X_i)_{i \in \mathbb{N}}$ *soddisfa la legge debole dei grandi numeri* se

$$\forall \varepsilon > 0: \quad \lim_{n \rightarrow +\infty} P(|\bar{X}_n - \mu| > \varepsilon) = 0. \quad (7.2)$$

Osservazione 7.2. Esiste anche una versione “forte” della legge dei grandi numeri, che afferma che

$$P\left(\omega \in \Omega: \lim_{n \rightarrow +\infty} \bar{X}_n(\omega) = \mu\right) = 1. \quad (7.3)$$

È possibile mostrare che, se vale la relazione (7.3), vale anche la relazione (7.2), il che spiega la ragione degli aggettivi “debole” e “forte”. Nel seguito, parlando di “legge dei grandi numeri”, ci riferiremo sempre alla versione “debole”. \square

La legge debole dei grandi numeri afferma dunque che per $n \rightarrow \infty$ la media campionaria \bar{X}_n converge, nel senso preciso descritto da (7.2), verso il valor medio μ . Come abbiamo discusso alla fine del Paragrafo 3.3.1, questo risultato mostra l'importanza e la rilevanza della nozione di valor medio.

Resta da stabilire sotto quali condizioni sulla successione $(X_i)_{i \in \mathbb{N}}$ sia valida la legge dei grandi numeri. L'ipotesi più comunemente assunta è quella in cui le variabili $(X_i)_{i \in \mathbb{N}}$ sono tra loro indipendenti e hanno tutte la stessa distribuzione. Diremo, in tal caso, che $(X_i)_{i \in \mathbb{N}}$ è una successione di variabili aleatorie *indipendenti e identicamente distribuite* (i.i.d.). Possiamo allora enunciare:

Teorema 7.3 (Legge debole dei grandi numeri in L^1). Ogni successione $(X_i)_{i \in \mathbb{N}}$ di variabili aleatorie i.i.d., che ammettono valor medio finito, soddisfa la legge debole dei grandi numeri.

Il Teorema 7.3 costituisce la versione “classica” della legge debole dei grandi numeri. Noi ne dimostreremo una versione alternativa, con ipotesi diverse:

- rafforziamo la condizione che le variabili X_i abbiano valor medio finito, richiedendo che anche il *momento secondo* sia finito: $E(X_i^2) < \infty$, ossia $X_i \in L^2$, per ogni $i \in \mathbb{N}$ (si ricordi che $L^2 \subseteq L^1$, per la Proposizione 3.60);
- indeboliamo l’ipotesi che le variabili siano indipendenti, richiedendo solo che siano *scorrelate*, ossia $\text{Cov}(X_i, X_j) = 0$ se $i \neq j$ (si ricordi il Corollario 3.74);
- indeboliamo l’ipotesi che le variabili abbiano la stessa distribuzione, richiedendo solo che abbiano lo stesso valor medio e la stessa varianza.

Dimostriamo dunque il seguente risultato.

Proposizione 7.4 (Legge debole dei grandi numeri in L^2). *Siano $(X_i)_{i \in \mathbb{N}}$ variabili aleatorie reali con momento secondo finito, scorrelate, con lo stesso valor medio e la stessa varianza:*

$$X_i \in L^2, \quad E(X_i) = \mu \in \mathbb{R}, \quad \text{Var}(X_i) = \sigma^2 \in [0, \infty), \quad \forall i \in \mathbb{N};$$

$$\text{Cov}(X_i, X_j) = 0, \quad \forall i \neq j \in \mathbb{N}.$$

Allora la successione $(X_i)_{i \in \mathbb{N}}$ soddisfa la legge debole dei grandi numeri.

Dimostrazione. Si noti che la variabile aleatoria \bar{X}_n è in L^2 , in quanto combinazione lineare di variabili aleatorie in L^2 (si ricordi la Proposizione 3.61). Pertanto \bar{X}_n ammette valor medio e varianza finiti, che ora calcoliamo.

Per linearità del valor medio, da (7.1) si ottiene immediatamente

$$E(\bar{X}_n) = \mu. \quad (7.4)$$

Inoltre, per le proprietà della varianza descritte nella Proposizione 3.69,

$$\text{Var}(\bar{X}_n) = \frac{1}{n^2} \left(\sum_{i=1}^n \text{Var}(X_i) + \sum_{1 \leq i \neq j \leq n} \text{Cov}(X_i, X_j) \right) = \frac{\sigma^2}{n}, \quad (7.5)$$

avendo usato il fatto che $\text{Var}(X_i) = \sigma^2$ e $\text{Cov}(X_i, X_j) = 0$ se $i \neq j$.

Applicando la disuguaglianza di Chebyshev alla variabile aleatoria \bar{X}_n (si ricordi il Teorema 3.77), otteniamo

$$P(|\bar{X}_n - \mu| \geq \varepsilon) \leq \frac{\text{Var}(\bar{X}_n)}{\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2}, \quad (7.6)$$

da cui la tesi segue immediatamente. \square

Osservazione 7.5. Nella Proposizione 7.4, l’ipotesi che le variabili X_i abbiano la stessa varianza può essere notevolmente indebolita. Infatti, ponendo $\sigma_i^2 := \text{Var}(X_i)$, è sufficiente richiedere che $\frac{1}{n^2} \sum_{i=1}^n \sigma_i^2 \rightarrow 0$ per $n \rightarrow \infty$, come il lettore può facilmente verificare. In particolare, è sufficiente che tutte le σ_i^2 siano maggiorate da una stessa costante finita. \square

Sottolineiamo che il cuore della dimostrazione consiste nell'osservare che la variabile aleatoria \bar{X}_n ha lo stesso valor medio μ delle singole variabili X_i , ma *la sua varianza σ^2/n è molto più piccola*, se n è grande. Ricordando l'interpretazione della varianza, o meglio della sua radice quadrata, come misura della larghezza della distribuzione, descritta nel Paragrafo 3.3.3, è intuitivamente chiaro che la distribuzione di \bar{X}_n tende a concentrarsi attorno al valore μ , per n grande.

Le ipotesi che le variabili X_i siano scorrelate e abbiano lo stesso valor medio e la stessa varianza sono automaticamente soddisfatte per una successione $(X_i)_{i \in \mathbb{N}}$ di variabili aleatorie i.i.d.. Pertanto possiamo enunciare il seguente corollario (più debole del Teorema 7.3, ma di cui abbiamo fornito la dimostrazione).

Corollario 7.6. *Siano $(X_i)_{i \in \mathbb{N}}$ variabili aleatorie reali i.i.d. con momento secondo finito. Allora $(X_i)_{i \in \mathbb{N}}$ soddisfa la legge debole dei grandi numeri.*

Consideriamo infine un caso molto speciale, ma particolarmente importante. Supponiamo che $(C_i)_{i \in \mathbb{N}}$ sia una successione di eventi *indipendenti* di uno spazio di probabilità (Ω, \mathcal{A}, P) , tutti con la stessa probabilità: $P(C_i) = p \in [0, 1]$, per ogni $i \in \mathbb{N}$. In altri termini, gli eventi C_i costituiscono una successione infinita di prove ripetute e indipendenti con probabilità di successo p . Le corrispondenti funzioni indicatrici $X_i := \mathbb{1}_{C_i}$ sono variabili aleatorie indipendenti con la stessa distribuzione di Bernoulli: $X_i \sim \text{Be}(p)$ per ogni $i \in \mathbb{N}$. Le ipotesi del Corollario 7.6 sono verificate, dunque la successione $(X_i)_{i \in \mathbb{N}}$ soddisfa la legge debole dei grandi numeri. In questo caso conviene riscrivere la media campionaria nel modo seguente:

$$\bar{X}_n = \frac{S_n}{n}, \quad \text{dove} \quad S_n := \sum_{i=1}^n \mathbb{1}_{C_i},$$

così che la relazione (7.2) può essere riformulata nel modo seguente:

$$\forall \varepsilon > 0: \quad \lim_{n \rightarrow +\infty} P\left(\left|\frac{S_n}{n} - p\right| > \varepsilon\right) = 0. \quad (7.7)$$

La variabile aleatoria S_n ha un significato molto chiaro: essa conta *quanti tra gli eventi C_1, \dots, C_n si sono verificati*. Di conseguenza, la legge debole dei grandi numeri afferma che, se n è sufficientemente grande, *la frazione di successi ottenuti è vicina alla probabilità di singolo successo p , con grande probabilità*. Ciò fornisce, in particolare, una giustificazione a posteriori della nozione intuitiva di probabilità di un evento, come frazione asintotica del numero di volte in cui l'evento si verifica in una successione di ripetizioni indipendenti dell'esperimento aleatorio.

Osservazione 7.7. Sappiamo che la variabile aleatoria S_n ha distribuzione $\text{Bin}(n, p)$, ossia $P(S_n = k) = \binom{n}{k} p^k (1-p)^{n-k}$ (si ricordi la relazione (3.14)). Pertanto, possiamo riscrivere la relazione (7.7) come segue:

$$\forall \varepsilon > 0: \quad \lim_{n \rightarrow \infty} \left\{ \sum_{k \in \{0, \dots, n\}: |k-np| > n\varepsilon} \binom{n}{k} p^k (1-p)^{n-k} \right\} = 0. \quad (7.8)$$

Abbiamo dunque una riformulazione “elementare” della legge debole dei grandi numeri per prove ripetute e indipendenti, in cui le variabili aleatorie non sono neppure menzionate. Sebbene sia possibile dimostrare la relazione (7.8) con stime dirette, la dimostrazione che abbiamo dato del Teorema 7.4, basata sull’uso di variabili aleatorie, valor medio e disuguaglianze, risulta più semplice, elegante e “trasparente”. \square

Osservazione 7.8 (Numeri ritardatari). La legge dei grandi numeri è spesso citata, a sproposito, per supportare la credenza che puntare sui numeri ritardatari al Lotto dia maggiori possibilità di vittoria. L’argomento tipico è il seguente.

Consideriamo ad esempio la “ruota di Venezia” del Lotto, in cui ogni settimana vengono estratti “a caso” 5 numeri distinti tra 1 e 90. La probabilità che un numero fissato, diciamo il 37, sia tra i cinque numeri estratti vale $p = \frac{1}{18}$ (esercizio). Supponiamo di osservare che per k settimane consecutive, con $k \gg 18$, non viene mai stato estratto il numero 37. Sappiamo che in un grande numero di settimane il numero 37 verrà estratto una frazione di volte vicina a $p = \frac{1}{18}$, con grande probabilità, per la legge dei grandi numeri. Si potrebbe allora essere tentati di concludere che nelle settimane successive alla k -esima il numero 37 dovrà essere estratto con maggiore frequenza, in modo da “riequilibrare” la media.

Questo ragionamento è scorretto. Infatti, se le estrazioni sono indipendenti, l’informazione che non è stato mai estratto il numero 37 nelle prime k settimane non cambia la probabilità che il 37 venga estratto nelle settimane successive. Ciò non è affatto in contraddizione con la legge dei grandi numeri (7.7), che continua a valere anche sapendo che $S_k = 0$, ossia rimpiazzando P con $P(\cdot | S_k = 0)$. La ragione intuitiva è che per k fissato, non importa quanto “grande”, il valore di S_k ha un’influenza trascurabile sul comportamento asintotico di S_n per $n \rightarrow \infty$. Lasciamo al lettore interessato il compito di formalizzare e dimostrare queste affermazioni. \square

7.1.2 Il metodo Monte Carlo per il calcolo di integrali

Per avere un’idea della portata applicativa della legge dei grandi numeri, consideriamo il seguente problema. Sia $f: [a, b] \rightarrow \mathbb{R}$ una funzione integrabile secondo Riemann (in particolare, f è limitata). Le funzioni il cui integrale è calcolabile esattamente con metodi analitici sono, in realtà, abbastanza poche, perciò è importante avere metodi numerici per il calcolo approssimato di tale integrale.

I metodi più comuni consistono nel “discretizzare” l’intervallo $[a, b]$, approssimando ad esempio $\int_a^b f(x) dx$ con la somma di Riemann

$$\frac{b-a}{N} \sum_{i=1}^N f\left(a + i \frac{b-a}{N}\right),$$

dove $(b-a)/N$ è il “passo” della discretizzazione. Tale approssimazione può essere migliorata, in particolare quando siano note ulteriori informazioni su f come, ad esempio, la sua derivata. Per una f sufficientemente regolare è possibile, fissata a priori un’arbitraria soglia di errore, determinare *esattamente* quanto grande dev’essere N per garantire di non superare tale soglia.

Gli algoritmi *stocastici*, che utilizzano generatori di numeri casuali, sono basati su un principio diverso. Vengono fissati due numeri: la soglia di errore e la massima probabilità tollerabile di commettere un errore maggiore della soglia data. Non si pretende dunque la *certezza* di commettere un errore piccolo, ma soltanto che sia estremamente improbabile commettere un errore maggiore della soglia fissata.

Tornando ai dettagli del problema in esame, siano $(X_i)_{i \in \mathbb{N}}$ variabili aleatorie indipendenti, con $X_i \sim U(a, b)$. Per ogni $N \in \mathbb{N}$, possiamo pensare a X_1, \dots, X_N come a N “numeri casuali” generati con distribuzione uniforme in $[a, b]$. Si noti che

$$\mathbb{E}[f(X_i)] = \frac{1}{b-a} \int_a^b f(x) dx, \quad \forall i \in \mathbb{N},$$

grazie alla Proposizione 6.21. Applicando la legge debole dei grandi numeri alle variabili aleatorie $f(X_1), \dots, f(X_N)$ (che sono i.i.d. e limitate, dunque con tutti i momenti finiti) otteniamo allora

$$\lim_{N \rightarrow +\infty} \mathbb{P} \left[\left| \frac{1}{N} \sum_{i=1}^N f(X_i) - \frac{1}{b-a} \int_a^b f(x) dx \right| > \varepsilon \right] = 0, \quad \forall \varepsilon > 0.$$

Quindi, se N è sufficientemente grande, la quantità aleatoria $\frac{1}{N} \sum_{i=1}^N f(X_i)$ fornisce con grande probabilità un’ottima approssimazione dell’integrale $\frac{1}{b-a} \int_a^b f(x) dx$.

Si può dire di più. Sia ε la massima soglia di errore che si vuole tollerare nel calcolo dell’integrale, e sia $\delta > 0$ la probabilità con cui si accetta di compiere un errore maggiore di ε . Vogliamo capire quanti numeri casuali è necessario generare, ossia quanto grande deve essere N , affinché

$$\mathbb{P} \left[\left| \frac{1}{N} \sum_{i=1}^N f(X_i) - \frac{1}{b-a} \int_a^b f(x) dx \right| > \varepsilon \right] \leq \delta. \quad (7.9)$$

Dato che f è limitata, possiamo introdurre la costante finita

$$M := \|f\|_\infty := \sup_{x \in [a, b]} |f(x)| < \infty,$$

grazie a cui possiamo stimare

$$\sigma^2 := \text{Var}[f(X_1)] \leq \mathbb{E}[(f(X_1))^2] \leq M^2.$$

Ricordando la relazione (7.6), il membro sinistro in (7.9) è maggiorato da $\sigma^2/(\varepsilon^2 N)$. Ne segue che la disuguaglianza (7.9) è verificata se

$$\frac{M^2}{\varepsilon^2 N} \leq \delta \quad \Longleftrightarrow \quad N \geq \frac{M^2}{\delta \varepsilon^2}. \quad (7.10)$$

Abbiamo dunque ottenuto una stima esplicita: se generiamo almeno $M^2/\delta \varepsilon^2$ numeri casuali, sappiamo che con probabilità maggiore o uguale a $1 - \delta$ la quantità aleatoria $\frac{1}{N} \sum_{i=1}^N f(X_i)$ dista non più di ε dall’integrale $\frac{1}{b-a} \int_a^b f(x) dx$.

Questo metodo per il calcolo approssimato di integrali definiti ha il vantaggio di essere molto facile da implementare, in quanto richiede solo un generatore di numeri casuali con distribuzione uniforme. Tuttavia, benché le disuguaglianze in (7.10) possano essere migliorate, per ottenere una precisione accettabile è necessario generare molti numeri casuali, il che rende questo metodo meno efficiente degli algoritmi “deterministici”. Questo discorso cambia radicalmente quando si tratta di calcolare integrali di funzioni di molte variabili. In tal caso, esistono varianti multidimensionali dell’algoritmo appena descritto che risultano, in dimensione elevata, più efficienti degli algoritmi deterministici.

7.1.3 Il teorema di approssimazione di Weierstrass

Come applicazione della legge dei grandi numeri – o, più direttamente, della disuguaglianza di Chebyshev – presentiamo una dimostrazione elegante e *costruttiva*, dovuta a Bernstein, di un celebre teorema di approssimazione di Weierstrass, che afferma che ogni funzione continua definita su un intervallo chiuso e limitato di \mathbb{R} può essere approssimata uniformemente da una successione di polinomi.

A meno di semplici trasformazioni, non costa nulla supporre che l’intervallo di definizione della funzione sia $[0, 1]$. A ogni funzione $f : [0, 1] \rightarrow \mathbb{R}$ associamo la seguente successione di polinomi, detti *polinomi di Bernstein di f* :

$$p_n^f(x) := \sum_{k=0}^n \binom{n}{k} f\left(\frac{k}{n}\right) x^k (1-x)^{n-k}.$$

Teorema 7.9 (Weierstrass-Bernstein). *Per ogni funzione $f : [0, 1] \rightarrow \mathbb{R}$ continua, la successione dei polinomi di Bernstein di f converge uniformemente verso f :*

$$\lim_{n \rightarrow +\infty} \sup_{x \in [0,1]} |f(x) - p_n^f(x)| = 0.$$

Dimostrazione. Ricordiamo che una funzione continua su un intervallo chiuso e limitato di \mathbb{R} è limitata e uniformemente continua, cioè

$$M := \sup_{x \in [0,1]} |f(x)| < +\infty, \quad (7.11)$$

$$\forall \varepsilon > 0 \exists \delta_\varepsilon > 0 : |y - x| \leq \delta_\varepsilon \implies |f(y) - f(x)| \leq \frac{\varepsilon}{2}, \quad (7.12)$$

dove in (7.11) abbiamo scelto $\frac{\varepsilon}{2}$ anziché ε per comodità futura.

Fissiamo $x \in [0, 1]$. Siano X_1, \dots, X_n variabili aleatorie indipendenti, definite su un opportuno spazio di probabilità (Ω, \mathcal{P}) , con $X_i \sim \text{Be}(x)$. Allora

$$\mu := E(X_i) = x, \quad \sigma^2 := \text{Var}(X_i) = x(1-x).$$

Ricordando la formula (3.30), possiamo rappresentare il polinomio di Bernstein

$p_n^f(x)$ in termini della media campionaria $\bar{X}_n := \frac{1}{n}(X_1 + \dots + X_n)$:

$$p_n^f(x) = E[f(\bar{X}_n)]. \quad (7.13)$$

Intuitivamente, se n è grande, per la legge dei grandi numeri la variabile aleatoria \bar{X}_n è con grande probabilità vicino al suo valor medio x , pertanto grazie a (7.13) dovrebbe essere chiaro che $p_n^f(x)$ è vicino a $f(x)$, essendo f continua.

Per formalizzare il ragionamento, introduciamo l'evento

$$A_{n,\varepsilon} := \{|\bar{X}_n - x| \geq \delta_\varepsilon\} \subseteq \Omega,$$

dove δ_ε è lo stesso che in (7.12). Per la relazione (7.6),

$$P(A_{n,\varepsilon}) \leq \frac{x(1-x)}{n\delta_\varepsilon^2} \leq \frac{1}{4n\delta_\varepsilon^2}, \quad (7.14)$$

dove si è usato il fatto che la funzione $x \mapsto x(1-x)$ ha massimo in $x = 1/2$. Grazie alla relazione (7.13), si ha

$$|f(x) - p_n^f(x)| = |f(x) - E[f(\bar{X}_n)]| = |E[f(x) - f(\bar{X}_n)]| \leq E[|f(x) - f(\bar{X}_n)|],$$

per la linearità del valor medio e la proprietà (3.34). Dato che $\mathbb{1}_{A_{n,\varepsilon}} + \mathbb{1}_{A_{n,\varepsilon}^c} = 1$, si ha

$$E[|f(x) - f(\bar{X}_n)|] = E[|f(x) - f(\bar{X}_n)| \cdot \mathbb{1}_{A_{n,\varepsilon}}] + E[|f(x) - f(\bar{X}_n)| \cdot \mathbb{1}_{A_{n,\varepsilon}^c}].$$

Consideriamo separatamente i due termini nel membro destro. Ricordando (7.11), si ha $|f(x) - f(\bar{X}_n)| \leq |f(x)| + |f(\bar{X}_n)| \leq 2M$, pertanto grazie a (7.14)

$$E[|f(x) - f(\bar{X}_n)| \cdot \mathbb{1}_{A_{n,\varepsilon}}] \leq 2ME[\mathbb{1}_{A_{n,\varepsilon}}] = 2MP(A_{n,\varepsilon}) \leq \frac{M}{2n\delta_\varepsilon^2}.$$

Ora osserviamo che, se $\omega \in A_{n,\varepsilon}^c$, si ha per costruzione $|\bar{X}_n(\omega) - x| < \delta_\varepsilon$ e quindi $|f(x) - f(\bar{X}_n)| \leq \frac{\varepsilon}{2}$, grazie alla relazione (7.12), per cui

$$E[|f(x) - f(\bar{X}_n)| \cdot \mathbb{1}_{A_{n,\varepsilon}^c}] \leq \frac{\varepsilon}{2} E[\mathbb{1}_{A_{n,\varepsilon}^c}] = \frac{\varepsilon}{2} P(A_{n,\varepsilon}^c) \leq \frac{\varepsilon}{2}.$$

Mettendo insieme i vari pezzi, otteniamo

$$|f(x) - p_n^f(x)| \leq \frac{M}{2n\delta_\varepsilon^2} + \frac{\varepsilon}{2}, \quad \forall x \in [0, 1], \varepsilon > 0, n \in \mathbb{N}.$$

Osserviamo che il membro destro non dipende da x . Di conseguenza, definendo $n_\varepsilon := M/(\delta_\varepsilon^2 \varepsilon)$, abbiamo mostrato che per ogni $\varepsilon > 0$ esiste $n_\varepsilon < \infty$ tale che

$$\sup_{x \in [0, 1]} |f(x) - p_n^f(x)| \leq \varepsilon, \quad \forall n \geq n_\varepsilon,$$

che è esattamente la tesi. \square

7.1.4 Un esempio con variabili aleatorie correlate

Supponiamo di avere n oggetti, ciascuno dei quali viene inserito casualmente in uno di r cassette. Questo modello, introdotto nell'Esempio 1.41 e ripreso negli Esempi 3.98 e 3.105, può essere descritto semplicemente considerando n variabili aleatorie U_1, \dots, U_n indipendenti e con la stessa distribuzione, uniforme nell'insieme $\{1, \dots, r\}$. Intuitivamente, la variabile aleatoria U_i descrive il cassetto in cui viene inserito l'oggetto i -esimo.

In questo paragrafo vogliamo capire quanti sono tipicamente i cassette che contengono un numero fissato k di oggetti, precisando un problema posto alla fine del Paragrafo 3.5.5. Fissiamo d'ora in avanti $k \in \mathbb{N}$ e introduciamo le variabili aleatorie di Bernoulli X_1, \dots, X_r , dove X_i assume il valore 1 se l' i -esimo cassetto contiene esattamente k oggetti, mentre assume il valore 0 altrimenti. Volendo esprimere X_i in funzione di U_1, \dots, U_n , possiamo scrivere

$$X_i := \mathbb{1}_{\{\omega \in \Omega: |\{\ell \in \{1, \dots, n\}: U_\ell(\omega) = i\}| = k\}}.$$

La variabile aleatoria $S := X_1 + \dots + X_r$, già introdotta nel Paragrafo 3.5.5, conta il numero di cassette che contengono esattamente k oggetti. Di conseguenza, la *frazione* di cassette che contengono esattamente k oggetti è data dalla media campionaria

$$\bar{X}_r = \frac{S}{r} = \frac{1}{r} \sum_{i=1}^r X_i, \quad (7.15)$$

di cui vogliamo studiare le proprietà quando r è grande.

Le variabili aleatorie X_1, \dots, X_r hanno la stessa distribuzione: $X_i \sim \text{Be}(p_{n,r,k})$, dove $p_{n,r,k}$ è la probabilità che k oggetti finiscano in un dato cassetto, ossia (ricordando l'Esempio 3.105)

$$p_{n,r,k} = \text{pBin}(n, \frac{1}{r})(k) = \binom{n}{k} \frac{1}{r^k} \left(1 - \frac{1}{r}\right)^{n-k}. \quad (7.16)$$

Sottolineiamo che le variabili aleatorie X_1, \dots, X_r *non sono indipendenti, e nemmeno scorrelate*, come vedremo. Inoltre, il valor medio $\mu = E(X_i) = p_{n,r,k}$ dipende da r e n (oltre che da k).

Consideriamo tuttavia il caso particolarmente interessante in cui $n, r \gg 1$ e $n/r \approx \lambda \in (0, \infty)$. In questo regime, per la “legge dei piccoli numeri” (Proposizione 3.108) si ha che

$$p_{n,r,k} \approx \text{pPois}(\lambda)(k) = e^{-\lambda} \frac{\lambda^k}{k!},$$

dunque il valor medio $E(X_i) = p_{n,r,k}$ si stabilizza; inoltre le variabili X_i tendono a diventare scorrelate, come mostriamo precisamente più sotto. Di conseguenza, possiamo dimostrare una versione della legge debole dei grandi numeri anche in questo contesto.

Proposizione 7.10. *Sia $(n_r)_{r \in \mathbb{N}}$ una successione a valori in \mathbb{N} tale che*

$$\lim_{r \rightarrow \infty} \frac{n_r}{r} = \lambda \in (0, \infty).$$

Allora, per ogni $k \in \mathbb{N}_0$ fissato, vale che

$$\forall \varepsilon > 0: \quad \lim_{r \rightarrow \infty} \text{P} \left(\left| \frac{S}{r} - e^{-\lambda} \frac{\lambda^k}{k!} \right| \geq \varepsilon \right) = 0. \quad (7.17)$$

Dimostrazione. Il cuore della dimostrazione consiste nell'ottenere una stima sulla covarianza $\text{Cov}(X_i, X_j)$ per $i \neq j$. Una volta ottenuta questa stima, la dimostrazione procede in modo del tutto analogo a quella della Proposizione 7.4.

Per ragioni di simmetria, $E(X_i X_j) = E(X_1 X_2)$ per ogni $i \neq j$. Inoltre, ricordando la definizione delle variabili X_i , la variabile aleatoria $X_1 X_2$ assume il valore 1 se e solo se i cassetti "1" e "2" contengono entrambi k oggetti, e assume il valore 0 altrimenti. Pertanto, *il valore atteso $E(X_1 X_2)$ coincide con la probabilità che i cassetti "1" e "2" contengano ciascuno k oggetti.*

Calcoliamo tale probabilità. Indicando con $I, J \subseteq \{1, \dots, n\}$ rispettivamente gli oggetti contenuti nei cassetti "1" e "2", e ricordando che l'oggetto i è contenuto nel cassetto U_i , si ha

$$E(X_1 X_2) = \sum_{\substack{I, J \subseteq \{1, \dots, n\} \\ |I|=k, |J|=k, I \cap J = \emptyset}} P(U_i = 1 \forall i \in I, U_j = 2 \forall j \in J, U_k \notin \{1, 2\} \forall k \in \{1, \dots, n\} \setminus (I \cup J)).$$

Dato che le variabili aleatorie U_ℓ hanno distribuzione uniforme nell'insieme $\{1, \dots, r\}$, si ha

$$P(U_\ell = 1) = \frac{1}{r}, \quad P(U_\ell = 2) = \frac{1}{r}, \quad P(U_\ell \notin \{1, 2\}) = \frac{r-2}{r},$$

pertanto, per l'indipendenza di U_1, \dots, U_n , si ottiene

$$E(X_1 X_2) = \sum_{\substack{I, J \subseteq \{1, \dots, n\} \\ |I|=k, |J|=k, I \cap J = \emptyset}} \left(\frac{1}{r}\right)^{|I|} \left(\frac{1}{r}\right)^{|J|} \left(\frac{r-2}{r}\right)^{n-|I|-|J|} = \binom{n}{k} \binom{n-k}{k} \frac{(r-2)^{n-2k}}{r^n},$$

dove $\binom{n}{k}$ è il numero di sottoinsiemi $I \subseteq \{1, \dots, n\}$ di k elementi, mentre $\binom{n-k}{k}$ è il numero di sottoinsiemi $J \subseteq \{1, \dots, n\} \setminus I$ di k elementi.

Riscriviamo l'espressione ottenuta per $E(X_1 X_2)$ nel modo seguente:

$$\begin{aligned} E(X_1 X_2) &= \binom{n}{k} \frac{1}{r^k} \left(1 - \frac{1}{r}\right)^{n-k} \binom{n-k}{k} \frac{1}{(r-1)^k} \left(1 - \frac{1}{r-1}\right)^{n-2k} \\ &= P_{\text{Bin}(n, \frac{1}{r})}(k) P_{\text{Bin}(n-k, \frac{1}{r-1})}(k). \end{aligned}$$

Questa formula risulta piuttosto intuitiva: infatti già sapevamo che la probabilità che un cassetto contenga k oggetti vale $P_{\text{Bin}(n, \frac{1}{r})}(k)$; se il primo cassetto contiene k oggetti, restano $n-k$ oggetti e $r-1$ cassetti disponibili, pertanto la probabilità (condizionale) che il secondo cassetto contenga k oggetti risulta essere $P_{\text{Bin}(n-k, \frac{1}{r-1})}(k)$. Dalla formula ottenuta, segue che

$$\text{Cov}(X_1, X_2) = E(X_1 X_2) - E(X_1) E(X_2) = P_{\text{Bin}(n, \frac{1}{r})}(k) (P_{\text{Bin}(n-k, \frac{1}{r-1})}(k) - P_{\text{Bin}(n, \frac{1}{r})}(k)). \quad (7.18)$$

Sottolineiamo che in generale $\text{Cov}(X_1, X_2) \neq 0$.

Vogliamo ora dare una stima su $|\text{Cov}(X_1, X_2)|$. Invece di fare calcoli espliciti, usiamo i risultati ottenuti nel Paragrafo 4.1: per il Teorema 4.5 e il Lemma 4.4 si ha

$$\begin{aligned} |P_{\text{Bin}(n-k, \frac{1}{r-1})}(k) - P_{\text{Pois}(\frac{n-k}{r-1})}(k)| &\leq \frac{n-k}{(r-1)^2} \leq \frac{n}{(r-1)^2}, \\ |P_{\text{Pois}(\frac{n-k}{r-1})}(k) - P_{\text{Pois}(\frac{n}{r})}(k)| &\leq \left| \frac{n}{r} - \frac{n-k}{r-1} \right| \leq \frac{n+kr}{r(r-1)}, \\ |P_{\text{Pois}(\frac{n}{r})}(k) - P_{\text{Bin}(n, \frac{1}{r})}(k)| &\leq \frac{n}{r^2}. \end{aligned}$$

Dalla relazione (7.18), stimando $p_{\text{Bin}(n, \frac{1}{r})}(k) \leq 1$ e applicando la disuguaglianza triangolare, si ottiene

$$\begin{aligned} |\text{Cov}(X_1, X_2)| &\leq |p_{\text{Bin}(n-k, \frac{1}{r-1})}(k) - p_{\text{Pois}(\frac{n-k}{r-1})}(k)| + |p_{\text{Pois}(\frac{n-k}{r-1})}(k) - p_{\text{Pois}(\frac{n}{r})}(k)| \\ &\quad + |p_{\text{Pois}(\frac{n}{r})}(k) - p_{\text{Bin}(n, \frac{1}{r})}(k)| \leq \frac{n}{(r-1)^2} + \frac{n+kr}{r(r-1)} + \frac{n}{r^2}. \end{aligned} \quad (7.19)$$

Sia ora $n = n_r$ come nelle ipotesi: $n_r/r \rightarrow \lambda \in (0, \infty)$ per $r \rightarrow \infty$, ossia $n_r \sim \lambda r$. Asintoticamente, il primo e il terzo termine nel membro destro in (7.19) sono equivalenti a λ/r , mentre il secondo è equivalente a $(\lambda + k)/r$. Pertanto, per ogni costante $C > 3\lambda + k$, esiste $r_0 < \infty$ tale che

$$|\text{Cov}(X_1, X_2)| \leq \frac{C}{r}, \quad \forall r \geq r_0. \quad (7.20)$$

Consideriamo infine la media campionaria \bar{X}_r in (7.15). Ricordando la relazione (7.15), dato che $\text{Cov}(X_i, X_j) = \text{Cov}(X_1, X_2)$ per ogni $i \neq j$, si ottiene

$$\begin{aligned} \text{Var}(\bar{X}_r) &= \frac{1}{r^2} \left(\sum_{i=1}^r \text{Var}(X_i) + \sum_{1 \leq i \neq j \leq r} \text{Cov}(X_i, X_j) \right) \\ &= \frac{1}{r} \text{Var}(X_1) + \frac{r-1}{r} \text{Cov}(X_1, X_2) \leq \frac{1}{r} + \frac{C}{r}, \end{aligned}$$

per ogni $r \geq r_0$, avendo usato la relazione (7.20) e il fatto che $\text{Var}(X_1) \leq E(X_1^2) \leq 1$, perché $|X_1| \leq 1$. Per la relazione (7.16)

$$E(\bar{X}_r) = E(X_1) = p_{\text{Bin}(n_r, \frac{1}{r})}(k), \quad (7.21)$$

quindi, applicando la disuguaglianza di Chebyshev (Teorema 3.77), per ogni $\varepsilon > 0$

$$P(|\bar{X}_r - p_{\text{Bin}(n_r, \frac{1}{r})}(k)| \geq 2\varepsilon) \leq \frac{\text{Var}(\bar{X}_r)}{(2\varepsilon)^2} \leq \frac{C+1}{(2\varepsilon)^2 r} \rightarrow 0 \quad \text{per } r \rightarrow \infty.$$

Ricordiamo ora la Proposizione 3.108 (legge dei piccoli numeri):

$$\lim_{r \rightarrow \infty} p_{\text{Bin}(n_r, \frac{1}{r})}(k) = p_{\text{Pois}(\lambda)}(k) = e^{-\lambda} \frac{\lambda^k}{k!}, \quad \forall k \in \mathbb{N}_0,$$

da cui segue l'inclusione di eventi $\{|\bar{X}_r - e^{-\lambda} \frac{\lambda^k}{k!}| \geq \varepsilon\} \subseteq \{|\bar{X}_r - p_{\text{Bin}(n_r, \frac{1}{r})}(k)| \geq 2\varepsilon\}$, per r sufficientemente grande. Questo completa la dimostrazione della relazione (7.17). \square

Esercizi

Esercizio 7.1. Sia $(X_i)_{i \in \mathbb{N}}$ una successione di variabili aleatorie, con valor medio μ finito, che soddisfa la legge debole dei grandi numeri (si veda (7.2)). Si mostri che per ogni funzione $f: \mathbb{R} \rightarrow \mathbb{R}$ limitata e continua nel punto μ si ha

$$\lim_{n \rightarrow +\infty} E[f(\bar{X}_n)] = f(\mu).$$

7.2 Il teorema limite centrale

La legge debole dei grandi numeri afferma che, sotto opportune ipotesi, $\bar{X}_n \rightarrow \mu$, o equivalentemente $(\bar{X}_n - \mu) \rightarrow 0$, nel senso preciso dato dalla relazione (7.2).

È naturale chiedersi “con che velocità” la differenza $(\bar{X}_n - \mu)$ tende a zero. Per rispondere a questa domanda si può cercare di amplificare tale differenza, ossia di moltiplicarla per un’opportuna successione reale $(c_n)_{n \in \mathbb{N}}$ che tende all’infinito, in modo che il prodotto $Z_n = c_n(\bar{X}_n - \mu)$ “converga verso un limite non degenerare, ossia non zero né infinito”. Abbiamo usato le virgolette perché, essendo la successione Z_n aleatoria, non è chiaro in che senso si debba interpretare il concetto di limite.

Vedremo in questo paragrafo che, sotto le ipotesi del Corollario 7.6, quanto descritto sopra è proprio quello che accade, con la scelta $c_n \approx \sqrt{n}$ e con una variabile normale come limite. Questo è il contenuto del celebre teorema limite centrale.

7.2.1 Enunciato e discussione

Sia $(X_i)_{i \in \mathbb{N}}$ una successione i.i.d. di variabili aleatorie reali, con momento secondo finito. Il valor medio sarà indicato con $\mu := E(X_i)$ e la varianza, *che assumeremo strettamente positiva*, con $\sigma^2 := \text{Var}(X_i) \in (0, \infty)$.

La media campionaria \bar{X}_n , definita in (7.1), ha valor medio $E(\bar{X}_n) = \mu$ e varianza $\text{Var}(\bar{X}_n) = \sigma^2/n$, come mostrato in (7.4) e (7.5). Pertanto, per la linearità del valor medio e per le proprietà della varianza (si ricordi la Proposizione 3.69),

$$E(\bar{X}_n - \mu) = 0, \quad \text{Var}(\bar{X}_n - \mu) = \text{Var}(\bar{X}_n) = \frac{\sigma^2}{n}.$$

Intuitivamente, per n grande, $(\bar{X}_n - \mu)$ è una quantità “piccola”, avendo media nulla e varianza che tende a zero. Possiamo allora amplificarla, moltiplicandola per il fattore \sqrt{n}/σ , così da ottenere una quantità con varianza unitaria: definendo

$$Z_n := \frac{\sqrt{n}}{\sigma}(\bar{X}_n - \mu), \quad (7.22)$$

sempre per la Proposizione 3.69 si ha

$$E(Z_n) = 0, \quad \text{Var}(Z_n) = 1. \quad (7.23)$$

A questo punto è naturale chiedersi se la successione Z_n converga in qualche senso per $n \rightarrow \infty$. Il teorema limite centrale fornisce una risposta affermativa, mostrando che la funzione di ripartizione F_{Z_n} converge verso un limite non degenerare, che risulta essere la funzione di ripartizione della normale standard $N(0, 1)$.

Teorema 7.11 (Teorema limite centrale). *Sia $(X_i)_{i \in \mathbb{N}}$ una successione di variabili aleatorie i.i.d. che ammettono momento secondo finito, con media μ e con varianza σ^2 non nulla. Allora, indicando con Z_n la variabile aleatoria definita in (7.22) e con Z una variabile aleatoria $N(0, 1)$, si ha*

$$\lim_{n \rightarrow +\infty} P(Z_n \leq x) = P(Z \leq x), \quad \forall x \in \mathbb{R}. \quad (7.24)$$

Il fatto che la variabile aleatoria limite Z abbia sempre *la stessa distribuzione*, qualunque sia la distribuzione delle variabili X_i (con momento secondo finito), costituisce un fenomeno di *universalità* di eccezionale interesse: nel limite descritto dal Teorema 7.11, i dettagli fini della distribuzione delle X_i diventano irrilevanti e contano soltanto il valor medio μ e la varianza σ^2 (che compaiono nella definizione di Z_n , cf. (7.22)). Il fatto che $Z \sim N(0, 1)$ spiega l'importanza della distribuzione normale e il fatto che sia così diffusa in diverse aree della scienza.

Non è affatto evidente *a priori* per quale motivo la variabile aleatoria limite Z che appare in (7.24) debba avere distribuzione proprio *normale*. Per capire meglio, consideriamo il caso speciale in cui le variabili di partenza siano già esse normali, ossia $X_i \sim N(\mu, \sigma^2)$ per ogni $i \in \mathbb{N}$. Dato che la classe delle variabili normali è stabile per somma di variabili indipendenti e per combinazioni lineari affini, grazie al Corollario 6.35, segue facilmente che $Z_n \sim N(0, 1)$. Quindi la relazione (7.24) è banalmente verificata, perché Z_n ha esattamente la stessa distribuzione di Z .

In altri termini, la distribuzione $N(0, 1)$ è un “punto fisso” della trasformazione che alla distribuzione delle X_i associa la distribuzione di Z_n . Il teorema limite centrale mostra che, anche quando le X_i hanno distribuzione non normale, purché di varianza finita e non nulla, al crescere di n la distribuzione di Z_n tende ad avvicinarsi sempre più (nel senso (7.24) della convergenza delle funzioni di ripartizione) alla distribuzione $N(0, 1)$, che funge dunque da “centro di attrazione”.

Presentiamo la dimostrazione del Teorema 7.11 nel Paragrafo 7.2.3, facendo l'ipotesi aggiuntiva che le variabili X_i ammettano anche momento terzo finito. Questa condizione, tipicamente verificata negli esempi concreti, può essere rimossa con tecniche che vanno al di là degli scopi di questo libro. Prima di vedere la dimostrazione, nel prossimo paragrafo mostriamo in dettaglio, attraverso esempi concreti, come il teorema limite centrale possa essere usato per calcolare approssimativamente le probabilità di eventi legati alla media campionaria \bar{X}_n .

Osservazione 7.12. Invece di studiare la media campionaria \bar{X}_n , talvolta risulta più intuitivo considerare la *somma* S_n delle prime n variabili, definita da

$$S_n := \sum_{i=1}^n X_i. \quad (7.25)$$

I due approcci sono equivalenti, in quanto le proprietà di \bar{X}_n sono in corrispondenza con quelle di S_n , grazie alla semplice relazione $\bar{X}_n = \frac{1}{n}S_n$. In particolare, ricordando le relazioni (7.4) e (7.5),

$$E(S_n) = n\mu, \quad \text{Var}(S_n) = n\sigma^2.$$

Anche la variabile aleatoria Z_n definita in (7.22), che compare nel teorema limite centrale, può essere facilmente espressa in termini di S_n :

$$Z_n = \frac{S_n - \mu n}{\sigma\sqrt{n}}, \quad (7.26)$$

Si noti che le relazioni (7.22) e (7.26) hanno la *medesima struttura*:

$$Z_n = \frac{Y - E(Y)}{\sqrt{\text{Var}(Y)}}, \quad \text{con } Y = \bar{X}_n \text{ o } Y = S_n.$$

In generale, la procedura che a una variabile aleatoria $Y \in L^2$ associa la variabile $Z := (Y - E(Y))/\sqrt{\text{Var}(Y)}$ è detta *standardizzazione*, perché la variabile Z ottenuta ha sempre valor medio nullo e varianza unitaria, come si verifica facilmente. \square

Osservazione 7.13. Un modo euristico di descrivere la legge dei grandi numeri consiste nell'affermare che *i valori tipici della media campionaria \bar{X}_n sono vicini a μ* , o equivalentemente *i valori tipici della somma S_n sono vicini a $n\mu$* . Il teorema limite centrale permette di quantificare il senso dell'aggettivo "vicini":

$$\begin{aligned} & i \text{ valori tipici della media campionaria } \bar{X}_n \text{ sono dell'ordine di } \mu \pm \sigma \frac{1}{\sqrt{n}}, \\ & i \text{ valori tipici della somma } S_n \text{ sono dell'ordine di } \mu n \pm \sigma \sqrt{n}. \end{aligned} \quad (7.27)$$

Sottolineiamo che anche queste sono affermazioni euristiche, il cui significato preciso è dato dal Teorema 7.11, insieme alle relazioni (7.22) e (7.26). Tuttavia, *se ben comprese*, le affermazioni in (7.27) forniscono un'ottima intuizione sul comportamento asintotico di \bar{X}_n e S_n , come mostra l'esempio seguente. \square

Esempio 7.14. Sia S_n il numero di teste che si ottengono lanciando n volte una moneta equilibrata. Quali scostamenti di S_n dal valor medio $n/2$ è lecito aspettarsi, tipicamente? Il teorema limite centrale può aiutare a fornire una risposta.

Scriviamo $S_n = X_1 + \dots + X_n$, dove $(X_k)_{k \in \mathbb{N}}$ sono variabili aleatorie i.i.d. $\text{Be}(1/2)$. Si noti che $\sigma^2 = \text{Var}(X_1^2) = \frac{1}{4}$ e $\mu = E(X_1) = \frac{1}{2}$. Ricordando la formula (7.26), la relazione (7.24) mostra che per ogni $\alpha > 0$

$$P\left(S_n > \frac{n}{2} + \frac{\alpha}{2}\sqrt{n}\right) = P(Z_n > \alpha) \xrightarrow{n \rightarrow \infty} P(Z > \alpha).$$

Questo significa che scostamenti del numero di teste dal valore medio $\frac{n}{2}$ che crescono *proporzionalmente alla radice quadrata \sqrt{n} del numero di lanci* hanno all'incirca la stessa probabilità. Concretamente, per $\alpha = 1.5$, con l'ausilio di un computer si possono calcolare esattamente le probabilità seguenti:

$$\begin{aligned} P(S_{100} > 50 + \frac{1.5}{2}\sqrt{100}) &= P(S_{100} > 50 + 7) \simeq 6.66\%, \\ P(S_{1000} > 500 + \frac{1.5}{2}\sqrt{1000}) &= P(S_{1000} > 500 + 23) \simeq 6.86\%, \\ P(S_{10000} > 5000 + \frac{1.5}{2}\sqrt{10000}) &= P(S_{10000} > 5000 + 75) \simeq 6.55\%, \end{aligned}$$

in buon accordo col valore limite $P(Z > 1.5) = 6.68\%$.[†] In altri termini, uno scostamento dal valor medio di 7 teste in 100 lanci è all'incirca altrettanto probabile di uno

[†] L'approssimazione può essere sensibilmente migliorata usando un accorgimento detto correzione di continuità, descritto nel Paragrafo 7.2.2.

scostamento di 23 teste in 1 000 lanci, o di 75 teste in 10 000 lanci. Questi valori per lo scostamento crescono lentamente col numero dei lanci, perché $\sqrt{n} \ll n$.

D'altro canto, scostamenti dal valor medio *proporzionali al numero n di lanci*, e non alla sua radice quadrata, hanno probabilità $P(S_n > \frac{n}{2} + \theta n)$ che decadono molto rapidamente per $n \rightarrow \infty$: a titolo di esempio, per $\theta = 0.3$,

$$\begin{aligned} P(S_{100} > 50 + 3) &\simeq 24\%, \\ P(S_{1000} > 500 + 30) &\simeq 2.7\%, \\ P(S_{10000} > 5000 + 300) &\simeq 10^{-9}. \end{aligned}$$

Si noti quanto sia *estremamente raro* (meno di una probabilità su un miliardo) che una moneta equilibrata cada sulla testa più di 5 300 volte, in 10 000 lanci. \square

7.2.2 Il metodo dell'approssimazione normale

Tra le ragioni della rilevanza applicativa del teorema limite centrale vi è la possibilità di effettuare il calcolo approssimato di probabilità di interesse, il cui calcolo esatto risulterebbe arduo. Prima di descrivere questo *metodo dell'approssimazione normale*, presentiamo alcuni raffinamenti del teorema limite centrale.

Per comodità, d'ora in avanti indichiamo con $Z \sim N(0, 1)$ una variabile aleatoria normale standard e denotiamo con $\Phi: \mathbb{R} \rightarrow (0, 1)$ la sua funzione di ripartizione:

$$\Phi(x) := P(Z \leq x) = \int_{-\infty}^x \frac{e^{-\frac{1}{2}t^2}}{\sqrt{2\pi}} dt.$$

Si tratta di una funzione molto importante e, sebbene non sia esprimibile tramite funzioni “elementari”, molte sue proprietà sono note: ad esempio, è una funzione continua e derivabile su tutto \mathbb{R} , per il Teorema 6.9 (i), con derivata esplicita:

$$\Phi'(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2},$$

da cui segue che Φ è strettamente crescente. Inoltre, per le proprietà (5.4) delle funzioni di ripartizione, $\Phi(x)$ ha limite 0 per $x \rightarrow -\infty$ e 1 per $x \rightarrow +\infty$, dunque $\Phi: \mathbb{R} \rightarrow (0, 1)$ è una funzione biunivoca.

I valori della funzione Φ possono essere calcolati numericamente, usando un software con funzioni statistiche o una calcolatrice scientifica avanzata. In alternativa, la tabella a pagina 385 riporta i valori di $\Phi(x)$ per $0 \leq x \leq 3.5$, su una griglia di passo 0.01. Per $x \geq 3.5$ si ha $\Phi(x) \geq 0.9997 \simeq 1$, mentre i valori per $x < 0$ possono essere ricavati da quelli per $x > 0$ mediante la relazione

$$\Phi(x) = 1 - \Phi(-x), \quad \forall x \in \mathbb{R}, \quad (7.28)$$

che il lettore può dimostrare per esercizio.

Sia $(X_i)_{i \in \mathbb{N}}$ una successione di variabili aleatorie i.i.d. con media μ e varianza σ^2 finita e non nulla, come nelle ipotesi del Teorema 7.11, e poniamo $Z_n := \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}$, come in (7.22). Il teorema limite centrale afferma che, per ogni $x \in \mathbb{R}$,

$$\lim_{n \rightarrow +\infty} P(Z_n \leq x) = \Phi(x).$$

È facile vedere che anche

$$\lim_{n \rightarrow +\infty} P(Z_n < x) = \Phi(x). \quad (7.29)$$

Infatti, essendo ovviamente $P(Z_n < x) \leq P(Z_n \leq x)$, si ha che

$$\limsup_{n \rightarrow +\infty} P(Z_n < x) \leq \Phi(x). \quad (7.30)$$

Inoltre, per ogni $\varepsilon > 0$, si ha $P(Z_n < x) \geq P(Z_n \leq x - \varepsilon)$, e perciò

$$\liminf_{n \rightarrow +\infty} P(Z_n < x) \geq \Phi(x - \varepsilon). \quad (7.31)$$

Poiché (7.31) vale per ogni $\varepsilon > 0$ e Φ è continua, segue che

$$\liminf_{n \rightarrow +\infty} P(Z_n < x) \geq \Phi(x), \quad (7.32)$$

che, assieme a (7.30), dimostra (7.29).

Se ora I è un intervallo (limitato o illimitato) di \mathbb{R} , la probabilità $P(Z_n \in I)$ si può scrivere come differenza di termini della forma $P(Z_n \leq x)$ o $P(Z_n < x)$, a seconda che gli estremi siano inclusi o esclusi, con $x \in [-\infty, +\infty]$. Si ha pertanto che

$$\lim_{n \rightarrow +\infty} P(Z_n \in I) = P(Z \in I) \quad (7.33)$$

per ogni intervallo I di \mathbb{R} . È rilevante osservare che tale convergenza è *uniforme* rispetto all'intervallo I , cioè si ha

$$\lim_{n \rightarrow +\infty} \sup\{|P(Z_n \in I) - P(Z \in I)| : I \text{ intervallo di } \mathbb{R}\} = 0. \quad (7.34)$$

Il limite (7.34) è una conseguenza (Esercizio 7.2) del seguente raffinamento del teorema limite centrale.

Proposizione 7.15. *Nelle ipotesi del Teorema 7.11, si ha*

$$\lim_{n \rightarrow +\infty} \sup_{x \in \mathbb{R}} |P(Z_n \leq x) - \Phi(x)| = 0.$$

Dimostrazione. Abbiamo già notato che Φ è una funzione strettamente crescente, e dunque invertibile, fra \mathbb{R} e $(0, 1)$. Per $N \in \mathbb{N}$ e $k = 1, 2, \dots, 2N - 1$, possiamo dunque porre

$$x_k := \Phi^{-1} \left(\frac{k}{2N} \right).$$

Poiché $\{x_1, x_2, \dots, x_{2N-1}\}$ è un insieme finito di punti, per il Teorema 7.11 si ha che esiste \bar{n} , che dipende solo da N , tale che per ogni $n \geq \bar{n}$ e per ogni punto x_k

$$|P(Z_n \leq x_k) - \Phi(x_k)| \leq \frac{1}{2N}.$$

Sia ora $x \in \mathbb{R}$ fissato, ma arbitrario. Ponendo $x_0 := -\infty$ e $x_{2N} := +\infty$, sia $k \in \{0, 1, \dots, 2N - 1\}$ l'unico indice per cui $x_k < x \leq x_{k+1}$. Per $n \geq \bar{n}$, se $P(Z_n \leq x) \geq \Phi(x)$ si ha

$$\begin{aligned} |P(Z_n \leq x) - \Phi(x)| &= P(Z_n \leq x) - \Phi(x) \leq P(Z_n \leq x_{k+1}) - \Phi(x_k) \\ &\leq |P(Z_n \leq x_{k+1}) - \Phi(x_{k+1})| + \Phi(x_{k+1}) - \Phi(x_k) \\ &\leq \frac{1}{2N} + \left(\frac{k+1}{2N} - \frac{k}{2N} \right) = \frac{1}{N}. \end{aligned}$$

Il caso in cui $P(Z_n \leq x) < \Phi(x)$ si tratta in modo analogo:

$$\begin{aligned} |P(Z_n \leq x) - \Phi(x)| &= \Phi(x) - P(Z_n \leq x) \leq \Phi(x_{k+1}) - P(Z_n \leq x_k) \\ &\leq \Phi(x_{k+1}) - \Phi(x_k) + |\Phi(x_k) - P(Z_n \leq x_k)| \\ &\leq \left(\frac{k+1}{2N} - \frac{k}{2N} \right) + \frac{1}{2N} = \frac{1}{N}. \end{aligned}$$

In definitiva, abbiamo mostrato che per ogni N esiste \bar{n} tale che per ogni $x \in \mathbb{R}$ si ha

$$|P(Z_n \leq x) - \Phi(x)| \leq \frac{1}{N},$$

che è proprio la convergenza uniforme nell'enunciato □

Una delle conseguenze della Proposizione 7.15, è che consente di approssimare $P(Z_n \leq x_n)$ con $\Phi(x_n)$ anche nel caso in cui x_n dipenda da n : l'errore nell'approssimazione è comunque limitato da

$$\|F_{Z_n} - \Phi\|_\infty := \sup_{x \in \mathbb{R}} |F_{Z_n}(x) - \Phi(x)|. \quad (7.35)$$

Per un uso accorto di tale approssimazione, è utile avere delle stime il più possibile esplicite per $\|F_{Z_n} - \Phi\|_\infty$. Quello che segue è uno dei risultati più generali in questa direzione; ne omettiamo la dimostrazione.

Teorema 7.16 (Berry-Esseen). *In aggiunta alle ipotesi del Teorema 7.11, si assuma che $\rho := E(|X_1 - \mu|^3) < +\infty$. Allora*

$$\|F_{Z_n} - \Phi\|_\infty \leq \frac{\rho}{2\sigma^3\sqrt{n}}. \quad (7.36)$$

Si noti che, essendo $\rho \geq \sigma^3$ (Esercizio 7.3), la disuguaglianza (7.36) risulta poco utile per valori “piccoli” di n .

Osservazione 7.17. Per n fissato, la quantità $\|F_{Z_n} - \Phi\|_\infty$ dipende in modo sostanziale dalla distribuzione delle X_i . In effetti, come mostreremo nell'Esempio 7.20, per $\varepsilon > 0$ e $n \in \mathbb{N}$ si può scegliere la distribuzione per le variabili X_i in modo che $\|F_{Z_n} - \Phi\|_\infty \geq \frac{1}{2} - \varepsilon$. Pertanto, ogni stima per $\|F_{Z_n} - \Phi\|_\infty$ che tenda a zero per $n \rightarrow \infty$ deve necessariamente dipendere da qualche proprietà della distribuzione delle X_i , come la quantità ρ del Teorema 7.16. \square

Sulla base di questi risultati, illustriamo finalmente il *metodo dell'approssimazione normale*. Sia $(X_i)_{i \in \mathbb{N}}$ una successione di variabili aleatorie i.i.d. con media μ e varianza σ^2 finita e non nulla. Per un assegnato valore $s \in \mathbb{R}$ si voglia calcolare

$$P(S_n \leq s),$$

dove $S_n := X_1 + X_2 + \dots + X_n$. L'idea è di riscrivere l'evento $\{S_n \leq s\}$ in termini di Z_n , data dall'espressione (7.26): si ha infatti l'uguaglianza di eventi

$$\{S_n \leq s\} = \left\{ \frac{S_n - \mu n}{\sigma \sqrt{n}} \leq \frac{s - \mu n}{\sigma \sqrt{n}} \right\} = \left\{ Z_n \leq \frac{s - \mu n}{\sigma \sqrt{n}} \right\}.$$

Di conseguenza, dalla Proposizione 7.15 si ottiene l'approssimazione cercata:

$$P(S_n \leq s) = P\left(Z_n \leq \frac{s - \mu n}{\sigma \sqrt{n}}\right) \simeq \Phi\left(\frac{s - \mu n}{\sigma \sqrt{n}}\right), \quad (7.37)$$

ed esattamente la stessa approssimazione vale per $P(S_n < s)$. Per la relazione (7.35), l'errore nell'approssimazione (7.37) è al massimo $\|F_{Z_n} - \Phi\|_\infty$, una quantità che può essere stimata grazie al Teorema 7.16 di Berry-Esseen.

Esempio 7.18. Si lanci 1000 volte un dado equilibrato. Qual è la probabilità che il punteggio totale sia minore o uguale a 3450?

Sia $(X_i)_{i \in \mathbb{N}}$ una successione di variabili aleatorie i.i.d., ciascuna con distribuzione $\text{Unif}\{1, \dots, 6\}$. Con calcoli elementari, in parte già fatti nel Paragrafo 3.5.1, si trova

$$\mu = E(X_i) = \frac{7}{2}, \quad \sigma^2 = \text{Var}(X_i) = \frac{35}{12}, \quad \rho = E(|X - \mu|^3) = \frac{51}{8}.$$

Con le notazioni introdotte sopra, l'approssimazione normale (7.37) dà

$$P(S_{1000} \leq 3450) \simeq \Phi\left(\frac{3450 - \frac{7}{2} 1000}{\sqrt{\frac{35}{12}} \sqrt{1000}}\right) \simeq \Phi(-0.93) = 1 - \Phi(0.93) \simeq 0.1762,$$

avendo usato la relazione (7.28) e il valore $\Phi(0.93) \simeq 0.8238$, ricavato dalla tavola a pagina 385. Per il teorema di Berry-Esseen, l'errore di approssimazione è non maggiore di

$$\frac{\rho}{2\sigma^3 \sqrt{1000}} \simeq 0.0202.$$

Dunque, con gli strumenti a disposizione, possiamo concludere che

$$P(S_{1000} \leq 3450) \in [0.1762 - 0.0202, 0.1762 + 0.0202] = [0.1560, 0.1964].$$

Come vedremo più avanti, in molti casi il teorema di Berry-Esseen *sovrastima* significativamente il margine di errore. \square

Esempio 7.19. Quante volte è necessario lanciare un dado equilibrato affinché il punteggio totale sia maggiore o uguale a 3000 con probabilità di almeno 0.8?

Nelle notazioni dell'esempio precedente, dobbiamo determinare n affinché

$$P(S_n \geq 3000) \geq 0.8. \quad (7.38)$$

Usando l'approssimazione normale (7.37) (con la disuguaglianza stretta),

$$P(S_n \geq 3000) = 1 - P(S_n < 3000) \simeq 1 - \Phi\left(\frac{3000 - \frac{7}{2}n}{\sqrt{\frac{35}{12}}\sqrt{n}}\right) = \Phi\left(\frac{\frac{7}{2}n - 3000}{\sqrt{\frac{35}{12}}\sqrt{n}}\right).$$

Per la monotonia di Φ , la disequazione $\Phi(x) \geq 0.8$ equivale a $x \geq \Phi^{-1}(0.8)$, dunque da (7.38) ricaviamo approssimativamente

$$\frac{\frac{7}{2}n - 3000}{\sqrt{\frac{35}{12}}\sqrt{n}} \geq \Phi^{-1}(0.8) \quad \Longleftrightarrow \quad \frac{7}{2}n - \Phi^{-1}(0.8)\sqrt{\frac{35}{12}}\sqrt{n} - 3000 \geq 0.$$

Abbiamo dunque ottenuto una *disequazione di secondo grado in \sqrt{n}* con coefficienti numerici espliciti, ad esclusione di $\Phi^{-1}(0.8)$ che ora stimiamo. Dalla tavola pagina 385 si ricavano i valori di z per cui $\Phi(z)$ è vicino a 0.8:

$$\Phi(0.84) = 0.7995, \quad \Phi(0.85) = 0.8023,$$

da cui segue per monotonia che $0.84 \leq \Phi^{-1}(0.8) \leq 0.85$. Prendendo la media, approssimiamo $\Phi^{-1}(0.8) \simeq 0.845$ ed esplicitiamo la disequazione

$$\frac{7}{2}(\sqrt{n})^2 - 0.845\sqrt{\frac{35}{12}}\sqrt{n} - 3000 \geq 0,$$

le cui soluzioni positive (le uniche che ci interessano) sono date da

$$\sqrt{n} \geq \frac{0.8416\sqrt{\frac{35}{12}} + \sqrt{(0.8416)^2\frac{35}{12} + 4 \cdot \frac{7}{2} \cdot 3000}}{7} \simeq 29.48,$$

cioè $n \geq (29.48)^2 \simeq 869.07$. Considerando il fatto che n è intero, otteniamo infine $n \geq 870$. Anche in questo caso bisognerebbe considerare l'errore nell'approssimazione, ma ci accontenteremo questa volta del risultato approssimato. \square

Esempio 7.20. Siano $n \geq 1$ e $\varepsilon > 0$ fissati ma arbitrari, e $(X_i)_{i \in \mathbb{N}}$ una successione di variabili aleatorie i.i.d. con $X_i \sim \text{Be}(p)$, in cui p è sufficientemente piccolo da soddisfare $(1-p)^n \geq 1-\varepsilon$ e $np < 1$. Si ha che

$$P(Z_n \leq 0) = P(S_n \leq np) = P(S_n = 0) = (1-p)^n \geq 1-\varepsilon.$$

Essendo $\Phi(0) = \frac{1}{2}$, abbiamo

$$\|F_{Z_n} - \Phi\|_\infty \geq \frac{1}{2} - \varepsilon.$$

Questo mostra che, per n fissato, l'accuratezza dell'approssimazione normale non può essere uniforme rispetto alla distribuzione delle X_i . \square

Concludiamo descrivendo un semplice accorgimento, detto *correzione di continuità*, che permette spesso di migliorare l'accuratezza dell'approssimazione normale quando le variabili aleatorie X_i sono a valori interi (come nei precedenti esempi). In questo anche $S_n = X_1 + \dots + X_n$ è a valori interi, pertanto si è interessati a calcolare probabilità del tipo $P(S_n \leq m)$ con m intero. Una stima uniforme dell'errore che si commette applicando l'approssimazione normale (7.37) è pertanto

$$E_n := \sup_{m \in \mathbb{Z}} \left| P(S_n \leq m) - \Phi\left(\frac{m - \mu n}{\sigma \sqrt{n}}\right) \right|. \quad (7.39)$$

Osserviamo ora che, poiché S_n è a valori interi, si ha $P(S_n \leq m) = P(S_n < m+1)$. Applicando l'approssimazione normale (7.37) a quest'ultima probabilità, si ha

$$P(S_n \leq m) \simeq \Phi\left(\frac{m+1 - \mu n}{\sigma \sqrt{n}}\right) \quad \text{anziché} \quad P(S_n \leq m) \simeq \Phi\left(\frac{m - \mu n}{\sigma \sqrt{n}}\right).$$

In molti casi tipici, risulta che la prima approssimazione sovrastima, mentre la seconda sottostima, la vera probabilità $P(S_n \leq m)$. L'accorgimento detto *correzione di continuità* consiste nel prendere una soluzione intermedia: osservando che si ha anche $P(S_n \leq m) = P(S_n \leq m + \frac{1}{2})$, si approssima

$$P(S_n \leq m) \simeq \Phi\left(\frac{m + \frac{1}{2} - \mu n}{\sigma \sqrt{n}}\right). \quad (7.40)$$

In questo caso l'errore massimo che si commette è dato da

$$\hat{E}_n := \sup_{m \in \mathbb{Z}} \left| P(S_n \leq m) - \Phi\left(\frac{m + \frac{1}{2} - \mu n}{\sigma \sqrt{n}}\right) \right|. \quad (7.41)$$

Non è difficile mostrare (Esercizio 7.4) che sia E_n che \hat{E}_n sono minori di $\|F_{Z_n} - \Phi\|_\infty$, la quantità definita in (7.35) che compare nel teorema di Berry-Esseen.

Resta da capire se l'errore massimo \hat{E}_n con la correzione di continuità sia effettivamente migliore di quello E_n dato dall'approssimazione normale "standard". Ci li-

mitiamo al caso speciale, di grande interesse applicativo, in cui $X_i \sim \text{Be}(p)$ e dunque $S_n \sim \text{Bin}(n, p)$, fornendo alcune osservazioni numeriche.

- Se $p = \frac{1}{2}$, per $n = 20$ si calcola $\hat{E}_{20} \simeq 0.001$ mentre $E_{20} \simeq 0.088$, quindi la correzione di continuità migliora drasticamente la stima. (Il teorema di Berry-Esseen fornisce una stima peggiore, pari circa a 0.112.)
- Nei casi “asimmetrici” in cui $p \neq \frac{1}{2}$, il vantaggio della correzione di continuità è meno pronunciato, ma comunque rilevante: ad esempio, per $p = 0.1$ e $n = 50$, si ha $\hat{E}_{50} \simeq 0.024$ contro $E_{50} \simeq 0.093$ (mentre il limite superiore all’errore fornito dal teorema di Berry-Esseen vale circa 0.194).

Secondo un criterio empirico molto usato, l’approssimazione normale *con correzione di continuità* fornisce una “buona stima” quando $np \geq 5$ e $n(1-p) \geq 5$. Non siamo a conoscenza di risultati rigorosi che giustificino tale criterio, ma evidenze numeriche mostrano che in tal caso si ha $\hat{E}_n \leq 0.029$

7.2.3 Dimostrazione del teorema limite centrale

La dimostrazione del Teorema 7.11 verrà divisa in più passi. Inoltre, come si è già detto, dimostreremo il teorema sotto l’ipotesi aggiuntiva che le variabili aleatorie X_i ammettano momento terzo finito. Iniziamo con alcuni lemmi ausiliari.

Lemma 7.21. Denotiamo con \mathcal{C}_b^3 l’insieme delle funzioni da \mathbb{R} in \mathbb{R} le cui prime tre derivate esistono e sono funzioni continue e limitate su \mathbb{R} . Sia $(Z_n)_{n \geq 1}$ una successione di variabili aleatorie, e Z una variabile aleatoria la cui funzione di ripartizione F_Z è continua. Supponiamo che per ogni $g \in \mathcal{C}_b^3$ si abbia che

$$\lim_{n \rightarrow +\infty} E[g(Z_n)] = E[g(Z)]. \quad (7.42)$$

Allora

$$\lim_{n \rightarrow +\infty} F_{Z_n}(x) = F_Z(x), \quad \forall x \in \mathbb{R}.$$

Dimostrazione. Fissiamo arbitrariamente $x \in \mathbb{R}$. Per ogni $k \geq 1$, è possibile costruire due funzioni g_k e \tilde{g}_k in \mathcal{C}_b^3 tali che, come mostrato in Figura 7.1,

$$\mathbb{1}_{(-\infty, x - \frac{1}{k}]}(z) \leq \tilde{g}_k(z) \leq \mathbb{1}_{(-\infty, x]}(z) \leq g_k(z) \leq \mathbb{1}_{(-\infty, x + \frac{1}{k}]}(z), \quad \forall z \in \mathbb{R}.$$

Ad esempio, si può scegliere $g_k(z) = f_{[x, x + \frac{1}{k}]}(z)$ e $\tilde{g}_k(z) = f_{[x - \frac{1}{k}, x]}(z)$, dove la funzione $f_{[a, b]}$ è definita per $a < b$ nel modo seguente:

$$f_{[a, b]}(z) := f_{[0, 1]} \left(\frac{z - a}{b - a} \right), \quad f_{[0, 1]}(z) := \begin{cases} 1 & \text{se } z \leq 0 \\ 1 - 140 \int_0^z t^3 (1 - t)^3 dt & \text{se } 0 \leq z \leq 1 \\ 0 & \text{se } z \geq 1 \end{cases},$$

dove il fattore di normalizzazione 140 è dovuto al fatto che $\int_0^1 t^3 (1 - t)^3 dt = \frac{1}{140}$.

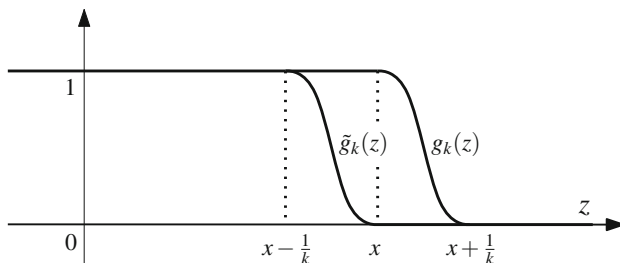


Figura 7.1 Una rappresentazione grafica delle funzioni g_k e \tilde{g}_k , che approssimano dall'alto e dal basso la funzione indicatrice $\mathbb{1}_{(-\infty, x]}(\cdot)$

Per definizione di funzione di ripartizione, per ogni variabile aleatoria W

$$F_W(x) = P(W \leq x) = E(\mathbb{1}_{\{W \leq x\}}) = E(\mathbb{1}_{(-\infty, x]}(W)).$$

Di conseguenza, per monotonia del valor medio,

$$\begin{aligned} F_W\left(x - \frac{1}{k}\right) &= E\left[\mathbb{1}_{(-\infty, x - \frac{1}{k}]}(W)\right] \leq E[\tilde{g}_k(W)] \leq E[\mathbb{1}_{(-\infty, x]}(W)] \\ &= F_W(x) \leq E[g_k(W)] \leq E\left[\mathbb{1}_{(-\infty, x + \frac{1}{k}]}(W)\right] = F_W\left(x + \frac{1}{k}\right), \end{aligned}$$

in particolare

$$F_W\left(x - \frac{1}{k}\right) \leq E[\tilde{g}_k(W)] \leq F_W(x) \leq E[g_k(W)] \leq F_W\left(x + \frac{1}{k}\right).$$

Usiamo ora queste disuguaglianze, per $W = Z_n$ e $W = Z$, insieme con l'ipotesi (7.42) per $g = g_k$, ottenendo che per ogni $k \in \mathbb{N}$ fissato

$$\begin{aligned} \limsup_{n \rightarrow +\infty} F_{Z_n}(x) &\leq \lim_{n \rightarrow +\infty} E[g_k(Z_n)] = E[g_k(Z)] \leq F_Z\left(x + \frac{1}{k}\right), \\ \liminf_{n \rightarrow +\infty} F_{Z_n}(x) &\geq \lim_{n \rightarrow +\infty} E[\tilde{g}_k(Z_n)] = E[\tilde{g}_k(Z)] \geq F_Z\left(x - \frac{1}{k}\right), \end{aligned}$$

quindi

$$F_Z\left(x - \frac{1}{k}\right) \leq \liminf_{n \rightarrow +\infty} F_{Z_n}(x) \leq \limsup_{n \rightarrow +\infty} F_{Z_n}(x) \leq F_Z\left(x + \frac{1}{k}\right). \quad (7.43)$$

Essendo F_Z continua,

$$\lim_{k \rightarrow +\infty} F_Z\left(x - \frac{1}{k}\right) = \lim_{k \rightarrow +\infty} F_Z\left(x + \frac{1}{k}\right) = F_Z(x),$$

pertanto prendendo il limite $k \rightarrow +\infty$ in (7.43) troviamo

$$\liminf_{n \rightarrow +\infty} F_{Z_n}(x) = \limsup_{n \rightarrow +\infty} F_{Z_n}(x) = F_Z(x), \quad \forall x \in \mathbb{R},$$

che conclude la dimostrazione. \square

Veniamo dunque a un lemma cruciale.

Lemma 7.22. *Siano V, Y, Z tre variabili aleatorie indipendenti, tali che Y, Z ammettono momento terzo finito, e inoltre $E(Y) = E(Z)$, $E(Y^2) = E(Z^2)$. Allora per ogni $g \in \mathcal{C}_b^3$, ponendo $C := \sup_{x \in \mathbb{R}} |g'''(x)|$, vale la disuguaglianza*

$$|E[g(V+Y)] - E[g(V+Z)]| \leq \frac{C}{6} [E(|Y|^3) + E(|Z|^3)].$$

Dimostrazione. La formula di Taylor per funzioni di classe \mathcal{C}^3 con resto integrale ci dà, per ogni $x, h \in \mathbb{R}$

$$g(x+h) = g(x) + g'(x)h + \frac{1}{2}g''(x)h^2 + R_2(x, h),$$

dove

$$R_2(x, h) = \frac{1}{2} \int_x^{x+h} (x+h-t)^2 g'''(t) dt.$$

In particolare

$$|R_2(x, h)| \leq \frac{C}{6} |h|^3. \quad (7.44)$$

Si ricava facilmente che

$$g(x+h) - g(x+k) = g'(x)[h-k] + \frac{1}{2}g''(x)[h^2-k^2] + R_2(x, h) - R_2(x, k). \quad (7.45)$$

Ponendo $x = V, h = Y, k = Z$ e prendendo il valor medio, otteniamo

$$\begin{aligned} E[g(V+Y)] - E[g(V+Z)] &= E[g'(V)(Y-Z)] + \frac{1}{2}E[g''(V)(Y^2-Z^2)] \\ &\quad + E[R_2(V, Y) - R_2(V, Z)]. \end{aligned}$$

Essendo V, Y, Z indipendenti e $E(Y) = E(Z)$, $E(Y^2) = E(Z^2)$, si ha che

$$\begin{aligned} E[g'(V)(Y-Z)] &= E[g'(V)]E[(Y-Z)] = 0, \\ E[g''(V)(Y^2-Z^2)] &= E[g''(V)]E[(Y^2-Z^2)] = 0, \end{aligned}$$

avendo usato il fatto che g' e g'' sono funzioni limitate, dunque $g'(V)$ e $g''(V)$ ammettono valor medio finito. Ricordando (7.44), otteniamo

$$\begin{aligned} |E[g(V+Y)] - E[g(V+Z)]| &= |E[R_2(V, Y) - R_2(V, Z)]| \\ &\leq E[|R_2(V, Y)|] + E[|R_2(V, Z)|] \leq \frac{C}{6} [E(|Y|^3) + E(|Z|^3)], \end{aligned}$$

ossia la tesi. \square

Il seguente risultato rappresenta il “cuore” della dimostrazione del Teorema 7.11.

Proposizione 7.23. *Siano Y_1, Y_2, \dots, Y_n variabili aleatorie i.i.d. con momento terzo finito e con $E(Y_1) = 0$, $E(Y_1^2) = 1$. Analogamente, siano W_1, W_2, \dots, W_n variabili aleatorie i.i.d. con momento terzo finito e con $E(W_1) = 0$, $E(W_1^2) = 1$. Ponendo $C := \sup_{x \in \mathbb{R}} |g'''(x)|$, per ogni $g \in \mathcal{C}_b^3$ si ha*

$$\left| E \left[g \left(\frac{Y_1 + \dots + Y_n}{\sqrt{n}} \right) \right] - E \left[g \left(\frac{W_1 + \dots + W_n}{\sqrt{n}} \right) \right] \right| \leq \frac{C}{6} \frac{E(|Y_1|^3) + E(|W_1|^3)}{\sqrt{n}}.$$

Dimostrazione. Sia $Y := (Y_1, Y_2, \dots, Y_n)$ e $W := (W_1, W_2, \dots, W_n)$. Il risultato da dimostrare dipende solo dalle distribuzioni individuali di Y e W , ma non dalla distribuzione congiunta di (Y, W) . Non è perciò restrittivo assumere che Y e W siano indipendenti, cioè che tutte le variabili aleatorie $Y_1, Y_2, \dots, Y_n, W_1, W_2, \dots, W_n$ siano indipendenti. L’idea chiave consiste nello scrivere la seguente somma telescopica:

$$\begin{aligned} & E \left[g \left(\frac{Y_1 + \dots + Y_n}{\sqrt{n}} \right) \right] - E \left[g \left(\frac{W_1 + \dots + W_n}{\sqrt{n}} \right) \right] \\ &= \sum_{k=0}^{n-1} \left\{ E \left[g \left(\frac{Y_1 + \dots + Y_k + Y_{k+1} + W_{k+2} + \dots + W_n}{\sqrt{n}} \right) \right] \right. \\ &\quad \left. - E \left[g \left(\frac{Y_1 + \dots + Y_k + W_{k+1} + W_{k+2} + \dots + W_n}{\sqrt{n}} \right) \right] \right\} \\ &= \sum_{k=0}^{n-1} \left\{ E \left[g \left(V_k + \frac{Y_{k+1}}{\sqrt{n}} \right) \right] - E \left[g \left(V_k + \frac{W_{k+1}}{\sqrt{n}} \right) \right] \right\}, \end{aligned}$$

dove abbiamo posto

$$V_k := \frac{Y_1 + \dots + Y_k + W_{k+2} + \dots + W_n}{\sqrt{n}}.$$

Per il Lemma 7.22

$$\left| E \left[g \left(V_k + \frac{Y_{k+1}}{\sqrt{n}} \right) \right] - E \left[g \left(V_k + \frac{W_{k+1}}{\sqrt{n}} \right) \right] \right| \leq \frac{C}{6} \frac{E(|Y_1|^3) + E(|W_1|^3)}{n\sqrt{n}},$$

pertanto

$$\begin{aligned} & \left| E \left[g \left(\frac{Y_1 + \dots + Y_n}{\sqrt{n}} \right) \right] - E \left[g \left(\frac{W_1 + \dots + W_n}{\sqrt{n}} \right) \right] \right| \\ & \leq \sum_{k=0}^{n-1} \left| E \left[g \left(V_k + \frac{Y_{k+1}}{\sqrt{n}} \right) \right] - E \left[g \left(V_k + \frac{W_{k+1}}{\sqrt{n}} \right) \right] \right| \\ & \leq n \frac{C}{6} \frac{E(|Y_1|^3) + E(|W_1|^3)}{n\sqrt{n}} = \frac{C}{6} \frac{E(|Y_1|^3) + E(|W_1|^3)}{\sqrt{n}}, \end{aligned}$$

che è quanto volevamo dimostrare. \square

Osservazione 7.24. Dalla proposizione precedente segue il fatto, assolutamente non banale e non intuitivo, che

$$\lim_{n \rightarrow +\infty} \left\{ E \left[g \left(\frac{Y_1 + \dots + Y_n}{\sqrt{n}} \right) \right] - E \left[g \left(\frac{W_1 + \dots + W_n}{\sqrt{n}} \right) \right] \right\} = 0 \quad (7.46)$$

indipendentemente dalle distribuzioni delle Y_i e delle W_i (purché soddisfino le ipotesi della Proposizione 7.23). \square

Possiamo finalmente completare la dimostrazione del Teorema 7.11 (con l'ipotesi aggiuntiva che le variabili X_i ammettano momento terzo finito). Si noti che le variabili normali non sono ancora apparse nella dimostrazione. Le introduciamo ora: sia $(W_i)_{i \in \mathbb{N}}$ una successione di variabili aleatorie i.i.d. con $W_i \sim N(0, 1)$. Come abbiamo già accennato in precedenza, segue dalla Corollario 6.35 che

$$\frac{W_1 + \dots + W_n}{\sqrt{n}} \sim N(0, 1), \quad \forall n \in \mathbb{N},$$

quindi, se $Z \sim N(0, 1)$ e $g \in \mathcal{C}_b^3$,

$$E \left[g \left(\frac{W_1 + \dots + W_n}{\sqrt{n}} \right) \right] = E[g(Z)]. \quad (7.47)$$

Usando le notazioni nell'enunciato del Teorema 7.11, poniamo ora

$$Y_i := \frac{X_i - \mu}{\sigma}, \quad \forall i \in \mathbb{N}.$$

Osserviamo che

$$Z_n = \frac{Y_1 + \dots + Y_n}{\sqrt{n}}, \quad \forall n \in \mathbb{N}, \quad (7.48)$$

come segue facilmente da (7.25) e (7.26). È immediato verificare che $E(Y_i) = 0$, $E(Y_i^2) = 1$, pertanto sono soddisfatte le ipotesi della Proposizione 7.23. Possiamo dunque applicare la relazione (7.46), che grazie a (7.47) e (7.48) si scrive come

$$\lim_{n \rightarrow +\infty} E[g(Z_n)] = E[g(Z)], \quad \forall g \in \mathcal{C}_b^3.$$

Non resta che applicare il Lemma 7.21, e il Teorema 7.11 è dimostrato.

Osservazione 7.25. Una lettura attenta della dimostrazione appena completata rivela che l'ipotesi che le variabili aleatorie X_i siano indipendenti è stata usata più volte, mentre quella che siano identicamente distribuite non è mai stata usata pienamente e può essere notevolmente indebolita: ad esempio, tutto funziona se

$$E(X_i) \text{ e } E(X_i^2) \text{ non dipendono da } i \in \mathbb{N}, \quad \sup_{i \in \mathbb{N}} E(|X_i|^3) < +\infty.$$

Questa osservazione amplia il raggio di validità del teorema limite centrale e rafforza dunque il valore di universalità della distribuzione normale, come distribuzione ap-

prossimata della somma di variabili aleatorie indipendenti, non necessariamente con la stessa distribuzione. \square

7.2.4 Un teorema limite locale per variabili esponenziali

Concludiamo questo capitolo dando un rafforzamento “locale” del teorema limite centrale, per una successione $(X_i)_{i \in \mathbb{N}}$ variabili aleatorie reali i.i.d. con distribuzione esponenziale. Assumiamo per semplicità che $X_i \sim \text{Exp}(1) = \text{Gamma}(1, 1)$, ma tutto si estende immediatamente al caso di $X_i \sim \text{Exp}(\lambda)$.

Ricordiamo che $\mu = E(X_i) = 1$, $\sigma^2 = \text{Var}(X_i) = 1$, cf. (6.36), e poniamo

$$Z_n := \frac{S_n - n}{\sqrt{n}}, \quad (7.49)$$

dove $S_n := X_1 + \dots + X_n$, cf. (7.26), (7.25). Segue dalla Proposizione 6.29 che $S_n \sim \text{Gamma}(n, 1)$, in particolare S_n è una variabile aleatoria assolutamente continua e di conseguenza anche Z_n lo è, per la Proposizione 6.18. Indicando con f_{Z_n} la sua densità, possiamo riformulare il teorema limite centrale (7.24) come segue:

$$\lim_{n \rightarrow \infty} \int_{-\infty}^x f_{Z_n}(y) dy = \int_{-\infty}^x \frac{e^{-y^2/2}}{\sqrt{2\pi}} dy, \quad \forall x \in \mathbb{R}. \quad (7.50)$$

A questo punto, è naturale chiedersi se ci sia la convergenza delle densità:

$$\lim_{n \rightarrow \infty} f_{Z_n}(x) = \frac{e^{-x^2/2}}{\sqrt{2\pi}}, \quad \forall x \in \mathbb{R}. \quad (7.51)$$

Mostriamo in questo paragrafo che la risposta è affermativa, ma sottolineiamo che non si tratta di una conseguenza “automatica” della relazione (7.50). In effetti, la dimostrazione che presentiamo consiste in un calcolo diretto, basato sulla conoscenza esplicita della densità $f_{Z_n}(x)$. Con tecniche più avanzate, al di là degli scopi di questo libro, è possibile estendere il risultato (7.51) a un’ampia classe di distribuzioni assolutamente continue. Come sottoprodotto della dimostrazione, otterremo una derivazione indipendente della formula di Stirling (1.20) (incluso il prefattore $C = \sqrt{2\pi}$, che era rimasto indeterminato nella dimostrazione della Proposizione 1.34).

Dimostrazione della relazione (7.51). Dato che $S_n \sim \text{Gamma}(n, 1)$, si ha

$$f_{S_n}(x) = \frac{x^{n-1}}{\Gamma(n)} e^{-x} \mathbb{1}_{[0, \infty)}(x) = \frac{x^{n-1}}{(n-1)!} e^{-x} \mathbb{1}_{(0, \infty)}(x).$$

Di conseguenza, per la Proposizione 6.18, segue da (7.49) che

$$f_{Z_n}(x) = \sqrt{n} f_{S_n}(\sqrt{n}x + n) = \sqrt{n} \frac{(\sqrt{n}x + n)^{n-1}}{(n-1)!} e^{-(\sqrt{n}x + n)} \mathbb{1}_{(-\sqrt{n}, \infty)}(x).$$

Nonostante l’apparenza ostica, questa relazione può essere efficacemente studiata.

Definiamo la successione reale

$$D_n := \frac{n^n e^{-n} \sqrt{2\pi n}}{n!}.$$

Notiamo che la formula di Stirling (1.20) equivale a $\lim_{n \rightarrow \infty} D_n = 1$, un fatto che dimostreremo tra poco. Possiamo riscrivere $f_{Z_n}(x)$ nel modo seguente:

$$f_{Z_n}(x) = D_n \frac{1}{\sqrt{2\pi}} \left(1 + \frac{x}{\sqrt{n}}\right)^{n-1} e^{-\sqrt{n}x} \mathbb{1}_{(-\sqrt{n}, \infty)}(x). \quad (7.52)$$

Definiamo ora per $t > -1$

$$R(t) := \log(1+t) - t + \frac{1}{2}t^2,$$

e si noti che $R(t) = O(t^3)$ per $t \rightarrow 0$, come segue dalla formula di Taylor per il logaritmo. Con qualche manipolazione algebrica si ha

$$\left(1 + \frac{x}{\sqrt{n}}\right)^n e^{-\sqrt{n}x} = e^{n \log(1 + \frac{x}{\sqrt{n}})} e^{-\sqrt{n}x} = e^{n[\frac{x}{\sqrt{n}} - \frac{x^2}{2n} + R(\frac{x}{\sqrt{n}})]} e^{-\sqrt{n}x} = e^{-\frac{x^2}{2}} e^{nR(\frac{x}{\sqrt{n}})}.$$

Definendo la funzione

$$g_n(x) := \left(1 + \frac{x}{\sqrt{n}}\right)^{-1} e^{nR(\frac{x}{\sqrt{n}})},$$

possiamo allora riscrivere la formula (7.52) nel modo seguente:

$$f_{Z_n}(x) = D_n g_n(x) \frac{e^{-x^2/2}}{\sqrt{2\pi}} \mathbb{1}_{(-\sqrt{n}, \infty)}(x). \quad (7.53)$$

Si noti che è “apparsa” la densità della normale standard. Dato che $R(t) = O(t^3)$, per ogni $x \in \mathbb{R}$ fissato si ha $nR(\frac{x}{\sqrt{n}}) = O(\frac{x^3}{\sqrt{n}})$ e dunque $\lim_{n \rightarrow \infty} g_n(x) = 1$. Di conseguenza, per completare la dimostrazione della relazione (7.51) resta solo da mostrare che $\lim_{n \rightarrow \infty} D_n = 1$ (ossia la formula di Stirling).

Fissiamo $\varepsilon > 0$ e sia $L := 1/\sqrt{\varepsilon}$. Applicando la disuguaglianza di Chebyshev (Teorema 3.77) possiamo scrivere

$$P(|Z_n| > L) = P(|S_n - E(S_n)| > L\sqrt{n}) \leq \frac{\text{Var}(S_n)}{L^2 n} = \frac{1}{L^2} = \varepsilon, \quad (7.54)$$

e analogamente, se $Z \sim N(0, 1)$,

$$P(|Z| > L) = P(|Z - E(Z)| > L) \leq \frac{\text{Var}(Z)}{L^2} = \frac{1}{L^2} = \varepsilon. \quad (7.55)$$

Dal fatto che $R(t) = O(t^3)$, e dunque $nR(\frac{x}{\sqrt{n}}) = O(\frac{x^3}{\sqrt{n}})$, segue che $\lim_{n \rightarrow \infty} g_n(x) = 1$

uniformemente per $x \in [-L, L]$. Quindi esiste $\bar{n} = \bar{n}(\varepsilon) < \infty$ tale che

$$1 - \varepsilon \leq g_n(x) \leq 1 + \varepsilon, \quad \forall n \geq \bar{n}, \forall x \in [-L, L].$$

Possiamo assumere che $\sqrt{\bar{n}} > L$. Di conseguenza, per $n \geq \bar{n}$, segue da (7.53) che

$$D_n(1 - \varepsilon) \int_{-L}^L \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx \leq \int_{-L}^L f_{Z_n}(x) dx \leq D_n(1 + \varepsilon) \int_{-L}^L \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx.$$

L'integrale nel primo e terzo membro non è altro che $P(|Z| \leq L)$, mentre quello nel secondo membro è $P(|Z_n| \leq L)$. Dalla relazione precedente otteniamo quindi

$$\frac{P(|Z_n| \leq L)}{(1 + \varepsilon)P(|Z| \leq L)} \leq D_n \leq \frac{P(|Z_n| \leq L)}{(1 - \varepsilon)P(|Z| \leq L)}, \quad \forall n \geq \bar{n}.$$

Ricordando le relazioni (7.54) e (7.55), si ha $1 - \varepsilon \leq P(|Z_n| \leq L) \leq 1$ e analogamente per $P(|Z| \leq L)$, pertanto abbiamo mostrato che

$$\frac{1 - \varepsilon}{1 + \varepsilon} \leq D_n \leq \frac{1}{(1 - \varepsilon)^2}, \quad \forall n \geq \bar{n},$$

quindi

$$\frac{1 - \varepsilon}{1 + \varepsilon} \leq \liminf_{n \rightarrow \infty} D_n \leq \limsup_{n \rightarrow \infty} D_n \leq \frac{1}{(1 - \varepsilon)^2}.$$

Prendendo il limite $\varepsilon \downarrow 0$ in questa relazione, segue che $\lim_{n \rightarrow \infty} D_n = 1$. \square

Esercizi

Esercizio 7.2. Si mostri che il limite (7.34) segue dalla Proposizione 7.15.

Esercizio 7.3. Con riferimento al Teorema 7.16, si dimostri che $\rho \geq \sigma^3$.

Esercizio 7.4. Si mostri che le quantità E_n ed \hat{E}_n , definite in (7.39) e (7.41), sono entrambe minori della quantità $\|F_{Z_n} - \Phi\|_\infty$ definita in (7.35).

7.3 Esercizi di riepilogo

Esercizio 7.5. Il gruppo promotore di un referendum ritiene che il 60% della popolazione sia disposta a firmare per la relativa raccolta di firme. Si assuma che le persone a cui viene richiesto di firmare siano scelte a caso. Dovendo raccogliere 30000 firme, quante persone è necessario interpellare affinché la soglia delle 30000 firme sia raggiunta con probabilità di almeno 0.95?

Esercizio 7.6. Un grande studio fotografico riceve l'incarico di eseguire un servizio che prevede l'uso di speciali lampade ad alta luminosità. La durata di tali lampade

ha distribuzione esponenziale di media uguale a 100 ore, e costano 100 Euro l'una. Le durate di lampade distinte si possono considerare indipendenti. Per il servizio si prevede siano necessarie 10000 ore di luce prodotta da tali lampade. Inoltre, a causa degli elevati costi di trasporto, è conveniente acquistare le lampade necessarie in un unico ordine.

- (i) Usando l'approssimazione normale, si determini il minimo numero di lampade che è necessario acquistare affinché le 10000 ore di luce siano garantite con probabilità 0.95.
- (ii) Un'altra ditta di lampade propone un prodotto la cui durata ha distribuzione esponenziale di media 200 ore, al costo di 190 Euro per lampada. Ritenete sia conveniente acquistare da questa ditta anziché da quella del punto precedente? (Anche in questo caso le lampade vengono acquistate nel numero minimo necessario a garantire 10000 ore di luce con probabilità 0.95).

Esercizio 7.7. Un gioco consiste nell'estrarre a caso due carte da un mazzo di carte da Poker (52 carte, 4 semi); si vince se nessuna delle carte estratte è di quadri.

- (i) Si determini la probabilità di successo in questo gioco.
- (ii) Per $n \geq 1$, sia p_n la probabilità che in $2n$ ripetizioni del gioco il numero di successi sia almeno n . Si determini, approssimativamente, il valore p_{50} .
- (iii) Si determini il minimo valore di n per cui $p_n \geq 1 - 10^{-4}$.

Esercizio 7.8. Un congegno è costituito da una componente elettrica che viene rimpiazzata non appena smette di funzionare. Dunque, se T_1, T_2, \dots, T_n sono i tempi di vita di n componenti che si hanno a disposizione, il tempo di vita totale del congegno è $T = T_1 + T_2 + \dots + T_n$. Si supponga che $T_i \sim \text{Exp}(1)$, e che le T_i siano indipendenti. Utilizzando l'approssimazione normale, si calcolino:

- (i) se $n = 100$ la probabilità $P(T < 90)$;
- (ii) il valore minimo di n per cui $P(T < 90) \leq 0.05$.

Esercizio 7.9. La lunghezza dei chiodini prodotti da una certa ditta ha una distribuzione incognita, la cui media e varianza indichiamo con μ e σ^2 . Il valore di σ^2 è noto e pari a 0.25 mm^2 , mentre il valore di μ (espresso in mm) è incognito e vogliamo stimarlo empiricamente.

A tal fine, misuriamo le lunghezze X_1, \dots, X_n di n chiodini scelti a caso e ne indichiamo la media aritmetica con $\bar{X}_n := (X_1 + \dots + X_n)/n$. Se n è grande, per la legge dei grandi numeri sappiamo che \bar{X}_n sarà vicino a μ . Per rendere più quantitativa questa affermazione, scegliamo un numero reale $\delta > 0$ e consideriamo l'intervallo I_δ di ampiezza δ centrato in \bar{X}_n , vale a dire

$$I_\delta := (\bar{X}_n - \delta, \bar{X}_n + \delta).$$

Si determini δ_n in modo che la probabilità che l'intervallo I_{δ_n} contenga μ valga approssimativamente 0.95 per n grande.

[Sugg. Si esprima l'evento $\{\mu \in I_{\delta_n}\}$ nella forma $\{a < \bar{X}_n < b\}$ per opportuni a, b .]

Esercizio 7.10. In una elezione votano un milione di persone, che devono scegliere tra i due candidati A e B . Il voto di un certo numero n di elettori è sotto il controllo di una organizzazione malavitosa, che garantisce che essi votino per il candidato A . Tutti gli altri elettori votano “a caso”, scegliendo con ugual probabilità i due candidati, ognuno indipendentemente dagli altri.

- (i) Supponiamo che l'organizzazione malavitosa controlli $n = 2000$ voti. Qual è la probabilità (approssimata) che il candidato A vinca le elezioni?
- (ii) Qual è il numero minimo n di individui che l'organizzazione malavitosa deve controllare, per garantire che la probabilità di vittoria di A sia almeno del 99%?

Esercizio 7.11. Indicando con $(X_n)_{n \in \mathbb{N}}$ una successione i.i.d. con distribuzioni marginali $X_n \sim U[-1, 1]$, si determini per ogni $t \in \mathbb{R}$ il limite

$$\lim_{n \rightarrow \infty} P \left(\frac{(X_1 + \dots + X_n)^2}{n} > t \right).$$

Esercizio 7.12. Usando opportunamente il teorema limite centrale, si calcoli

$$\lim_{n \rightarrow +\infty} e^{-n} \sum_{k=0}^{n+\sqrt{n}} \frac{n^k}{k!}. \quad (7.56)$$

[Sugg. Date $(X_n)_{n \in \mathbb{N}}$ variabili aleatorie i.i.d. con distribuzione $\text{Pois}(1)$, si esprima la somma in (7.56) in funzione di tali variabili aleatorie.]

Esercizio 7.13. Siano $(X_n)_{n \in \mathbb{N}}$ variabili aleatorie i.i.d. $U(-\frac{1}{2}, +\frac{1}{2})$. Indichiamo con $B_n(r, \delta)$ la “buccia” di spessore 2δ della palla di raggio r centrata nell'origine di \mathbb{R}^n :

$$B_n(r, \delta) = \left\{ (x_1, \dots, x_n) \in \mathbb{R}^n : r - \delta \leq \sqrt{x_1^2 + \dots + x_n^2} \leq r + \delta \right\}.$$

- (i) Si spieghi perché le variabili aleatorie $(Y_n)_{n \in \mathbb{N}}$, definite da $Y_n := X_n^2$, sono i.i.d., con valor medio $\mu = \frac{1}{12}$ e deviazione standard $\sigma = \frac{1}{6\sqrt{5}}$.
- (ii) Applicando il teorema limite centrale alla successione $(Y_n)_{n \in \mathbb{N}}$, si deduca che

$$\lim_{n \rightarrow \infty} P \left((X_1, \dots, X_n) \in B_n \left(\sqrt{\frac{n}{12}}, \delta \right) \right) = P \left(-\delta\sqrt{60} \leq Z \leq \delta\sqrt{60} \right),$$

per ogni $\delta > 0$, dove $Z \sim N(0, 1)$. In particolare, usando la tavola a pagina 385, si mostri che per $\delta = \frac{1}{3}$ il valore del limite è ≥ 0.99 .

- (iii) (*) Si mostri che, per ogni $n \in \mathbb{N}$, il vettore aleatorio (X_1, \dots, X_n) ha distribuzione uniforme continua nell'insieme $(-\frac{1}{2}, +\frac{1}{2})^n \subseteq \mathbb{R}^n$ (si ricordi l'Esempio 6.39). Ricordando la definizione (6.46) di misura n -dimensionale, si osservi che $\text{mis}((-\frac{1}{2}, +\frac{1}{2})^n) = 1$ e si deduca che per n sufficientemente grande

$$\text{mis} \left(B_n \left(\sqrt{\frac{n}{12}}, \frac{1}{3} \right) \cap \left(-\frac{1}{2}, +\frac{1}{2} \right)^n \right) \geq 99\%.$$

Possiamo dunque affermare che, per n grande, “quasi tutto il volume di un cubo n -dimensionale è contenuto nella buccia sottile di una palla”, visto che lo spessore $2\delta = \frac{2}{3}$ è costante, mentre il raggio $r = \sqrt{\frac{n}{12}}$ diverge per $n \rightarrow +\infty$!

7.4 Note bibliografiche

La prima dimostrazione, per variabili aleatorie di Bernoulli, della legge dei grandi numeri è dovuta a J. Bernoulli (1713), in [4], benché il nome “legge dei grandi numeri” venne proposto da S.D. Poisson in [46].

Le origini del metodo di Monte Carlo sono raccontate da N. Metropolis in [41]. Una parte rilevante delle moderne applicazioni del metodo di Monte Carlo sono basate su un'estensione della legge dei grandi numeri per particolari successioni di variabili aleatorie *dipendenti*, dette *catene di Markov*. Per una monografia sull'argomento si veda il libro di P. Bremaud [12].

La prima versione del teorema limite centrale è dovuta a A. De Moivre (1738), in [18]. Egli dimostrò un risultato di approssimazione normale per variabili aleatorie binomiali con $p = \frac{1}{2}$, indicando, benché in modo non completo, le modifiche necessarie per $p \neq \frac{1}{2}$. Un passo decisivo verso la formulazione moderna venne compiuto da P.-S. de Laplace (1812) in [34], una monografia che avrebbe fortemente influenzato gli sviluppi successivi del calcolo delle probabilità.

La dimostrazione del teorema limite centrale che abbiamo scelto di includere in questo testo è dovuta a J.W. Lindeberg [40]. Successivamente, P. Lévy [37, 38] fornì una dimostrazione basata sulle *funzioni caratteristiche* (dette anche *trasformate di Fourier*), che viene presentata nella quasi totalità dei testi avanzati di calcolo delle probabilità. Benché la dimostrazione di Lévy sia per certi versi più elegante e concisa di quella di Lindeberg, essa richiede strumenti analitici considerevolmente più avanzati. L'accessibilità della dimostrazione di Lindeberg è stata di recente sottolineata nell'articolo [15] di R.C. Dalang, a cui la nostra esposizione è ispirata.

Una storia dettagliata del teorema limite centrale è contenuta nella monografia di H. Fischer [24], che è tuttavia accessibile solo a lettori esperti.

Qualche commento merita infine il Teorema di Berry-Esseen, dimostrato indipendentemente da A.C. Berry [5] e da C.-G. Esseen [20], nella forma

$$\|F_{Z_n} - \Phi\|_\infty \leq \frac{C_0 \rho}{\sigma^3 \sqrt{n}},$$

per un'opportuna costante C_0 , indipendente dalla distribuzione di X_1 . La determinazione della costante ottima C_0 è, al momento, un problema aperto. C.-G. Esseen, in [21], dimostrò che $C_0 \geq \frac{\sqrt{10+3}}{6\sqrt{2\pi}} \simeq 0.40973$. L'enunciato del Teorema di Berry-Esseen che qui abbiamo presentato utilizza la disuguaglianza $C_0 \leq 0.5$, dimostrata in [33] (dove, per la precisione, si dimostra $C_0 \leq 0.4784$).

Applicazioni alla statistica matematica

Sommario. In questo capitolo applicheremo alcune delle nozioni fin qui introdotte alla *statistica matematica*. Nel primo paragrafo presentiamo diverse nozioni di base, senza alcuna pretesa di sistematicità o completezza. Successivamente, studiamo alcuni rilevanti problemi di stima parametrica. Segnaliamo che in questo capitolo *non* è richiesta la conoscenza dei vettori aleatori assolutamente continui, ad eccezione della dimostrazione del Lemma 8.12, che comunque può essere omessa.

8.1 Modelli statistici parametrici

Se il calcolo delle probabilità fornisce modelli matematici per fenomeni aleatori, la *statistica* ha lo scopo di confrontare il modello con “dati sperimentali”, traendone opportune conclusioni. Ad esempio, in *statistica parametrica*, viene proposta una classe di modelli dipendenti da un parametro; avendo a disposizione dati sperimentali, ci si chiede per esempio quale sia il valore del parametro per cui si ha il miglior accordo con i dati.

Definizione 8.1 (Modello statistico parametrico). Si dice *modello statistico (parametrico)* una famiglia di spazi di probabilità $(\Omega, \mathcal{A}, P_\theta)$ indicizzati da $\theta \in \Theta$, dove Θ viene detto l'*insieme dei parametri*. Si noti che l'insieme Ω e la σ -algebra \mathcal{A} sono sempre gli stessi: è la probabilità P_θ che dipende da θ .

Un modello statistico è dunque una classe di modelli per un fenomeno aleatorio. Nell'approccio *classico* alla statistica, si assume che *esista un valore di θ che fornisce il modello corretto*: tale valore va stimato sulla base di osservazioni.

Definizione 8.2 (Campione). Sia $\{(\Omega, \mathcal{A}, P_\theta)\}_{\theta \in \Theta}$ un modello statistico. Si dice *campione* ogni successione $(X_n)_{n \geq 1}$ di variabili aleatorie reali definite su Ω che, per ogni probabilità P_θ , $\theta \in \Theta$, sono indipendenti ed identicamente distribuite (con distribuzione che, naturalmente, può dipendere da θ). La sequenza finita (X_1, X_2, \dots, X_n) è detta *campione di taglia n*.

Le variabili X_n rappresentano le osservazioni riguardanti il fenomeno aleatorio. Il caso tipico, *che assumeremo sempre nel seguito*, è quello in cui la distribuzione marginale delle X_n rispetto a P_θ sia discreta oppure assolutamente continua:

- nel caso in cui le X_n siano variabili aleatorie discrete, denoteremo con $p(x; \theta) = P_\theta(X_1 = x)$ il valore in $x \in \mathbb{R}$ della loro comune densità discreta;
- analogamente, se le X_n sono variabili aleatorie assolutamente continue, denotiamo con $f(x; \theta)$ la loro comune densità, rispetto a P_θ .

Definizione 8.3 (Statistica campionaria). Sia $(X_n)_{n \geq 1}$ un campione (per un modello statistico fissato). Una successione $(Y_n)_{n \geq 1}$ di variabili aleatorie della forma

$$Y_n = h_n(X_1, X_2, \dots, X_n),$$

dove $h_n : \mathbb{R}^n \rightarrow \mathbb{R}$ è una funzione misurabile, viene detta *statistica campionaria*. La singola variabile aleatoria Y_n viene chiamata *statistica campionaria basata su un campione di taglia n*.

Nel contesto della stima parametrica, alcune statistiche campionarie verranno chiamate *stimatori*: non si tratta di una nuova nozione, ma un termine che suggerisce il ruolo “speciale” che, come vedremo, hanno alcune statistiche campionarie.

Definizione 8.4 (Stimatore corretto). Sia $\{(\Omega, \mathcal{A}, P_\theta)\}_{\theta \in \Theta}$ un modello statistico e sia $h : \Theta \rightarrow \mathbb{R}$ una funzione. Una statistica campionaria $(Y_n)_{n \geq 1}$ si dice *stimatore corretto* per $h(\theta)$ se, per ogni $\theta \in \Theta$ e per ogni $n \geq 1$, si ha $Y_n \in L^1(\Omega, \mathcal{A}, P_\theta)$ e

$$E_\theta(Y_n) = h(\theta),$$

dove E_θ denota il valor medio rispetto alla probabilità P_θ .

Esempio 8.5. Consideriamo un modello statistico per il quale $p(x; \theta)$ è la densità discreta di una variabile di Poisson di parametro $\theta \in \Theta := (0, +\infty)$. Poniamo

$$Y_n := \bar{X}_n.$$

Si vede facilmente che $(Y_n)_{n \geq 1}$ è uno stimatore corretto per $h(\theta) = \theta$. □

Esempio 8.6 (Campioni normali). Sia dato un modello statistico per il quale $f(x; \theta)$ è la densità di una variabile aleatoria normale di parametri $\theta = (\mu, \sigma^2) \in \Theta := \mathbb{R} \times (0, +\infty)$. Consideriamo le statistiche campionarie

$$\bar{X}_n := \frac{1}{n} \sum_{i=1}^n X_i, \quad S_n^2 := \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2. \quad (8.1)$$

La variabile aleatoria S_n^2 , che è definita per $n \geq 2$, viene chiamata *varianza campionaria* delle variabili aleatorie X_1, X_2, \dots, X_n . Il fatto che \bar{X}_n sia uno stimatore corretto

per μ segue immediatamente dalla linearità del valor medio. Inoltre

$$\begin{aligned}
 (n-1)E(S_n^2) &= \sum_{i=1}^n E[(X_i - \bar{X}_n)^2] = \sum_{i=1}^n [E(X_i^2) + E(\bar{X}_n^2) - 2E(X_i \bar{X}_n)] \\
 &= \left(\sum_{i=1}^n E(X_i^2) \right) + nE(\bar{X}_n^2) - 2E\left[\left(\sum_{i=1}^n X_i\right)\bar{X}_n\right] \\
 &= \left(\sum_{i=1}^n E(X_i^2) \right) - nE(\bar{X}_n^2) = n(\mu^2 + \sigma^2) - n\left(\mu^2 + \frac{\sigma^2}{n}\right) \\
 &= (n-1)\sigma^2,
 \end{aligned} \tag{8.2}$$

dunque S_n^2 è uno stimatore corretto per σ^2 . \square

Osservazione 8.7. L'Esempio 8.6 è particolarmente rilevante per le applicazioni; ne discuteremo più approfonditamente nel prossimo paragrafo. Tuttavia, nella verifica della correttezza degli stimatori, la *distribuzione normale non gioca alcun ruolo*. Si può quindi generalizzare il risultato nel modo seguente: dati un modello statistico $\{(\Omega, \mathcal{A}, P_\theta)\}_{\theta \in \Theta}$ e un campione $(X_n)_{n \geq 1}$ tali che $X_n \in L^2(\Omega, \mathcal{A}, P_\theta)$, per ogni $\theta \in \Theta$, poniamo $\mu(\theta) := E(X_n)$ e $\sigma^2(\theta) := \text{Var}(X_n)$; allora gli stimatori \bar{X}_n e S_n^2 sono corretti per $\mu(\theta)$ e $\sigma^2(\theta)$ rispettivamente. \square

Le nozioni fin qui introdotte si adattano bene a fenomeni aleatori in cui si possano eseguire misure ripetute e indipendenti. Le variabili aleatorie $(X_n)_{n \geq 1}$ del campione corrispondono in questo caso alle misure successive. Una sequenza di n dati x_1, x_2, \dots, x_n va intesa come una *realizzazione* $X_1(\omega), X_2(\omega), \dots, X_n(\omega)$ delle variabili aleatorie X_1, X_2, \dots, X_n . Se $Y_n = h_n(X_1, X_2, \dots, X_n)$ è uno stimatore per $h(\theta)$, il valore $h_n(x_1, x_2, \dots, x_n)$ viene detto una *stima* di $h(\theta)$ basata sui dati x_1, x_2, \dots, x_n . Risulta piuttosto naturale cercare stimatori la cui distribuzione sia il più possibile “concentrata” attorno alla funzione del parametro che stimano, in modo da far risultare “più probabile” che il valore $h_n(X_1, X_2, \dots, X_n)$ ottenuto con n osservazioni sia effettivamente “vicino” a $h(\theta)$.

In un dato modello statistico parametrico, vi sono usualmente molti stimatori corretti per una assegnata funzione $h(\theta)$ del parametro θ . Un indice, imperfetto ma efficace, di quanto la distribuzione di una variabile aleatoria reale sia concentrata attorno al valor medio è dato, come sappiamo, dalla varianza. Pertanto, se uno stimatore Y_n è corretto per $h(\theta)$, ossia $E_\theta(Y_n) = h(\theta)$, la varianza $\text{Var}_\theta(Y_n)$ è una misura dell'efficienza dello stimatore. Mostriamo ora che, se il modello statistico (e lo stimatore in esame) soddisfano opportune ipotesi di regolarità, la varianza di uno stimatore corretto non può scendere al di sotto di un certo valore.

Teorema 8.8 (Cramér-Rao). Sia Θ un intervallo di \mathbb{R} e sia $h : \Theta \rightarrow \mathbb{R}$ una funzione derivabile. Siano quindi $\{(\Omega, \mathcal{A}, P_\theta)\}_{\theta \in \Theta}$ un modello statistico e $(X_n)_{n \in \mathbb{N}}$ un relativo campione, tali che le derivate $\frac{d}{d\theta} p(x; \theta)$ (risp. $\frac{d}{d\theta} f(x; \theta)$ nel caso assolutamente continuo) esistano su tutto Θ .

Sia $Y_n = h_n(X_1, X_2, \dots, X_n)$ uno stimatore corretto per $h(\theta)$. Supponiamo che valgano le seguenti commutazioni di somma (risp. integrale) e derivata:

$$\sum_{x \in \mathbb{R}} \frac{d}{d\theta} p(x; \theta) = \frac{d}{d\theta} \sum_{x \in \mathbb{R}} p(x; \theta) = \frac{d}{d\theta} 1 = 0, \quad (8.3)$$

$$\begin{aligned} \sum_{x_1, x_2, \dots, x_n \in \mathbb{R}} h_n(x_1, x_2, \dots, x_n) \frac{d}{d\theta} [p(x_1; \theta) p(x_2; \theta) \cdots p(x_n; \theta)] \\ = \frac{d}{d\theta} \sum_{x_1, x_2, \dots, x_n \in \mathbb{R}} h_n(x_1, x_2, \dots, x_n) p(x_1; \theta) p(x_2; \theta) \cdots p(x_n; \theta) \quad (8.4) \\ = \frac{d}{d\theta} E(h_n(X_1, X_2, \dots, X_n)) = \frac{d}{d\theta} E_\theta(Y_n) = h'(\theta). \end{aligned}$$

(La formulazione delle analoghe condizioni per il caso assolutamente continuo è lasciata al lettore). Allora, indicando con Var_θ la varianza rispetto a P_θ ,

$$\text{Var}_\theta(Y_n) \geq V_n(\theta) := \frac{1}{n} \frac{[h'(\theta)]^2}{E_\theta \left[\left(\frac{d}{d\theta} \log p(X_1; \theta) \right)^2 \right]} \quad (8.5)$$

(con f al posto di p nel caso assolutamente continuo).

La quantità $V_n(\theta)$ che compare in (8.5) è detta *limite inferiore di Cramér-Rao*. Uno stimatore corretto $(Y_n)_{n \geq 1}$ che, per ogni n , raggiunge il limite inferiore di Cramér-Rao, ossia per cui vale l'uguaglianza in (8.5), viene dello *efficiente*.

Osservazione 8.9. Nella definizione (8.5) di $V_n(\theta)$ compare la quantità

$$E_\theta \left[\left(\frac{d}{d\theta} \log p(X_1; \theta) \right)^2 \right] = \sum_{x \in \mathbb{R}} \left(\frac{d}{d\theta} \log p(x; \theta) \right)^2 p(x; \theta) = \sum_{x \in \mathbb{R}} \left(\frac{\frac{d}{d\theta} p(x; \theta)}{p(x; \theta)} \right)^2 p(x; \theta).$$

Si noti che l'ultima espressione è sempre ben definita, *convenendo* che gli addendi della somma siano uguali a zero ogniqualevolta $p(x; \theta) = 0$ (un'analogha osservazione vale nel caso assolutamente continuo, con l'integrale al posto della somma). \square

Osservazione 8.10. Le condizioni (8.3) e (8.4) non sono troppo difficili da verificare in molti casi interessanti, anche se gli strumenti analitici necessari vanno oltre gli scopi di questo libro. Si può dimostrare che una condizione sufficiente perché valgano (8.3) e (8.4) è che

$$\{x \in \mathbb{R} : p(x; \theta) > 0\} \text{ non dipenda da } \theta,$$

e che per ogni $\theta \in \Theta$ esista $\varepsilon > 0$ tale che

$$\sum_{x \in \mathbb{R}} \sup_{t: |\theta - t| < \varepsilon} \left| \frac{d}{d\theta} p(x; t) \right| < +\infty,$$

$$\sum_{x_1, x_2, \dots, x_n \in \mathbb{R}} |h_n(x_1, x_2, \dots, x_n)| \sup_{t: |\theta - t| < \varepsilon} \left| \frac{d}{d\theta} [p(x_1; t) p(x_2; t) \cdots p(x_n; t)] \right| < +\infty.$$

Una condizione sufficiente del tutto analoga vale nel caso assolutamente continuo. \square

Dimostrazione (del Teorema 8.8). Forniamo la dimostrazione nel caso discreto. Le modifiche per il caso assolutamente continuo si limitano alle notazioni. Cominciamo col notare che vale l'identità

$$\frac{d}{d\theta} [p(x_1; \theta) p(x_2; \theta) \cdots p(x_n; \theta)] = \sum_{i=1}^n \frac{d}{d\theta} \log p(x_i; \theta) [p(x_1; \theta) p(x_2; \theta) \cdots p(x_n; \theta)], \quad (8.6)$$

la cui semplice verifica è lasciata al lettore. Per la condizione (8.4)

$$\begin{aligned} h'(\theta) &= \sum_{x_1, x_2, \dots, x_n \in \mathbb{R}} h_n(x_1, x_2, \dots, x_n) \frac{d}{d\theta} [p(x_1; \theta) p(x_2; \theta) \cdots p(x_n; \theta)] \\ &= \sum_{i=1}^n \sum_{x_1, x_2, \dots, x_n \in \mathbb{R}} h_n(x_1, x_2, \dots, x_n) \frac{d}{d\theta} \log p(x_i; \theta) [p(x_1; \theta) p(x_2; \theta) \cdots p(x_n; \theta)] \quad (8.7) \\ &= \sum_{i=1}^n E_{\theta} \left[Y_n \frac{d}{d\theta} \log p(X_i; \theta) \right] = \text{Cov}_{\theta} \left(Y_n, \sum_{i=1}^n \frac{d}{d\theta} \log p(X_i; \theta) \right), \end{aligned}$$

dove nell'ultima uguaglianza abbiamo usato il fatto che, per la condizione (8.3),

$$E_{\theta} \left[\frac{d}{d\theta} \log p(X_i; \theta) \right] = \sum_{x \in \mathbb{R}} \frac{d}{d\theta} p(x; \theta) = 0.$$

Per la disuguaglianza di Cauchy-Schwarz,

$$\begin{aligned} \left[\text{Cov}_{\theta} \left(Y_n, \sum_{i=1}^n \frac{d}{d\theta} \log p(X_i; \theta) \right) \right]^2 &\leq \text{Var}_{\theta}(Y_n) \text{Var}_{\theta} \left(\sum_{i=1}^n \frac{d}{d\theta} \log p(X_i; \theta) \right) \\ &= n \text{Var}_{\theta}(Y_n) \text{Var}_{\theta} \left(\frac{d}{d\theta} \log p(X_1; \theta) \right), \end{aligned}$$

che, assieme a (8.7), fornisce

$$[h'(\theta)]^2 \leq n \text{Var}_{\theta}(Y_n) \text{Var}_{\theta} \left(\frac{d}{d\theta} \log p(X_1; \theta) \right),$$

da cui la conclusione segue facilmente. \square

Esercizi

Esercizio 8.1. Si mostri che $Y_n := \bar{X}_n$ fornisce uno stimatore corretto e efficiente per θ per i modelli statistici determinati da:

- (i) $p(x; \theta) := e^{-\theta} \frac{\theta^x}{x!}$, con $x \in \mathbb{N}_0$ e $\theta \in \Theta = (0, +\infty)$;
- (ii) $f(x; \theta) := \frac{1}{\sqrt{2\pi\sigma^2}} \exp[-(x - \theta)^2 / 2\sigma^2]$, dove $\theta \in \mathbb{R}$ e $\sigma > 0$ è una costante che viene assunta nota.

Esercizio 8.2. Si consideri il modello statistico per cui

$$f(x; \theta) := \theta e^{-\theta x} \mathbb{1}_{(0, +\infty)}(x),$$

per $\theta \in \Theta = (0, +\infty)$.

- (i) Si mostri che $Y_n := \bar{X}_n$ è uno stimatore corretto ed efficiente per $h(\theta) = 1/\theta$.
- (ii) Si mostri che $Y_n := \frac{n-1}{\sum_{i=1}^n X_i}$ è uno stimatore corretto, ma non efficiente, per θ .

8.2 Intervalli di confidenza per campioni normali

La nozione che ora introduciamo formula in modo preciso l'idea di “margine di errore” nella stima di un parametro.

Definizione 8.11 (Intervallo di confidenza). Sia $\{(\Omega, \mathcal{A}, P_\theta)\}_{\theta \in \Theta}$ un modello statistico, $(X_n)_{n \geq 1}$ un corrispondente campione e $f: \Theta \rightarrow \mathbb{R}$ una funzione. Siano inoltre assegnate, per ogni n , due statistiche campionarie

$$U_n = u_n(X_1, X_2, \dots, X_n), \quad V_n = v_n(X_1, X_2, \dots, X_n),$$

basate su un campione di taglia n , tali che $U_n \leq V_n$.

Si dice che l'intervallo aleatorio $[U_n, V_n]$ è un *intervallo di confidenza* (basato su un campione di taglia n) per $f(\theta)$ di *livello di confidenza* $\gamma \in (0, 1)$ se

$$P_\theta(U_n \leq f(\theta) \leq V_n) \geq \gamma, \quad \forall \theta \in \Theta. \quad (8.8)$$

Con abuso di linguaggio, in corrispondenza di una sequenza x_1, x_2, \dots, x_n di n dati, l'intervallo $[u_n(x_1, x_2, \dots, x_n), v_n(x_1, x_2, \dots, x_n)]$ viene detto *intervallo di confidenza* per $f(\theta)$ di *livello di confidenza* γ .

L'effettiva determinazione di stimatori e intervalli di confidenza è, in generale, un problema difficile. Un'eccezione di grande rilevanza applicativa è data dai campioni normali, descritti nell'Esempio 8.6, per cui è possibile un'analisi completa e dettagliata, che presentiamo in questo paragrafo. Per modelli statistici più generali, risulterà conveniente indebolire la nozione di intervallo di confidenza, richiedendo

che (8.8) valga solo asintoticamente per $n \rightarrow \infty$, come vedremo nel Paragrafo 8.3. Uno dei possibili approcci per determinare intervalli di confidenza in questo contesto verrà presentato nel Paragrafo 8.4.

D'ora in avanti consideriamo dunque il modello statistico dell'Esempio 8.6, per il quale $f(x; \theta)$ è la densità di una variabile aleatoria normale di parametri $\theta = (\mu, \sigma^2) \in \Theta := \mathbb{R} \times (0, +\infty)$. Abbiamo già individuato gli stimatori corretti per μ e σ^2 , dati rispettivamente dalla media e varianza campionarie \bar{X}_n e S_n^2 , definite in (8.1). Il nostro scopo è ora di determinare intervalli di confidenza per μ e σ^2 , per ogni fissato livello di confidenza γ . Il primo passo è costituito dal seguente risultato. Si ricordi che le variabili aleatorie $\chi^2(n)$ sono state introdotte nel Paragrafo 6.5.1.

Lemma 8.12. *Siano $Z \sim N(0, 1)$ e $V \sim \chi^2(n) = \text{Gamma}(\frac{n}{2}, \frac{1}{2})$ variabili aleatorie indipendenti. Poniamo*

$$T := \frac{Z}{\sqrt{V/n}}.$$

Allora T è una variabile aleatoria assolutamente continua, con densità

$$f_T(x) = \frac{c_n}{(1 + \frac{x^2}{n})^{\frac{n+1}{2}}}, \quad \text{con} \quad c_n := \frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi} \Gamma(\frac{n}{2})}. \quad (8.9)$$

Dimostrazione. Usiamo la Proposizione 6.49. Per ogni n fissato, la funzione

$$\varphi(x, v) := \left(\frac{x}{\sqrt{v/n}}, v \right)$$

è un diffeomorfismo di $\mathbb{R} \times (0, +\infty)$ in sé, la cui funzione inversa φ^{-1} è data da

$$\varphi^{-1}(t, w) := \left(t\sqrt{w/n}, w \right).$$

Posto $(T, W) := \varphi(Z, V)$, per la Proposizione 6.49:

$$\begin{aligned} f_{T,W}(t, w) &= f_Z(t\sqrt{w/n}) f_V(w) |\det \varphi^{-1}(t, w)| \\ &= \frac{1}{\sqrt{2\pi} 2^{n/2} \Gamma(\frac{n}{2})} e^{-\frac{1}{2} \frac{w}{n} t^2} w^{\frac{n}{2}-1} e^{-w/2} \sqrt{w/n} \mathbb{1}_{(0, +\infty)}(w) \\ &= \frac{1}{\sqrt{\pi} 2^{\frac{n+1}{2}} \Gamma(\frac{n}{2}) \sqrt{n}} w^{\frac{n-1}{2}} e^{-\frac{1}{2} (1 + \frac{t^2}{n}) w} \mathbb{1}_{(0, +\infty)}(w). \end{aligned}$$

Per calcolare $f_T(t) = \int_{\mathbb{R}} f_{T,W}(t, w) dw$, osserviamo che con un cambio di variabili

$$\begin{aligned} \int_0^{+\infty} w^{\frac{n-1}{2}} e^{-\frac{1}{2} (1 + \frac{t^2}{n}) w} dw &= \left(\frac{1}{2} (1 + \frac{t^2}{n}) \right)^{-\frac{n+1}{2}} \int_0^{+\infty} x^{\frac{n-1}{2}} e^{-x} dx \\ &= \frac{2^{\frac{n+1}{2}}}{(1 + \frac{t^2}{n})^{\frac{n+1}{2}}} \Gamma\left(\frac{n+1}{2}\right), \end{aligned}$$

avendo ricordato la definizione della funzione $\Gamma(\cdot)$. Di conseguenza

$$f_T(t) = \int_{\mathbb{R}} f_{T,W}(t, w) dw = \frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi} \Gamma(\frac{n}{2})} \frac{1}{(1 + \frac{t^2}{n})^{\frac{n+1}{2}}},$$

come volevasi dimostrare. \square

Una variabile aleatoria reale assolutamente continua T , la cui densità sia data da (8.9), è detta *t-di-Student a n gradi di libertà*, e scriveremo $T \sim t(n)$.

Proposizione 8.13. *Siano \bar{X}_n e S_n^2 rispettivamente la media campionaria e la varianza campionaria per una successione $(X_k)_{k \in \mathbb{N}}$ di variabili aleatorie i.i.d. con distribuzione $N(\mu, \sigma^2)$. Allora, per ogni $n \geq 2$,*

$$\frac{\bar{X}_n - \mu}{S_n} \sqrt{n} \sim t(n-1).$$

Dimostrazione. Ricordiamo la Proposizione 6.68: per ogni $n \in \mathbb{N}$

$$Z_n := \frac{\bar{X}_n - \mu}{\sigma} \sqrt{n} \sim N(0, 1), \quad (8.10)$$

$$V_n := \frac{1}{\sigma^2} \sum_{i=1}^n (X_i - \bar{X}_n)^2 = \frac{(n-1)S_n^2}{\sigma^2} \sim \chi^2(n-1), \quad (8.11)$$

e inoltre Z_n e V_n sono variabili aleatorie indipendenti. Osservando che

$$\frac{\bar{X}_n - \mu}{S_n} \sqrt{n} = \frac{Z_n}{\sqrt{V_n/(n-1)}},$$

la conclusione segue dal Lemma 8.12. \square

Prima di enunciare e dimostrare il risultato principale di questo paragrafo, introduciamo alcune notazioni. Sia $T \sim t(n)$. La sua funzione di ripartizione F_T è strettamente crescente, dato che la sua derivata è la densità calcolata nel Lemma 8.12, che è ovunque strettamente positiva. Pertanto $F_T : \mathbb{R} \rightarrow (0, 1)$ è invertibile. Poniamo dunque, per $\alpha \in (0, 1)$,

$$t_\alpha(n) := F_T^{-1}(1 - \alpha). \quad (8.12)$$

Analogamente, se $Y \sim \chi^2(n)$, la funzione di ripartizione F_Y è invertibile come funzione da $[0, +\infty)$ in $(0, 1)$, e poniamo

$$\chi_\alpha^2(n) := F_Y^{-1}(1 - \alpha). \quad (8.13)$$

Mettiamo in guardia da possibili confusioni indotte dalle notazioni (che sono quelle standard in letteratura): $t_\alpha(n)$ e $\chi_\alpha^2(n)$ sono numeri reali, per ogni $\alpha \in (0, 1)$ e $n \in \mathbb{N}$, mentre $t(n)$ e $\chi^2(n)$ indicano delle distribuzioni.

Teorema 8.14 (Intervalli di confidenza per un campione normale). Sia $(X_n)_{n \geq 1}$ un campione con distribuzione $N(\mu, \sigma^2)$. Indicando con \bar{X}_n e S_n rispettivamente media e varianza campionaria, cf. (8.1), per ogni $\alpha \in (0, 1)$:

(i) l'intervallo aleatorio

$$\left[\bar{X}_n - \frac{S_n}{\sqrt{n}} t_{\alpha/2}(n-1), \bar{X}_n + \frac{S_n}{\sqrt{n}} t_{\alpha/2}(n-1) \right]$$

è un intervallo di confidenza per μ di livello di confidenza $1 - \alpha$;

(ii) l'intervallo aleatorio

$$\left[\frac{(n-1)S_n^2}{\chi_{\alpha/2}^2(n-1)}, \frac{(n-1)S_n^2}{\chi_{1-\alpha/2}^2(n-1)} \right]$$

è un intervallo di confidenza per σ^2 di livello di confidenza $1 - \alpha$.

Dimostrazione. Si noti che possiamo riscrivere

$$\begin{aligned} & \mathbb{P} \left(\bar{X}_n - \frac{S_n}{\sqrt{n}} t_{\alpha/2}(n-1) \leq \mu \leq \bar{X}_n + \frac{S_n}{\sqrt{n}} t_{\alpha/2}(n-1) \right) \\ &= \mathbb{P} \left(\left| \frac{\bar{X}_n - \mu}{S_n} \sqrt{n} \right| \leq t_{\alpha/2}(n-1) \right) = \mathbb{P}(|T_n| \leq t_{\alpha/2}(n-1)), \end{aligned}$$

avendo posto

$$T_n := \frac{\bar{X}_n - \mu}{S_n} \sqrt{n}.$$

Per la Proposizione 8.13 si ha $T_n \sim t(n-1)$. Dato che la densità (8.9) di T_n è una funzione pari, le variabili aleatorie T_n e $-T_n$ hanno la stessa distribuzione, dunque $F_T(-x) = 1 - F_T(x)$ per ogni $x \in \mathbb{R}$. Di conseguenza

$$\begin{aligned} \mathbb{P}(|T_n| \leq t_{\alpha/2}(n-1)) &= \mathbb{P}(T_n \leq t_{\alpha/2}(n-1)) - \mathbb{P}(T_n \leq -t_{\alpha/2}(n-1)) \\ &= F_{T_n}(t_{\alpha/2}(n-1)) - F_{T_n}(-t_{\alpha/2}(n-1)) \\ &= 2F_{T_n}(t_{\alpha/2}(n-1)) - 1 = 2(1 - \frac{\alpha}{2}) - 1 = 1 - \alpha, \end{aligned}$$

avendo usato nella penultima uguaglianza il fatto che $T_n \sim t(n-1)$ e la definizione di $t_{\alpha/2}(n-1)$. Questo completa la dimostrazione del punto (i).

Per il punto (ii), posto

$$V_n := \frac{(n-1)S_n^2}{\sigma^2},$$

sappiamo che $V_n \sim \chi^2(n-1)$ (si ricordi la relazione (8.11)). Perciò

$$\begin{aligned} P\left(\frac{(n-1)S_n^2}{\chi_{\alpha/2}^2(n-1)} \leq \sigma^2 \leq \frac{(n-1)S_n^2}{\chi_{1-\alpha/2}^2(n-1)}\right) \\ = P\left(\chi_{1-\alpha/2}^2(n-1) \leq V_n \leq \chi_{\alpha/2}^2(n-1)\right) \\ = F_{V_n}(\chi_{\alpha/2}^2(n-1)) - F_{V_n}(\chi_{1-\alpha/2}^2(n-1)) \\ = (1 - \frac{\alpha}{2}) - (1 - (1 - \frac{\alpha}{2})) = 1 - \alpha, \end{aligned}$$

e la dimostrazione è completata. \square

8.3 Proprietà asintotiche

Come abbiamo visto nel paragrafo precedente, nel caso di campioni normali è possibile determinare stimatori corretti per media e varianza, nonché intervalli di confidenza ad essi associati. In casi più generali, può essere assai difficile determinare stimatori e intervalli aleatori con le proprietà desiderate, cioè, rispettivamente, la correttezza ed un assegnato livello di confidenza. È pertanto opportuno indebolire tali proprietà, richiedendone la validità non *per ogni* n , ma solo nel limite per $n \rightarrow +\infty$. Nel Paragrafo 8.4 vedremo come tali nozioni asintotiche risultino utili in un'ampia classe di modelli statistici.

Cominciamo indebolendo la nozione di correttezza per uno stimatore.

Definizione 8.15 (Stimatore asintoticamente corretto). Sia $\{(\Omega, \mathcal{A}, P_\theta)\}_{\theta \in \Theta}$ un modello statistico e sia $h: \Theta \rightarrow \mathbb{R}$ una funzione. Una statistica campionaria $(Y_n)_{n \geq 1}$ si dice *stimatore asintoticamente corretto per* $h(\theta)$ se per ogni $\theta \in \Theta$ si ha $Y_n \in L^1(\Omega, \mathcal{A}, P_\theta)$ e

$$\lim_{n \rightarrow +\infty} E_\theta(Y_n) = h(\theta).$$

Esempio 8.16. Consideriamo un modello statistico per cui la densità $f(x; \theta)$ è data da

$$f(x; \theta) = \frac{1}{\theta} \mathbb{1}_{(0, \theta)}(x),$$

dove $\theta \in \Theta = (0, +\infty)$. In altre parole, $X_n \sim U(0, \theta)$. Definiamo

$$Y_n := \max(X_1, X_2, \dots, X_n).$$

Anzitutto calcoliamo la funzione di ripartizione di Y_n . Per $x \in (0, \theta)$ abbiamo

$$F_{Y_n}(x; \theta) = P_\theta(Y_n \leq x) = P_\theta(X_1 \leq x, X_2 \leq x, \dots, X_n \leq x) = [P_\theta(X_1 \leq x)]^n = \frac{x^n}{\theta^n}.$$

Derivando, si ottiene

$$f_{Y_n}(x; \theta) = \frac{nx^{n-1}}{\theta^n} \mathbb{1}_{(0, \theta)}(x).$$

Quindi

$$E_{\theta}(Y_n) = \frac{n}{\theta^n} \int_0^{\theta} x^n dx = \frac{n}{n+1} \theta,$$

da cui si vede che Y_n è uno stimatore asintoticamente corretto per θ . \square

Definiamo ora un'altra importante proprietà asintotica per uno stimatore.

Definizione 8.17 (Stimatore consistente). Sia $\{(\Omega, \mathcal{A}, P_{\theta})\}_{\theta \in \Theta}$ un modello statistico e sia $h : \Theta \rightarrow \mathbb{R}$ una funzione. Una statistica campionaria $(Y_n)_{n \geq 1}$ si dice *stimatore consistente per $h(\theta)$* se per ogni $\theta \in \Theta$ e ogni $\varepsilon > 0$

$$\lim_{n \rightarrow +\infty} P_{\theta}(|Y_n - h(\theta)| > \varepsilon) = 0.$$

In termini intuitivi, possiamo dire che se $(Y_n)_{n \geq 1}$ è uno stimatore consistente per $h(\theta)$, allora Y_n , per n grande, è una funzione delle osservazioni che, con grande probabilità rispetto a P_{θ} , assume valori vicini a $h(\theta)$.

Esempio 8.18. Consideriamo la statistica dell'Esempio 8.5. La consistenza dello stimatore deriva immediatamente dalla legge dei grandi numeri. \square

Esempio 8.19. Consideriamo la statistica dell'Esempio 8.16. Per costruzione si ha $P_{\theta}(Y_n \in (0, \theta)) = 1$, pertanto

$$P_{\theta}(|Y_n - \theta| > \varepsilon) = P_{\theta}(Y_n < \theta - \varepsilon).$$

Se $\varepsilon \geq \theta$ queste probabilità valgono zero, e non c'è nulla da dimostrare. Se invece $0 < \varepsilon < \theta$,

$$P_{\theta}(|Y_n - \theta| > \varepsilon) = P_{\theta}(Y_n < \theta - \varepsilon) = F_{Y_n}(\theta - \varepsilon; \theta) = \left(\frac{\theta - \varepsilon}{\theta}\right)^n,$$

da cui la consistenza segue immediatamente. \square

Definizione 8.20 (Intervallo di confidenza asintotico). Sia $\{(\Omega, \mathcal{A}, P_{\theta})\}_{\theta \in \Theta}$ un modello statistico, $(X_n)_{n \geq 1}$ un corrispondente campione e $f : \Theta \rightarrow \mathbb{R}$ una funzione. Siano inoltre assegnate, per ogni n , due statistiche campionarie

$$U_n = u_n(X_1, X_2, \dots, X_n), \quad V_n = v_n(X_1, X_2, \dots, X_n),$$

basate su un campione di taglia n , tali che $U_n \leq V_n$.

Si dice che la famiglia di intervalli aleatori $([U_n, V_n])_{n \in \mathbb{N}}$ è un *intervallo di confidenza asintotico per $f(\theta)$ di livello di confidenza $\gamma \in (0, 1)$* se

$$\liminf_{n \rightarrow +\infty} P_{\theta}(U_n \leq h(\theta) \leq V_n) \geq \gamma, \quad \forall \theta \in \Theta.$$

La prossima nozione risulta assai utile nella determinazione di intervalli di confidenza asintotici. In quanto segue, se $(Z_n)_{n \geq 1}$ è una successione di variabili aleatorie

definite su $(\Omega, \mathcal{A}, P_\theta)$ a valori reali, scriveremo

$$Z_n \xrightarrow{D} N(0, 1)$$

come abbreviazione per

$$\forall z \in \mathbb{R} : \lim_{n \rightarrow +\infty} P_\theta(Z_n \leq z) = P(Z \leq z), \quad \text{con } Z \sim N(0, 1).$$

Definizione 8.21 (Stimatore asintoticamente normale). Sia $\{(\Omega, \mathcal{A}, P_\theta)\}_{\theta \in \Theta}$ un modello statistico e sia $h : \Theta \rightarrow \mathbb{R}$ una funzione. Una statistica campionaria $(Y_n)_{n \geq 1}$ si dice *stimatore asintoticamente normale per $h(\theta)$* se per ogni $\theta \in \Theta$ esiste una costante positiva $\sigma(\theta)$, che può dipendere da θ , tale che

$$\sqrt{n} \frac{Y_n - h(\theta)}{\sigma(\theta)} \xrightarrow{D} N(0, 1). \quad (8.14)$$

La normalità asintotica di uno stimatore permette di quantificare l'*errore di stima* che si commette nello stimare $h(\theta)$ con Y_n . Sia $0 < \alpha < 1$ e definiamo

$$z_\alpha := \Phi^{-1}(1 - \alpha),$$

dove ricordiamo che $\Phi(x) := P(Z \leq x)$ con $Z \sim N(0, 1)$. Si verifica facilmente che

$$P(|Z| \leq z_{\alpha/2}) = 1 - \alpha.$$

Segue allora da (8.14) che

$$\lim_{n \rightarrow +\infty} P_\theta \left(\left| \sqrt{n} \frac{Y_n - h(\theta)}{\sigma(\theta)} \right| \leq z_{\alpha/2} \right) = 1 - \alpha, \quad (8.15)$$

o, equivalentemente,

$$\lim_{n \rightarrow +\infty} P_\theta \left(h(\theta) \in \left[Y_n - \frac{\sigma(\theta)}{\sqrt{n}} z_{\alpha/2}, Y_n + \frac{\sigma(\theta)}{\sqrt{n}} z_{\alpha/2} \right] \right) = 1 - \alpha.$$

Dunque, con probabilità che tende a $1 - \alpha$ per $n \rightarrow +\infty$, lo stimatore Y_n dista dal valore $h(\theta)$ al più $\frac{\sigma(\theta)}{\sqrt{n}} z_{\alpha/2}$. Quest'ultima quantità, tuttavia, dipende dal parametro incognito θ . Nei casi in cui

$$\bar{\sigma} := \sup_{t \in \Theta} \sigma(t) < +\infty,$$

si ha che

$$\begin{aligned} P_\theta \left(h(\theta) \in \left[Y_n - \frac{\bar{\sigma}}{\sqrt{n}} z_{\alpha/2}, Y_n + \frac{\bar{\sigma}}{\sqrt{n}} z_{\alpha/2} \right] \right) \\ \geq P_\theta \left(h(\theta) \in \left[Y_n - \frac{\sigma(\theta)}{\sqrt{n}} z_{\alpha/2}, Y_n + \frac{\sigma(\theta)}{\sqrt{n}} z_{\alpha/2} \right] \right), \end{aligned}$$

quindi

$$\liminf_{n \rightarrow +\infty} P_{\theta} \left(h(\theta) \in \left[Y_n - \frac{\bar{\sigma}}{\sqrt{n}} z_{\alpha/2}, Y_n + \frac{\bar{\sigma}}{\sqrt{n}} z_{\alpha/2} \right] \right) \geq 1 - \alpha.$$

La famiglia di intervalli, al variare di $n \in \mathbb{N}$,

$$\left[Y_n - \frac{\bar{\sigma}}{\sqrt{n}} z_{\alpha/2}, Y_n + \frac{\bar{\sigma}}{\sqrt{n}} z_{\alpha/2} \right]$$

è perciò *intervallo di confidenza asintotico* per $h(\theta)$ di livello di confidenza $1 - \alpha$.

Proposizione 8.22. Sia $h : \Theta \rightarrow \mathbb{R}$ una funzione. Ogni stimatore asintoticamente normale $(Y_n)_{n \geq 1}$ per $h(\theta)$ è uno stimatore consistente per $h(\theta)$.

Dimostrazione. Si tratta di una conseguenza della relazione (8.15). Infatti, per ogni $\varepsilon > 0$, $\theta \in \Theta$ e $\alpha \in (0, 1)$ fissati, per n sufficientemente grande si ha $\varepsilon > z_{\alpha/2} \sigma(\theta) / \sqrt{n}$, pertanto

$$\liminf_{n \rightarrow +\infty} P_{\theta} (|Y_n - h(\theta)| \leq \varepsilon) \geq \liminf_{n \rightarrow +\infty} P_{\theta} \left(|Y_n - h(\theta)| \leq \frac{z_{\alpha/2} \sigma(\theta)}{\sqrt{n}} \right) = 1 - \alpha.$$

Per l'arbitrarietà di α , questo implica

$$\lim_{n \rightarrow +\infty} P_{\theta} (|Y_n - h(\theta)| \leq \varepsilon) = 1,$$

cioè la consistenza. □

Esempio 8.23. Lo stimatore ottenuto nell'Esempio 8.5 è asintoticamente normale, con $\sigma(\theta) = \sqrt{\theta}$. Ciò deriva direttamente dal teorema limite centrale. Si noti che, in questo caso, $\sigma(\theta)$ non è una funzione limitata di θ . □

Esempio 8.24. Consideriamo lo stimatore $(Y_n)_{n \geq 1}$ degli Esempi 8.16 e 8.19. Abbiamo visto che $P(Y_n \leq \theta) = 1$. Pertanto, per ogni scelta di $\sigma(\theta)$, la variabile aleatoria

$$\sqrt{n} \frac{Y_n - \theta}{\sigma(\theta)}$$

assume solo valori negativi, e quindi (perché?) non si può avere

$$\sqrt{n} \frac{Y_n - \theta}{\sigma(\theta)} \xrightarrow{D} N(0, 1).$$

Di conseguenza, lo stimatore non è asintoticamente normale. □

Dato un modello statistico ed un associato campione, vi sono varie tecniche per determinare stimatori con “buone” proprietà. Uno degli approcci più generali, e l'unico che tratteremo in questo testo, è descritto nel prossimo paragrafo.

Esercizi

Esercizio 8.3. Si consideri il modello statistico in cui $p(x; \theta)$ è la densità di una variabile di Bernoulli di parametro $\theta \in \Theta = (0, 1)$. Si mostri che lo stimatore $Y_n := \bar{X}_n$ è asintoticamente normale, e si determini un intervallo di confidenza asintotico per $h(\theta) = \theta$.

8.4 Stimatori di massima verosimiglianza

Consideriamo un modello statistico $(\Omega, \mathcal{A}, P_\theta)$ con $\theta \in \Theta$, e sia $(X_n)_{n \geq 1}$ un campione. Per il momento assumiamo che le X_n siano variabili aleatorie discrete, con densità $p(x; \theta)$. Consideriamo n numeri reali x_1, x_2, \dots, x_n , che interpretiamo come i *valori osservati* delle variabili X_1, X_2, \dots, X_n . È ragionevole pensare che il valore di θ che *meglio si accorda* ai valori osservati sia quel valore di θ che massimizza, rispetto a θ , la probabilità

$$P_\theta(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n) = p(x_1; \theta) p(x_2; \theta) \cdots p(x_n; \theta).$$

Assumendo che tale punto di massimo esista e che sia unico, esso dipenderà dai valori x_1, x_2, \dots, x_n , cioè sarà una *funzione* di x_1, x_2, \dots, x_n , che denoteremo con $\hat{\theta}_n(x_1, x_2, \dots, x_n)$. Per costruzione si ha

$$P_{\hat{\theta}_n(x_1, x_2, \dots, x_n)}(X_1 = x_1, \dots, X_n = x_n) = \max_{\theta \in \Theta} P_\theta(X_1 = x_1, \dots, X_n = x_n).$$

La successione $(Y_n := \hat{\theta}_n(X_1, X_2, \dots, X_n))_{n \geq 1}$ è detta *stimatore di massima verosimiglianza per il parametro θ* del modello statistico in esame.

Per rendere più precisa la nozione appena introdotta, e per dimostrare alcune proprietà, d'ora in avanti assumeremo che il modello statistico $(\Omega, \mathcal{A}, P_\theta)$ con $\theta \in \Theta$ e il relativo campione $(X_n)_{n \geq 1}$ soddisfino le seguenti ipotesi.

- (A) L'insieme Θ è un intervallo aperto di \mathbb{R} (che può anche essere una semiretta aperta, o tutto \mathbb{R}).
- (B) Le variabili aleatorie del campione $(X_n)_{n \geq 1}$ sono discrete, con densità $p(x; \theta)$, o assolutamente continue, con densità $f(x; \theta)$. Inoltre esiste $I \in \mathcal{B}(\mathbb{R})$ tale che
 - per ogni $\theta \in \Theta$, $P_\theta(X_1 \in I) = 1$;
 - per ogni $x \in I$ e $\theta \in \Theta$, si ha che $p(x; \theta) > 0$ (risp. $f(x; \theta) > 0$);
 - per ogni $x \in I$, la funzione $\theta \mapsto p(x; \theta)$ (risp. $\theta \mapsto f(x; \theta)$) è continua.
- (C) Per ogni $n \geq 1$ e $\underline{x} = (x_1, \dots, x_n) \in I^n$, nel caso discreto la funzione

$$\theta \mapsto L_n(\underline{x}, \theta) := \frac{1}{n} \sum_{i=1}^n \log p(x_i; \theta) \quad (8.16)$$

ha un unico punto di massimo locale proprio,[†] indicato con $\hat{\theta}(x_1, x_2, \dots, x_n)$; analogamente nel caso assolutamente continuo, con $f(x_i; \theta)$ al posto di $p(x_i; \theta)$.

(D) Le distribuzioni di X_1 rispetto a P_θ e P_t sono diverse, per ogni $\theta \neq t$.

(E) Per ogni $\theta, t \in \Theta$

$$E_\theta \left[(\log p(X_1, t))^2 \right] < +\infty,$$

(con $f(X_1, t)$ al posto di $p(X_1, t)$ nel caso assolutamente continuo).

Osservazione 8.25. Essendo il logaritmo una funzione strettamente crescente, massimizzare $L_n(\underline{x}, \theta)$ equivale a massimizzare $p(x_1; \theta)p(x_2; \theta) \cdots p(x_n; \theta)$. \square

Osservazione 8.26. La condizione (C) sull'unicità del massimo locale è delicata nel caso assolutamente continuo, in quanto il suo verificarsi può dipendere dalla *scelta* della densità di X_n che, come sappiamo, non è unica (si ricordi, ad esempio, che una densità può essere modificata su un insieme finito restando densità della stessa variabile casuale). È importante sottolineare che, sotto la condizione (C), la funzione $\hat{\theta} : I^n \rightarrow \mathbb{R}$ è misurabile; questo è rilevante per la prossima Definizione 8.27. \square

Per verificare che $\hat{\theta} : I^n \rightarrow \mathbb{R}$ è misurabile, basta mostrare che per ogni $t \in \mathbb{R}$

$$\{\underline{x} : \hat{\theta}(\underline{x}) \leq t\} \in \mathcal{B}(\mathbb{R}^n).$$

A tale scopo, si osservi che, sotto la condizione (C), l'affermazione $\hat{\theta}(\underline{x}) \leq t$ è equivalente a: “nessun $\theta > t$ è un massimo locale proprio per la funzione $s \mapsto L_n(\underline{x}, s)$ ”. In formule

$$\{\underline{x} : \hat{\theta}(\underline{x}) \leq t\} = \bigcap_{\substack{\theta > t \\ \theta \in \Theta}} \bigcup_{k \in \mathbb{N}} \bigcup_{\substack{s \in \Theta \\ |s - \theta| < \frac{1}{k}}} \{\underline{x} : L_n(\underline{x}, s) \geq L_n(\underline{x}, \theta)\}.$$

Per l'ipotesi (B), per ogni $\underline{x} \in I^n$ la funzione $s \mapsto L_n(\underline{x}, s)$ è continua. Pertanto, nella formula precedente, è sufficiente che la disuguaglianza $L_n(\underline{x}, s) \geq L_n(\underline{x}, \theta)$ sia verificata per s, θ in un sottoinsieme numerabile denso di Θ :

$$\{\underline{x} : \hat{\theta}(\underline{x}) \leq t\} = \bigcap_{\substack{\theta > t \\ \theta \in \Theta \cap \mathbb{Q}}} \bigcup_{k \in \mathbb{N}} \bigcup_{\substack{s \in \Theta \cap \mathbb{Q} \\ |s - \theta| < \frac{1}{k}}} \{\underline{x} : L_n(\underline{x}, s) \geq L_n(\underline{x}, \theta)\}. \quad (8.17)$$

Per s fissato, la funzione $\underline{x} \mapsto L_n(\underline{x}, s)$ è misurabile, in quanto la densità di una variabile aleatoria è una funzione misurabile. Ne segue che

$$\{\underline{x} : L_n(\underline{x}, s) \geq L_n(\underline{x}, \theta)\} \in \mathcal{B}(\mathbb{R}^n).$$

Quindi, per (8.17), l'insieme $\{\underline{x} : \hat{\theta}(\underline{x}) \leq t\}$ si ottiene con intersezioni e unioni numerabili di elementi di $\mathcal{B}(\mathbb{R}^n)$, e quindi è un elemento di $\mathcal{B}(\mathbb{R}^n)$.

[†] Diciamo che una funzione $h : \Theta \rightarrow \mathbb{R}$ ha un *massimo locale proprio* in $\theta \in \Theta$ se esiste $\varepsilon > 0$ tale che $h(t) < h(\theta)$ per ogni $t \in \Theta$ con $|t - \theta| < \varepsilon$, $t \neq \theta$.

Definizione 8.27 (Stimatore di massima verosimiglianza). Consideriamo un modello statistico $\{(\Omega, \mathcal{A}, P_\theta)\}_{\theta \in \Theta}$ e un relativo campione $(X_n)_{n \geq 1}$, che soddisfano le ipotesi (A)–(C). Poniamo $\hat{\theta}_n(x_1, x_2, \dots, x_n) = 0$ per $(x_1, x_2, \dots, x_n) \notin I^n$, e definiamo per ogni $n \geq 1$

$$Y_n = \hat{\theta}_n(X_1, X_2, \dots, X_n). \quad (8.18)$$

La statistica campionaria $(Y_n)_{n \geq 1}$ si dice *stimatore di massima verosimiglianza*, o per brevità *stimatore MV*, per il parametro θ del modello statistico in esame.

Dalle ipotesi sopra elencate ricaveremo la consistenza dello stimatore MV. Per semplicità, limitiamo enunciati e dimostrazioni al caso discreto. Gli enunciati si modificano in modo ovvio nel caso assolutamente continuo; le dimostrazioni richiedono invece qualche attenzione in più, e saranno omesse.

Lemma 8.28. *Supponiamo che valgano le ipotesi (A)–(E). Allora per ogni $\theta, t \in \Theta$ la famiglia $\{p(x; \theta) \log p(x; t)\}_{x \in I}$ ammette somma finita. Inoltre, per ogni $\theta \neq t$,*

$$\sum_{x \in I} p(x; \theta) \log p(x; t) < \sum_{x \in I} p(x; \theta) \log p(x; \theta).$$

Dimostrazione. Per l'ipotesi (E), la variabile aleatoria $\log p(X_1, t)$ ammette momento secondo finito rispetto a P_θ , quindi ammette anche valor medio finito (si ricordi che $L^2 \subseteq L^1$). Per la Proposizione 3.52, ciò significa che $\{p(x; \theta) \log p(x; t)\}_{x \in I}$ ammette somma finita. Notiamo ora che

$$\sum_{x \in I} p(x; \theta) \log p(x; \theta) - \sum_{x \in I} p(x; \theta) \log p(x; t) = \sum_{x \in I} p(x; t) \frac{p(x; \theta)}{p(x; t)} \log \frac{p(x; \theta)}{p(x; t)}.$$

Ora usiamo il seguente fatto elementare, che deriva dalla stretta convessità della funzione $h(z) = z \log z - z + 1$: per ogni $z > 0$

$$z \log z - z + 1 \geq 0,$$

e $z \log z - z + 1 = 0$ solo per $z = 1$. Pertanto, per ogni $x \in I$

$$\frac{p(x; \theta)}{p(x; t)} \log \frac{p(x; \theta)}{p(x; t)} - \frac{p(x; \theta)}{p(x; t)} + 1 \geq 0,$$

dove la disuguaglianza è stretta ogni qual volta $p(x; \theta) \neq p(x; t)$. Grazie all'ipotesi (D), tale disuguaglianza stretta vale per almeno un $x \in I$. Perciò

$$\sum_{x \in I} p(x; t) \left[\frac{p(x; \theta)}{p(x; t)} \log \frac{p(x; \theta)}{p(x; t)} - \frac{p(x; \theta)}{p(x; t)} + 1 \right] > 0,$$

da cui la tesi segue facilmente. \square

Teorema 8.29 (Consistenza dello stimatore MV). *Se il modello statistico $\{(\Omega, \mathcal{A}, P_\theta)\}_{\theta \in \Theta}$ e il relativo campione $(X_n)_{n \geq 1}$ soddisfano le ipotesi (A)–(E), lo stimatore MV $(Y_n)_{n \geq 1}$, definito in (8.18), è uno stimatore consistente per θ .*

Dimostrazione. Siano $\theta, t \in \Theta$. Le variabili aleatorie $(\log p(X_n; t))_{n \geq 1}$ sono i.i.d. rispetto a P_θ , grazie alle Proposizioni 3.15 e 3.40, e, per l'ipotesi (E), esse ammettono momento secondo finito. Poniamo

$$l(\theta, t) := E_\theta [\log p(X_n; t)] = \sum_{x \in I} p(x; \theta) \log p(x; t).$$

Ricordando (8.16), per la legge debole dei grandi numeri (Corollario 7.6), ponendo $\underline{X} := (X_1, \dots, X_n)$, si ha

$$\lim_{n \rightarrow +\infty} P_\theta (|L_n(\underline{X}, t) - l(\theta, t)| > \varepsilon) = 0, \quad \forall \varepsilon > 0. \quad (8.19)$$

A questo punto possiamo illustrare l'idea della dimostrazione. Per il Lemma 8.28, $l(\theta, \theta) > l(\theta, t)$ per ogni $t \neq \theta$. Poiché, per (8.19), $L_n(\underline{X}, t) \simeq l(\theta, t)$ per n grande e ogni $\theta, t \in \Theta$, l'unico massimo locale proprio della funzione $t \mapsto L_n(\underline{X}, t)$ deve essere “vicino” a θ , che è il massimo assoluto della funzione $t \mapsto l(\theta, t)$. Poiché tale massimo locale è Y_n , si ottiene $Y_n \simeq \theta$ per n grande, che è quanto si vuole dimostrare. Per rendere rigoroso l'argomento appena illustrato, notiamo anzitutto che, per $\varepsilon > 0$ fissato ma arbitrario,

$$\begin{aligned} P_\theta (& |(L_n(\underline{X}, t) - L_n(\underline{X}, \theta)) - (l(\theta, t) - l(\theta, \theta))| > \varepsilon) \\ & \leq P_\theta (\{|L_n(\underline{X}, t) - l(\theta, t)| > \frac{\varepsilon}{2}\} \cup \{|L_n(\underline{X}, \theta) - l(\theta, \theta)| > \frac{\varepsilon}{2}\}) \\ & \leq P_\theta (|L_n(\underline{X}, t) - l(\theta, t)| > \frac{\varepsilon}{2}) + P_\theta (|L_n(\underline{X}, \theta) - l(\theta, \theta)| > \frac{\varepsilon}{2}), \end{aligned}$$

perciò, per (8.19),

$$\lim_{n \rightarrow +\infty} P_\theta (|(L_n(\underline{X}, t) - L_n(\underline{X}, \theta)) - (l(\theta, t) - l(\theta, \theta))| > \varepsilon) = 0. \quad (8.20)$$

Per il Lemma 8.28 si ha $l(\theta, t) - l(\theta, \theta) < 0$ $t \neq \theta$. Pertanto, usando (8.20) con $\varepsilon < l(\theta, \theta) - l(\theta, t)$, abbiamo che, per $t \neq \theta$,

$$\lim_{n \rightarrow +\infty} P_\theta (L_n(\underline{X}, t) - L_n(\underline{X}, \theta) < 0) = 1. \quad (8.21)$$

Consideriamo allora, per $\delta > 0$ fissato ma arbitrario, l'evento

$$A_n := \{L_n(\underline{X}, \theta - \delta) - L_n(\underline{X}, \theta) < 0\} \cap \{L_n(\underline{X}, \theta + \delta) - L_n(\underline{X}, \theta) < 0\}.$$

Per (8.21)

$$\lim_{n \rightarrow +\infty} P_\theta (A_n) = 1. \quad (8.22)$$

Per definizione, se $\omega \in A_n$, si ha che $L_n(\underline{X}(\omega), \theta)$ è strettamente maggiore sia di $L_n(\underline{X}(\omega), \theta - \delta)$, sia di $L_n(\underline{X}(\omega), \theta + \delta)$. Dato che la funzione $t \mapsto L_n(\underline{X}(\omega), t)$ è continua in t , per l'ipotesi (B), segue che essa ha un massimo locale nell'intervallo $(\theta - \delta, \theta + \delta)$, per ogni $\omega \in A_n$, che è necessariamente $\hat{\theta}(\underline{X}(\omega)) = Y_n(\omega)$, per l'ipotesi (C) (si ricordi (8.18)). Abbiamo dimostrato l'inclusione di eventi

$$A_n \subseteq \{\hat{\theta}_n(X_1, X_2, \dots, X_n) \in (\theta - \delta, \theta + \delta)\} = \{|Y_n - \theta| < \delta\},$$

Quindi, per (8.22)

$$\lim_{n \rightarrow +\infty} P_\theta(|Y_n - \theta| < \delta) = 1, \quad \forall \delta > 0,$$

ossia

$$\lim_{n \rightarrow +\infty} P_\theta(|Y_n - \theta| > \delta) = 0, \quad \forall \delta > 0,$$

che è proprio la consistenza. \square

Ci occupiamo ora di fornire condizioni sufficienti affinché lo stimatore MV sia asintoticamente normale. Faremo uso delle seguenti ulteriori ipotesi.

(F) La funzione $p(x; \theta)$ (risp. $f(x; \theta)$) soddisfa le seguenti condizioni:

- (i) Per ogni $x \in I$, la funzione $\theta \mapsto \log p(x; \theta)$ (risp. $\theta \mapsto \log f(x; \theta)$) è di classe \mathcal{C}^2 , ossia due volte derivabile con continuità.
- (ii) La funzione

$$B(x, \theta) := \frac{d^2 \log p(x; \theta)}{d\theta^2}. \quad (8.23)$$

è continua in θ uniformemente per $x \in I$, ossia per ogni $\theta \in \Theta$ e $\varepsilon > 0$ esiste $\delta_\varepsilon > 0$ tale che

$$|t - \theta| < \delta_\varepsilon \implies |B(x, t) - B(x, \theta)| < \varepsilon \text{ per ogni } x \in I.$$

(Nel caso assolutamente continuo la condizione è analoga, con $f(x; \theta)$ al posto di $p(x; \theta)$.)

- (iii) È lecita la seguente commutazione di somma e derivata: per $k = 1, 2$

$$\sum_{x \in I} \frac{d^k}{d\theta^k} p(x; \theta) = \frac{d^k}{d\theta^k} \sum_{x \in I} p(x; \theta) = \frac{d^k}{d\theta^k} 1 = 0.$$

(Per il caso assolutamente continuo, la condizione è analoga, con $f(x; \theta)$ al posto di $p(x; \theta)$ e l'integrale $\int_I(\dots) dx$ al posto di $\sum_{x \in I}(\dots)$.)

- (G) Per ogni $\theta \in \Theta$, la variabile aleatoria $B(X_1, \theta)$ ammette momento secondo rispetto a P_θ .

Osservazione 8.30. La condizione (iii) dell'ipotesi (F) è un'ipotesi abbastanza debole. Si può dimostrare che una condizione sufficiente è la seguente: per ogni $\theta \in \Theta$

esiste $\varepsilon > 0$ tale che, per $k = 1, 2$,

$$\sum_{x \in I} \sup_{t \in (\theta - \varepsilon, \theta + \varepsilon)} \left| \frac{d^k}{d\theta^k} p(x; t) \right| < +\infty \quad \square$$

Osservazione 8.31. La condizione (ii) dell'ipotesi (F) è invece assai restrittiva e, come vedremo, non è soddisfatta in molti casi interessanti. Essa può essere indebolita in modo considerevole, al prezzo di rendere più complicata la dimostrazione della normalità asintotica. Un caso particolare, ma interessante, in cui la dimostrazione che daremo può essere modificata in modo abbastanza indolore, è quello in cui $B(x, \theta)$ è della forma

$$B(x, \theta) = \varphi(x) + \psi(x)g(\theta), \quad (8.24)$$

dove g è una funzione continua e $E_\theta[|\psi(X_1)|] < +\infty$ per ogni $\theta \in \Theta$. \square

Teorema 8.32 (Normalità asintotica dello stimatore MV). *Supponiamo che il modello statistico $\{(\Omega, \mathcal{A}, P_\theta)\}_{\theta \in \Theta}$ e il relativo campione $(X_n)_{n \geq 1}$ soddisfino le ipotesi (A)–(G). Allora lo stimatore MV $(Y_n)_{n \geq 1}$, definito in (8.18), è uno stimatore asintoticamente normale per θ , con $\sigma(\theta)$ dato da*

$$\frac{1}{\sigma^2(\theta)} := -E_\theta[B(X_1, \theta)] = \sum_{x \in I} p(x; \theta) \left[\frac{d}{d\theta} \log p(x; \theta) \right]^2. \quad (8.25)$$

Osservazione 8.33. Con un po' di abuso di linguaggio, l'enunciato del Teorema 8.32 esprime il fatto che la distribuzione di Y_n è “vicina” a quella di una $N(\theta, \sigma^2(\theta)/n)$, quando n è abbastanza grande. Si noti che la varianza di quest'ultima Normale coincide con il limite inferiore di Cramér-Rao, che abbiamo visto nel Teorema 8.8; per questa ragione si dice che gli stimatori di massima verosimiglianza sono *asintoticamente efficienti*. Osserviamo che la convergenza della funzione di ripartizione, presente nella definizione di normalità asintotica, *non* garantisce che

$$\lim_{n \rightarrow +\infty} \frac{\text{Var}(Y_n)}{\sigma^2(t)/n} = 1.$$

Quest'ultima identità è comunque vera in molti casi interessanti, e può essere dimostrata con qualche ulteriore ipotesi. La questione non verrà ulteriormente approfondita in questo testo. \square

La dimostrazione del Teorema 8.32 richiederà alcuni lemmi tecnici. L'idea di fondo può tuttavia essere illustrata in poche righe, almeno a livello euristico. Le ipotesi di regolarità che abbiamo assunto permettono il seguente sviluppo di Taylor al primo ordine per la funzione $\frac{d}{d\theta} L_n$:

$$\frac{d}{d\theta} L_n(\underline{X}, Y_n) = \frac{d}{d\theta} L_n(\underline{X}, \theta) + \frac{d^2}{d\theta^2} L_n(\underline{X}, \theta)(Y_n - \theta) + o(Y_n - \theta). \quad (8.26)$$

Tale espansione è giustificata dal fatto che, per il Teorema 8.29, la variabile aleatoria $(Y_n - \theta)$ è “piccola per n grande”. Quindi, a meno di “termini di ordine inferiore” e osservando che $\frac{d}{d\theta} L_n(\underline{X}, Y_n) = 0$, per definizione di Y_n , possiamo dedurre da (8.26)

$$\sqrt{n}(Y_n - \theta) \simeq -\frac{\sqrt{n} \frac{d}{d\theta} L_n(\underline{X}, \theta)}{\frac{d^2}{d\theta^2} L_n(\underline{X}, \theta)} = -\frac{\sqrt{n} \frac{1}{n} \sum_{i=1}^n \frac{d}{d\theta} \log p(X_i, \theta)}{\frac{1}{n} \sum_{i=1}^n \frac{d^2}{d\theta^2} \log p(X_i, \theta)}. \quad (8.27)$$

In quest’ultima frazione, il numeratore è la media campionaria, moltiplicata per \sqrt{n} , delle variabili aleatorie i.i.d. $\log p(X_i, \theta)$ che, come vedremo, hanno media zero e varianza finita; quindi al numeratore si applica il teorema limite centrale. Il denominatore è anch’esso una media campionaria, delle variabili aleatorie i.i.d. $\frac{d^2}{d\theta^2} \log p(X_i, \theta)$ che vedremo avere media strettamente negativa e, per l’ipotesi (G), varianza finita. Si applica perciò la legge debole dei grandi numeri. Grazie al prossimo Lemma 8.34, tali informazioni si combinano per ottenere il limite della funzione di ripartizione di $\sqrt{n}(Y_n - \theta)$, da cui la tesi del Teorema 8.32 segue agevolmente.

Come annunciato, sono necessari alcuni lemmi preliminari.

Lemma 8.34. *In uno spazio di probabilità (Ω, \mathcal{A}, P) , si considerino due successioni $(Z_n)_{n \geq 1}$ e $(X_n)_{n \geq 1}$ di variabili aleatorie a valori reali, tali che*

$$Z_n \xrightarrow{D} N(0, 1) \quad e \quad \lim_{n \rightarrow +\infty} P(|X_n - 1| > \varepsilon) = 0, \quad \forall \varepsilon > 0.$$

Allora

$$X_n Z_n \xrightarrow{D} N(0, 1).$$

Dimostrazione. Consideriamo le seguenti inclusioni di eventi, valide per $x \geq 0$:

$$\begin{aligned} \{X_n Z_n \leq x\} &\subseteq \{X_n Z_n \leq x, |X_n - 1| \leq \varepsilon\} \cup \{|X_n - 1| > \varepsilon\} \\ &\subseteq \{Z_n \leq \frac{x}{1-\varepsilon}\} \cup \{|X_n - 1| > \varepsilon\}. \end{aligned} \quad (8.28)$$

Quindi

$$P(X_n Z_n \leq x) \leq P(Z_n \leq \frac{x}{1-\varepsilon}) + P(|X_n - 1| > \varepsilon),$$

da cui segue

$$\limsup_{n \rightarrow +\infty} P(X_n Z_n \leq x) \leq \Phi\left(\frac{x}{1-\varepsilon}\right).$$

Facendo tendere ε a zero, e usando il fatto che Φ è continua, si ottiene

$$\limsup_{n \rightarrow +\infty} P(X_n Z_n \leq x) \leq \Phi(x), \quad (8.29)$$

per ogni $x \geq 0$. Il caso $x < 0$ si tratta in modo analogo, rimpiazzando $1 - \varepsilon$ con $1 + \varepsilon$ in (8.28) e nelle successive disuguaglianze.

In modo analogo, per $x \geq 0$ si ha

$$\begin{aligned} \{X_n Z_n \leq x\} &\supseteq \{X_n Z_n \leq x, |X_n - 1| \leq \varepsilon\} \supseteq \{Z_n \leq \frac{x}{1+\varepsilon}, |X_n - 1| \leq \varepsilon\} \\ &= \{Z_n \leq \frac{x}{1+\varepsilon}\} \setminus \{Z_n \leq \frac{x}{1+\varepsilon}, |X_n - 1| > \varepsilon\}, \end{aligned} \quad (8.30)$$

pertanto

$$\begin{aligned} P(X_n Z_n \leq x) &\geq P(Z_n \leq \frac{x}{1+\varepsilon}) - P(Z_n \leq \frac{x}{1+\varepsilon}, |X_n - 1| > \varepsilon) \\ &\geq P(Z_n \leq \frac{x}{1+\varepsilon}) - P(|X_n - 1| > \varepsilon), \end{aligned}$$

da cui segue che

$$\liminf_{n \rightarrow +\infty} P(X_n Z_n \leq x) \geq \Phi\left(\frac{x}{1+\varepsilon}\right).$$

Facendo tendere ε a zero, ciò implica che

$$\liminf_{n \rightarrow +\infty} P(X_n Z_n \leq x) \geq \Phi(x), \quad (8.31)$$

per ogni $x \geq 0$. Il caso $x < 0$ è del tutto analogo: basta rimpiazzare $1 + \varepsilon$ con $1 - \varepsilon$ in (8.30) e nelle relazioni successive.

Mettendo assieme (8.29) e (8.31), otteniamo

$$\lim_{n \rightarrow +\infty} P(X_n Z_n \leq x) = \Phi(x), \quad \forall x \in \mathbb{R},$$

che è quanto volevamo dimostrare. \square

Lemma 8.35. *Supponiamo che valgano le ipotesi (A)-(G). Allora*

$$E_\theta[B(X_1, \theta)] = - \sum_{x \in I} p(x; \theta) \left[\frac{d}{d\theta} \log p(x; \theta) \right]^2 =: - \frac{1}{\sigma^2(\theta)}. \quad (8.32)$$

Dimostrazione. Ricordando la definizione (8.23) di $B(x, \theta)$ e la formula (3.30), possiamo scrivere

$$E_\theta[B(X_1, \theta)] = \sum_{x \in I} p(x; \theta) \frac{d^2 \log p(x; \theta)}{d\theta^2}. \quad (8.33)$$

Si noti che

$$\begin{aligned} \frac{d^2 \log p(x; \theta)}{d\theta^2} &= \frac{d}{d\theta} \left[\frac{\frac{d}{d\theta} p(x; \theta)}{p(x; \theta)} \right] = \frac{1}{p(x; \theta)} \frac{d^2}{d\theta^2} p(x; \theta) - \frac{\left[\frac{d}{d\theta} p(x; \theta) \right]^2}{p^2(x; \theta)} \\ &= \frac{1}{p(x; \theta)} \frac{d^2}{d\theta^2} p(x; \theta) - \left[\frac{d}{d\theta} \log p(x; \theta) \right]^2. \end{aligned}$$

Pertanto

$$\begin{aligned} \sum_{x \in I} p(x; \theta) \frac{d^2 \log p(x; \theta)}{d\theta^2} &= \sum_{x \in I} \frac{d^2}{d\theta^2} p(x; \theta) - \sum_{x \in I} p(x; \theta) \left[\frac{d}{d\theta} \log p(x; \theta) \right]^2 \\ &= - \sum_{x \in I} p(x; \theta) \left[\frac{d}{d\theta} \log p(x; \theta) \right]^2, \end{aligned}$$

in quanto, per l'ipotesi (F)-(iii), si ha

$$\sum_{x \in I} \frac{d^2}{d\theta^2} p(x; \theta) = 0.$$

Con ciò la dimostrazione è completa. \square

Dimostrazione (del Teorema 8.32). Per ogni $\underline{x} = (x_1, x_2, \dots, x_n) \in I^n$ e $\theta \in \Theta$ poniamo

$$H_n(\underline{x}, \theta) := \frac{d}{d\theta} L_n(\underline{x}, \theta) = \frac{1}{n} \sum_{i=1}^n \frac{d}{d\theta} \log p(x_i; \theta).$$

Inoltre definiamo (si ricordi (8.23))

$$B_n(\underline{x}, t, \theta) := \begin{cases} \frac{H_n(\underline{x}, t) - H_n(\underline{x}, \theta)}{t - \theta} & \text{per } t \neq \theta \\ \frac{d}{d\theta} H_n(\underline{x}, \theta) = \frac{1}{n} \sum_{i=1}^n \frac{d^2}{d\theta^2} \log p(x_i; \theta) = \frac{1}{n} \sum_{i=1}^n B(x_i, \theta) & \text{per } t = \theta. \end{cases} \quad (8.34)$$

Tanto H_n che B_n si possono definire per ogni $\underline{x} \in \mathbb{R}^n$, ponendole identicamente uguali a zero se $\underline{x} \notin I^n$. Per il teorema della media o di Lagrange, per ogni $t, \theta \in \Theta$ e ogni $\underline{x} \in \mathbb{R}^n$, esiste θ^* con $|\theta^* - \theta| \leq |t - \theta|$

$$B_n(\underline{x}, t, \theta) = \frac{d}{d\theta} H_n(\underline{x}, \theta^*) = \frac{1}{n} \sum_{i=1}^n \frac{d^2}{d\theta^2} \log p(x_i; \theta^*). \quad (8.35)$$

Fissiamo $\varepsilon > 0$. Per l'ipotesi (F)-(ii), esiste $\delta_\varepsilon > 0$ tale che se $|t - \theta| < \delta_\varepsilon$ si ha

$$\left| \frac{d^2}{d\theta^2} \log p(x; t) - \frac{d^2}{d\theta^2} \log p(x; \theta) \right| < \varepsilon, \quad \forall x \in I.$$

Pertanto, per ogni $\underline{x} \in \mathbb{R}^n$, se $|t - \theta| < \delta_\varepsilon$ si ha

$$|B_n(\underline{x}, t, \theta) - B_n(\underline{x}, \theta, \theta)| \leq \frac{1}{n} \sum_{i=1}^n \left| \frac{d^2}{d\theta^2} \log p(x_i; \theta^*) - \frac{d^2}{d\theta^2} \log p(x_i; \theta) \right| \leq \varepsilon.$$

Ricordando che $Y_n := \hat{\theta}_n(X_1, X_2, \dots, X_n)$, cf. (8.18), definiamo l'evento

$$A_n := \{|Y_n - \theta| \leq \delta_\varepsilon\}. \quad (8.36)$$

Per il Teorema 8.29 sappiamo che

$$\lim_{n \rightarrow +\infty} P_\theta(A_n) = 1. \quad (8.37)$$

Inoltre, per quanto appena visto, se $\omega \in A_n$

$$|B_n(\underline{X}(\omega), Y_n(\omega), \theta) - B_n(\underline{X}(\omega), \theta, \theta)| \leq \varepsilon,$$

il che significa che

$$\{|B_n(\underline{X}, Y_n, \theta) - B_n(\underline{X}, \theta, \theta)| \leq \varepsilon\} \supseteq A_n.$$

In definitiva, per (8.37), abbiamo mostrato che per ogni $\varepsilon > 0$

$$\lim_{n \rightarrow +\infty} P_\theta(|B_n(\underline{X}, Y_n, \theta) - B_n(\underline{X}, \theta, \theta)| \leq \varepsilon) = 1. \quad (8.38)$$

Si noti ora che

$$B_n(\underline{X}, \theta, \theta) = \frac{1}{n} \sum_{i=1}^n \frac{d^2}{d\theta^2} \log p(X_i; \theta).$$

Le variabili aleatorie $\frac{d^2}{d\theta^2} \log p(X_i; \theta)$ rispetto a P_θ sono i.i.d., ammettono momento secondo finito per l'ipotesi (G) e hanno media $-\frac{1}{\sigma^2(\theta)}$ per il Lemma 8.35. Perciò, per la legge debole di grandi numeri (Corollario 7.6), per ogni $\varepsilon > 0$

$$\lim_{n \rightarrow +\infty} P_\theta \left(\left| B_n(\underline{X}, \theta, \theta) + \frac{1}{\sigma^2(\theta)} \right| \leq \varepsilon \right) = 1. \quad (8.39)$$

Mettendo assieme (8.38) e (8.39), abbiamo che, per ogni $\varepsilon > 0$

$$\lim_{n \rightarrow +\infty} P_\theta \left(\left| B_n(\underline{X}, Y_n, \theta) + \frac{1}{\sigma^2(\theta)} \right| \leq \varepsilon \right) = 1, \quad (8.40)$$

che è equivalente (si veda l'Esercizio 8.6) a

$$\lim_{n \rightarrow +\infty} P_\theta \left(\left| \left(-\frac{1}{B_n(\underline{X}, Y_n, \theta) \sigma^2(\theta)} \right) - 1 \right| > \varepsilon \right) = 0. \quad (8.41)$$

Notiamo ora che, per $t, \theta \in \Theta$,

$$H_n(\underline{X}, t) = H_n(\underline{X}, \theta) + B_n(\underline{X}, t, \theta)(t - \theta).$$

Poiché Y_n è per costruzione un punto di massimo locale per $L_n(\underline{X}, \theta)$, si ha che $H_n(\underline{X}, Y_n) = 0$. Perciò, applicando la formula precedente con $t = Y_n$, otteniamo

$$H_n(\underline{X}, \theta) = -B_n(\underline{X}, Y_n, \theta)(Y_n - \theta). \quad (8.42)$$

Ricordiamo che

$$H_n(\underline{X}, \theta) = \frac{1}{n} \sum_{i=1}^n \frac{d}{d\theta} \log p(X_i; \theta),$$

e osserviamo che le variabili aleatorie $\frac{d}{d\theta} \log p(X_i; \theta)$ sono i.i.d., hanno valor medio

$$E_{\theta} \left[\frac{d}{d\theta} \log p(X_i; \theta) \right] = \sum_{x \in I} p(x; \theta) \frac{d}{d\theta} \log p(x; \theta) = \sum_{x \in I} \frac{d}{d\theta} p(x; \theta) = 0,$$

grazie all'ipotesi (F)-(iii), e varianza

$$E_{\theta} \left[\left(\frac{d}{d\theta} \log p(X_i; \theta) \right)^2 \right] = \sum_{x \in I} p(x; \theta) \left(\frac{d}{d\theta} \log p(x; \theta) \right)^2 = \frac{1}{\sigma^2(\theta)},$$

avendo usato la definizione (8.32) di $\sigma^2(\theta)$. Grazie al teorema limite centrale (Teorema 7.11) abbiamo dunque che

$$\sigma(\theta) \sqrt{n} H_n(\underline{X}, \theta) \xrightarrow{D} N(0, 1).$$

Applicando infine il Lemma 8.34 alle variabili aleatorie $Z_n = \sigma(\theta) \sqrt{n} H_n(\underline{X}, \theta)$ e $X_n = -1/(B_n(\underline{X}, Y_n, \theta) \sigma^2(\theta))$, grazie a (8.41) e (8.42) abbiamo

$$-\frac{1}{B_n(\underline{X}, Y_n, \theta) \sigma^2(\theta)} \sigma(\theta) \sqrt{n} H_n(\underline{X}, \theta) = \frac{\sqrt{n}}{\sigma(\theta)} (Y_n - \theta) \xrightarrow{D} N(0, 1),$$

che è quanto si voleva dimostrare. \square

Osservazione 8.36. Illustriamo brevemente la modifica necessaria per adattare questa dimostrazione al caso in cui non valga l'ipotesi (F)-(ii), ma la funzione $B(x, \theta)$ sia della forma speciale (8.24) illustrata nell'Osservazione 8.31, ossia

$$B(x, \theta) = \varphi(x) + \psi(x)g(\theta), \quad (8.43)$$

dove g è una funzione continua e $E_{\theta}[|\psi(X_1)|] < +\infty$ per ogni $\theta \in \Theta$. Osserviamo che l'ipotesi (F)-(ii) è stata usata solo per dimostrare la relazione (8.38), cioè

$$\lim_{n \rightarrow +\infty} P_{\theta}(|B_n(\underline{X}, Y_n, \theta) - B_n(\underline{X}, \theta, \theta)| \leq \varepsilon) = 1, \quad \forall \varepsilon > 0. \quad (8.44)$$

Per la definizione (8.34) e (8.35),

$$B_n(\underline{X}, \theta, \theta) = \frac{1}{n} \sum_{i=1}^n B(X_i, \theta),$$

e

$$|B_n(\underline{X}, Y_n, \theta) - B_n(\underline{X}, \theta, \theta)| = \frac{1}{n} \sum_{i=1}^n |B(X_i, \theta_n^*) - B(X_i, \theta)|,$$

dove $|\theta_n^* - \theta| \leq |Y_n - \theta|$. Pertanto, se vale la relazione (8.43),

$$|B_n(\underline{X}, Y_n, \theta) - B_n(\underline{X}, \theta, \theta)| = \left| \frac{1}{n} \sum_{i=1}^n \psi(X_i) \right| |g(\theta_n^*) - g(\theta)|. \quad (8.45)$$

Dato che g è continua, dato $\varepsilon > 0$ esiste $\delta_\varepsilon > 0$ tale che se $|t - \theta| \leq \delta_\varepsilon$, allora $|g(t) - g(\theta)| \leq \varepsilon$. Pertanto, posto $A_n = \{|Y_n - \theta| \leq \delta_\varepsilon\}$, ossia l'evento definito in (8.36), allora per $\omega \in A_n$

$$|g(\theta_n^*(\omega)) - g(\theta)| \leq \varepsilon.$$

Ricordiamo che $P_\theta(A_n) \rightarrow 1$, per il Teorema 8.29. Definiamo ora l'evento

$$C_n := \left\{ \left| \frac{1}{n} \sum_{i=1}^n \psi(X_i) - m \right| \leq 1 \right\},$$

dove $m = E_\theta(\psi(X_1))$. Le variabili aleatorie $\psi(X_i)$ sono i.i.d. con $E_\theta(|X_1|) < \infty$, quindi per la legge debole dei grandi numeri (Teorema 7.3) si ha $P_\theta(C_n) \rightarrow 1$. Se $\omega \in C_n$,

$$\left| \frac{1}{n} \sum_{i=1}^n \psi(X_i(\omega)) \right| \leq m + 1,$$

quindi, se $\omega \in A_n \cap C_n$

$$|B_n(\underline{X}, Y_n, \theta) - B_n(\underline{X}, \theta, \theta)| \leq (m + 1)\varepsilon.$$

Dato che $P_\theta(A_n) \rightarrow 1$ e $P_\theta(C_n) \rightarrow 1$, segue che $P_\theta(A_n \cap C_n) \rightarrow 1$. Per l'arbitrarietà di ε , la relazione (8.44) si deduce facilmente. \square

Vediamo ora alcuni esempi di stimatori MV.

Esempio 8.37. Cominciamo col considerare il modello statistico dell'Esempio 8.5:

$$p(x; \theta) = e^{-\theta} \frac{\theta^x}{x!} \quad \text{per } x \in \mathbb{N}_0 = \{0, 1, 2, \dots\},$$

con $\Theta = (0, +\infty)$. L'ipotesi (B) è verificata con $I = \mathbb{N}_0$. Notiamo che

$$\begin{aligned} L_n(\underline{x}, \theta) &= \frac{1}{n} \sum_{i=1}^n \log p(x_i; \theta) = \frac{1}{n} \sum_{i=1}^n [-\theta + x_i \log \theta - \log x_i!] \\ &= -\theta + \bar{x}_n \log \theta + \frac{1}{n} \sum_{i=1}^n \log x_i!, \end{aligned}$$

dove $\bar{x}_n = \frac{1}{n} \sum_{i=1}^n x_i$. Derivando rispetto a θ , si vede facilmente che l'unico punto in cui la derivata si annulla, che è un punto di massimo per $L_n(\underline{x}, \theta)$, è

$$\hat{\theta}_n(\underline{x}) = \bar{x}_n,$$

dunque lo stimatore MV $Y_n = \hat{\theta}_n(\underline{X}) = \bar{X}_n$ non è altro che la media campionaria.

Non è difficile mostrare che le ipotesi (C)–(G) sono verificate, eccetto la (F)–(ii). Infatti si trova

$$B(x, \theta) = -\frac{x}{\theta^2}.$$

Valgono però le ipotesi modificate illustrate nell'Osservazione 8.31, cf. (8.24). In questo caso, tuttavia, la consistenza e la normalità asintotica di $Y_n = \bar{X}_n$ seguono direttamente dalla legge dei grandi numeri e dal teorema limite centrale. \square

Esempio 8.38. Sia $T = (0, 1)$, e $X_1 \sim \text{Geo}(\theta)$, cioè, per $x \in I := \mathbb{N} = \{1, 2, \dots\}$

$$p(x; \theta) = \theta(1 - \theta)^{x-1},$$

con $\Theta = (0, 1)$. Otteniamo

$$L_n(\underline{x}, \theta) = \log \theta + (\bar{x}_n - 1) \log(1 - \theta),$$

perciò

$$\frac{d}{d\theta} L_n(\underline{x}, \theta) = \frac{1}{\theta} - (\bar{x}_n - 1) \frac{1}{1 - \theta}.$$

Ponendo $\frac{d}{d\theta} L_n(\underline{x}, \theta) = 0$, si trova

$$\hat{\theta}_n(\underline{x}) = \frac{1}{\bar{x}_n},$$

che, come è facile vedere, è effettivamente l'unico massimo locale.

Si noti che lo stimatore $Y_n = \hat{\theta}_n(\underline{X})$ non soddisfa pienamente l'Ipotesi (C), in quanto, se $\bar{x}_n = 1$, si ha $\hat{\theta}_n(\underline{x}) = 1 \notin \Theta$. Tuttavia, per ogni $\theta \in \Theta$

$$\lim_{n \rightarrow +\infty} P_\theta(\bar{X}_n = 1) = \lim_{n \rightarrow +\infty} P_\theta(X_1 = 1, X_2 = 1, \dots, X_n = 1) = \lim_{n \rightarrow +\infty} \theta^n = 0.$$

Questo permette di apportare opportune modifiche alle dimostrazioni dei Teoremi 8.29 e 8.32, in modo da adattarli a questo caso. Si vede infatti che le ipotesi (C)–(G), a parte questo aspetto, e l'ipotesi (F)–(ii), che va rimpiazzata con quella nell'Osservazione 8.31, sono verificate. Segue che $Y_n = 1/\bar{X}_n$ è uno stimatore consistente e asintoticamente normale per θ .

Calcoliamo ora

$$\frac{d^2}{d\theta^2} \log p(x; \theta) = -\frac{1}{\theta^2} - \frac{x-1}{(1-\theta)^2}.$$

Perciò, usando il Lemma 8.35,

$$\begin{aligned} \frac{1}{\sigma^2(\theta)} &= -E_\theta \left[\frac{d^2}{d\theta^2} \log p(X_1; \theta) \right] = \frac{1}{\theta^2} + \frac{1}{(1-\theta)^2} E_\theta(X_1 - 1) \\ &= \frac{1}{\theta^2} + \frac{1}{(1-\theta)^2} \frac{1-\theta}{\theta} = \frac{1}{(1-\theta)\theta^2}. \end{aligned}$$

Perciò

$$\sigma(\theta) = \theta \sqrt{1 - \theta}.$$

È facile verificare che $\sigma(\theta)$ assume il suo massimo in $(0, 1)$ per $\theta = \frac{2}{3}$, quindi

$$\bar{\sigma} = \sup_{\theta \in (0,1)} \sigma(\theta) = \frac{2}{3\sqrt{3}}.$$

Quest'ultima identità consente di determinare un intervallo di confidenza asintotico per θ , come indicato nel paragrafo precedente. \square

Esempio 8.39. Supponiamo che $X_1 \sim \Gamma(\alpha, \theta)$, dove $\alpha > 0$ si assume noto, e $\theta \in \Theta = (0, +\infty)$. In altre parole

$$f(x; \theta) = \frac{\theta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\theta x},$$

per $x \in I = (0, +\infty)$ (che rende soddisfatta l'ipotesi (B)). Si trova

$$\begin{aligned} \log f(x; \theta) &= \alpha \log \theta - \log \Gamma(\alpha) + (\alpha - 1) \log x - \theta x, \\ \frac{d}{d\theta} \log f(x; \theta) &= \frac{\alpha}{\theta} - x, \quad \frac{d^2}{d\theta^2} \log f(x; \theta) = -\frac{\alpha}{\theta^2}, \\ L_n(\underline{x}; \theta) &= \alpha \log \theta - \log \Gamma(\alpha) + (\alpha - 1) \frac{1}{n} \sum_{i=1}^n \log x_i - \theta \frac{1}{n} \sum_{i=1}^n x_i, \end{aligned}$$

da cui si può verificare agevolmente che le ipotesi (B)-(G) sono verificate, e

$$Y_n = \hat{\theta}_n(\bar{X}) = \frac{\alpha}{\frac{1}{n} \sum_{i=1}^n X_i} = \frac{\alpha}{\bar{X}_n},$$

che risulta dunque consistente e asintoticamente normale.

Calcoliamo dunque

$$\frac{1}{\sigma^2(\theta)} = -E_\theta \left[\frac{d^2}{d\theta^2} \log f(X_1; \theta) \right] = \frac{\alpha}{\theta^2},$$

da cui si ottiene $\sigma(\theta) = \frac{\theta}{\sqrt{\alpha}}$. \square

Osservazione 8.40. Negli esempi precedenti lo stimatore MV risulta essere una funzione “regolare” di una media campionaria. Questo permetterebbe di ottenere la consistenza in modo abbastanza elementare, ed anche la dimostrazione della normalità asintotica (Teorema 8.32) potrebbe essere semplificata. Ci sono situazioni, tuttavia, in cui le condizioni del Teorema 8.32 sono verificate, ma non è possibile esprimere esplicitamente lo stimatore MV come funzione di una media campionaria. Si consideri, ad esempio, un modello statistico per cui

$$f(x; \theta) = \frac{\sqrt{\theta}}{\pi} \frac{1}{1 + \theta x^2},$$

dove $\theta \in \Theta = (0, +\infty)$. Non è difficile mostrare che questo modello soddisfa le ipotesi (A)-(G), che abbiamo assunto nel Teorema 8.32 (vedi Esercizio 8.5). In particolare, la funzione

$$L_n(\underline{x}; \theta) = -\log \pi + \frac{1}{2} \log \theta - \frac{1}{n} \sum_{i=1}^n \log(1 + \theta x_i^2)$$

ammette un unico massimo locale, non esprimibile, tuttavia, esplicitamente come funzione di \underline{x} . Ciononostante, la normalità asintotica di tale stimatore è garantita dal Teorema 8.32, ed è possibile calcolare esplicitamente la sua varianza asintotica. \square

Esercizi

Esercizio 8.4. Si consideri il seguente modello statistico:

$$f(x; \theta) = (\theta + 1)x^\theta \mathbb{1}_{(0,1)}(x),$$

dove $\theta \in \Theta = (-1, +\infty)$. Si determini lo stimatore di massima verosimiglianza, si mostri che è asintoticamente normale e si calcoli la varianza asintotica $\sigma^2(t)$.

Esercizio 8.5. Si mostri che il modello statistico

$$f(x; \theta) = \frac{\sqrt{\theta}}{\pi} \frac{1}{1 + \theta x^2},$$

dove $\theta \in \Theta = (0, +\infty)$, soddisfa alle ipotesi (A)–(G), e calcolare la varianza asintotica $\sigma(\theta)$ del corrispondente stimatore MV.

Esercizio 8.6. Si dimostri la seguente affermazione, deducendo poi da essa l'equivalenza fra (8.40) e (8.41): se $(X_i)_{i \in \mathbb{N}}$ è una successione di variabili aleatorie reali per cui esiste $c > 0$ tale che

$$\lim_{i \rightarrow +\infty} P(|X_i - c| \leq \varepsilon) = 1,$$

per ogni $\varepsilon > 0$, allora

$$\lim_{i \rightarrow +\infty} P\left(\left|1 - \frac{c}{X_i}\right| \leq \varepsilon\right) = 1.$$

8.5 Note bibliografiche

In questo capitolo abbiamo illustrato alcune interessanti e utili applicazioni della probabilità alla statistica matematica. Gli argomenti presentati, in particolare la stima per campioni normali e gli stimatori di massima verosimiglianza, sono centrali in statistica matematica, e ci hanno permesso di vedere “all’opera” alcuni degli strumenti introdotti nei capitoli precedenti. Tuttavia, un’introduzione organica alla statistica matematica non può prescindere da altri contenuti essenziali quali, ad esempio, la *verifica di ipotesi* e l’*approccio bayesiano*. Le classiche monografie di J.A. Rice [47] e R.V. Hogg, J.W. McKean e A.T. Craig [31] forniscono un’introduzione assai accessibile; un’ottima lettura, di livello più avanzato, è la monografia di A.A. Borovkov [8].

Appendice

A.1 Somme infinite

Forniamo qui le dimostrazioni delle principali proprietà delle somme infinite, che abbiamo menzionato nel capitolo introduttivo “Nozioni Preliminari” e abbiamo largamente usato in questo testo. Per comodità di lettura, ricordiamo la definizione di somma infinita.

Definizione A.1. Sia I un insieme arbitrario, e $(a_i)_{i \in I}$ una famiglia di elementi di $[-\infty, +\infty]$ indicizzata dagli elementi di I .

(i) Se $a_i \geq 0$ per ogni $i \in I$, poniamo

$$\sum_{i \in I} a_i := \sup_{A \subseteq I, |A| < \infty} \sum_{j \in A} a_j \in [0, +\infty]. \quad (\text{A.1})$$

Si noti che $\sum_{i \in I} a_i$ può assumere il valore $+\infty$, e certamente lo assume non appena uno degli addendi vale $+\infty$.

(ii) Nel caso generale di una famiglia $(a_i)_{i \in I}$ di elementi di $[-\infty, +\infty]$, non necessariamente positivi, diremo che tale famiglia *ammette somma* se almeno una delle due somme $\sum_{i \in I} a_i^+$, $\sum_{i \in I} a_i^-$ è finita (si osservi che sono entrambe somme a termini positivi, dunque sono sempre ben definite). In tal caso, la somma $\sum_{i \in I} a_i$ è definita da

$$\sum_{i \in I} a_i := \sum_{i \in I} a_i^+ - \sum_{i \in I} a_i^- \in [-\infty, +\infty]. \quad (\text{A.2})$$

Si noti che ogni famiglia con tutti i termini positivi, o tutti negativi, ammette somma.

Se una famiglia $(a_i)_{i \in I}$ ammette somma e se $\sum_{i \in I} a_i \in (-\infty, +\infty)$, diremo che la famiglia *ammette somma finita*. Chiaramente ciò accade se e solo se *entrambe* le somme $\sum_{i \in I} a_i^+$ e $\sum_{i \in I} a_i^-$ sono finite. Come mostreremo formalmente nella dimo-

strazione della Proposizione A.3, questo è equivalente a $\sum_{i \in I} |a_i| < \infty$. In particolare gli a_i devono essere tutti numeri reali.

Osserviamo che se I è un insieme finito, ogni famiglia di *numeri reali* $(a_i)_{i \in I}$ ammette somma finita, e la Definizione A.1 si riduce alla somma di numeri reali: ciò segue dal fatto che, nel caso di somme a termini positivi, l'estremo superiore in (A.1) coincide con la somma finita $\sum_{i \in I} a_i$, e tale coincidenza si estende a somme generali per la relazione (A.2).

Dimostriamo ora le proprietà delle somme infinite.

Proposizione A.2 (Monotonia). *Siano $(a_i)_{i \in I}$ e $(b_i)_{i \in I}$ famiglie di elementi di $[-\infty, +\infty]$, indicizzate dallo stesso insieme I . Se entrambe le famiglie ammettono somma (finita o infinita) e sono tali che $a_i \leq b_i$ per ogni $i \in I$, allora*

$$\sum_{i \in I} a_i \leq \sum_{i \in I} b_i. \quad (\text{A.3})$$

Dimostrazione. Se $a_i \geq 0$ e $b_i \geq 0$ e $a_i \leq b_i$ per ogni $i \in I$, allora per ogni $A \subseteq I$ finito

$$\sum_{i \in A} a_i \leq \sum_{i \in A} b_i.$$

Prendendo l'estremo superiore su A si ottiene la relazione (A.3). In generale, senza ipotesi sul segno, se $a_i \leq b_i$ allora

$$a_i^+ \leq b_i^+, \quad a_i^- \geq b_i^-,$$

da cui, per la monotonia appena dimostrata per somme a termini positivi,

$$\sum_{i \in I} a_i^+ \leq \sum_{i \in I} b_i^+, \quad \sum_{i \in I} a_i^- \geq \sum_{i \in I} b_i^-.$$

Da ciò segue facilmente:

$$\sum_{i \in I} a_i = \sum_{i \in I} a_i^+ - \sum_{i \in I} a_i^- \leq \sum_{i \in I} b_i^+ - \sum_{i \in I} b_i^- = \sum_{i \in I} b_i,$$

che conclude la dimostrazione. \square

Proposizione A.3 (Linearità). *Siano $(a_i)_{i \in I}$ e $(b_i)_{i \in I}$ famiglie di elementi di $[-\infty, +\infty]$, indicizzate dallo stesso insieme I . Se entrambe le famiglie ammettono somma finita, per ogni $\alpha, \beta \in \mathbb{R}$ la famiglia $(\alpha a_i + \beta b_i)_{i \in I}$ ammette somma finita e*

$$\sum_{i \in I} (\alpha a_i + \beta b_i) = \alpha \left(\sum_{i \in I} a_i \right) + \beta \left(\sum_{i \in I} b_i \right). \quad (\text{A.4})$$

Questa relazione vale anche se le famiglie sono positive, senza richiedere che ammettano somma finita, purché $\alpha, \beta \geq 0$.

Dimostrazione. Cominciamo col mostrare che se $(a_i)_{i \in I}$ e $(b_i)_{i \in I}$ sono famiglie positive e $\alpha, \beta \geq 0$, allora vale (A.4). Se $A \subseteq I$ è finito, allora

$$\sum_{i \in A} (\alpha a_i + \beta b_i) = \alpha \left(\sum_{i \in A} a_i \right) + \beta \left(\sum_{i \in A} b_i \right) \leq \alpha \left(\sum_{i \in I} a_i \right) + \beta \left(\sum_{i \in I} b_i \right). \quad (\text{A.5})$$

Prendendo l'estremo superiore sui sottoinsiemi $A \subseteq I$ finiti, si trova

$$\sum_{i \in I} (\alpha a_i + \beta b_i) \leq \alpha \left(\sum_{i \in I} a_i \right) + \beta \left(\sum_{i \in I} b_i \right).$$

Poniamo ora $s_a := \sum_{i \in I} a_i$ e $s_b := \sum_{i \in I} b_i$. Per ogni $\rho < \alpha s_a + \beta s_b$ possiamo scrivere $\rho = \alpha \rho_a + \beta \rho_b$ con $\rho_a < s_a$ e $\rho_b < s_b$. Per definizione di s_a e s_b , esistono allora $A_1, A_2 \subseteq I$ finiti tali che

$$\sum_{i \in A_1} a_i > \rho_a, \quad \sum_{i \in A_2} b_i > \rho_b.$$

Se $A := A_1 \cup A_2$ abbiamo allora anche $\sum_{i \in A} a_i > \rho_a$ e $\sum_{i \in A} b_i > \rho_b$, da cui segue

$$\sum_{i \in I} (\alpha a_i + \beta b_i) \geq \sum_{i \in A} (\alpha a_i + \beta b_i) = \alpha \sum_{i \in A} a_i + \beta \sum_{i \in A} b_i > \alpha \rho_a + \beta \rho_b = \rho.$$

Poiché tale disuguaglianza vale per ogni $\rho < \alpha s_a + \beta s_b$, abbiamo che

$$\sum_{i \in I} (\alpha a_i + \beta b_i) \geq \alpha s_a + \beta s_b,$$

che, assieme a (A.5) completa la dimostrazione di (A.4) per questo caso.

Si noti che, per quanto appena dimostrato, se $(a_i)_{i \in I}$ è un'arbitraria famiglia di elementi di $[-\infty, +\infty]$,

$$\sum_{i \in I} a_i^+ + \sum_{i \in I} a_i^- = \sum_{i \in I} (a_i^+ + a_i^-) = \sum_{i \in I} |a_i|.$$

Quindi la famiglia $(a_i)_{i \in I}$ ammette somma finita, ossia entrambe le somme $\sum_{i \in I} a_i^+$ e $\sum_{i \in I} a_i^-$ sono finite, se e solo se $\sum_{i \in I} |a_i| < +\infty$, come enunciato in precedenza.

Dimostriamo ora che (A.4) vale ogniquale volta $(a_i)_{i \in I}$ e $(b_i)_{i \in I}$ ammettono somma finita. Si noti, anzitutto, che, per quanto appena dimostrato sulle somme a termini positivi, e per la proprietà di monotonia,

$$\begin{aligned} \sum_{i \in I} |\alpha a_i + \beta b_i| &\leq \sum_{i \in I} (|\alpha| |a_i| + |\beta| |b_i|) = \sum_{i \in I} |\alpha| |a_i| + \sum_{i \in I} |\beta| |b_i| \\ &= |\alpha| \sum_{i \in I} |a_i| + |\beta| \sum_{i \in I} |b_i| < +\infty, \end{aligned}$$

dove il fatto che $\sum_{i \in I} |\alpha| |a_i| = |\alpha| \sum_{i \in I} |a_i|$ si dimostra elementarmente usando la definizione (A.1). Pertanto, la famiglia $(\alpha a_i + \beta b_i)_{i \in I}$ ammette somma finita. Osserviamo poi che, per (A.2), si ha che

$$\sum_{i \in I} (-a_i) = - \sum_{i \in I} a_i.$$

Ne segue che, a meno di cambiare segno a a_i o b_i , non è restrittivo assumere $\alpha \geq 0$ e $\beta \geq 0$. In tal caso abbiamo:

$$\alpha a_i + \beta b_i = (\alpha a_i + \beta b_i)^+ - (\alpha a_i + \beta b_i)^-,$$

e anche

$$\alpha a_i + \beta b_i = \alpha a_i^+ - \alpha a_i^- + \beta b_i^+ - \beta b_i^-,$$

da cui si ricava

$$(\alpha a_i + \beta b_i)^+ + \alpha a_i^- + \beta b_i^- = (\alpha a_i + \beta b_i)^- + \alpha a_i^+ + \beta b_i^+.$$

Sommando su i in quest'ultima identità, e usando la linearità per somme a termini positivi, si ha

$$\sum_{i \in I} (\alpha a_i + \beta b_i)^+ + \alpha \sum_{i \in I} a_i^- + \beta \sum_{i \in I} b_i^- = \sum_{i \in I} (\alpha a_i + \beta b_i)^- + \alpha \sum_{i \in I} a_i^+ + \beta \sum_{i \in I} b_i^+,$$

che, riordinando i termini, fornisce

$$\begin{aligned} \sum_{i \in I} (\alpha a_i + \beta b_i) &:= \sum_{i \in I} (\alpha a_i + \beta b_i)^+ - \sum_{i \in I} (\alpha a_i + \beta b_i)^- \\ &= \alpha \sum_{i \in I} a_i^+ - \alpha \sum_{i \in I} a_i^- + \beta \sum_{i \in I} b_i^+ - \beta \sum_{i \in I} b_i^- \\ &= \alpha \left(\sum_{i \in I} a_i \right) + \beta \left(\sum_{i \in I} b_i \right), \end{aligned}$$

che è quanto si voleva dimostrare. \square

Proposizione A.4 (Somma a blocchi). Sia $(a_i)_{i \in I}$ una famiglia di elementi di $[-\infty, +\infty]$ che ammette somma e sia $(I_j)_{j \in J}$ una partizione di I (si noti che nulla viene assunto sulla cardinalità di J).

Allora, per ogni $j \in J$ fissato, la sottofamiglia $(a_i)_{i \in I_j}$ ammette somma; ponendo $s_j := \sum_{i \in I_j} a_i$, anche la famiglia $(s_j)_{j \in J}$ ammette somma e si ha

$$\sum_{i \in I} a_i = \sum_{j \in J} s_j = \sum_{j \in J} \left(\sum_{i \in I_j} a_i \right). \quad (\text{A.6})$$

Dimostrazione. Cominciamo a trattare il caso in cui $a_i \geq 0$ per ogni $i \in I$. Di conseguenza le somme $s := \sum_{i \in I} a_i$ e $s_j := \sum_{i \in I_j} a_i$, per $j \in J$, sono ben definite in $[0, +\infty]$. Per la Definizione A.1, per ogni $\rho < s$ esiste $A \subseteq I$ finito tale che $\sum_{i \in A} a_i \geq \rho$. Poniamo ora $A_j := A \cap I_j$, e sia J_A il sottoinsieme di J , necessariamente finito, definito da $J_A := \{j \in J : A_j \neq \emptyset\}$. Per l'associatività delle somme finite, abbiamo

$$\rho \leq \sum_{i \in A} a_i = \sum_{j \in J_A} \sum_{i \in A_j} a_i \leq \sum_{j \in J_A} s_j \leq \sum_{j \in J} s_j,$$

dove si è usato il fatto ovvio che $\sum_{i \in A_j} a_i \leq s_j$. Poiché la disuguaglianza precedente

vale per ogni $\rho < s$, abbiamo dunque

$$s \leq \sum_{j \in J} s_j. \quad (\text{A.7})$$

Mostriamo ora la disuguaglianza inversa di (A.7). Se uno degli s_j vale $+\infty$, la disuguaglianza è verificata banalmente. Possiamo dunque assumere che $s_j < +\infty$ per ogni $j \in J$. Scegliamo $\rho < \sum_{j \in J} s_j$; esiste dunque $B \subseteq J$ finito tale che $\rho < s_B := \sum_{j \in B} s_j$. Dato che $\varepsilon := \frac{s_B - \rho}{|B|} > 0$, per ogni $j \in B$ esiste $A_j \subseteq I_j$ finito tale che

$$\sum_{i \in A_j} a_i \geq s_j - \varepsilon = s_j - \frac{s_B - \rho}{|B|}.$$

Perciò, se definiamo l'insieme finito $A := \bigcup_{j \in B} A_j$, per l'associatività delle somme finite abbiamo

$$s \geq \sum_{i \in A} a_i = \sum_{j \in B} \sum_{i \in A_j} a_i \geq \sum_{j \in B} \left(s_j - \frac{s_B - \rho}{|B|} \right) = \sum_{j \in B} s_j - |B| \frac{s_B - \rho}{|B|} = \rho.$$

Poiché la disuguaglianza precedente vale per ogni $\rho < \sum_{j \in J} s_j$, abbiamo

$$s \geq \sum_{j \in J} s_j,$$

che, assieme a (A.7), dimostra che per somme a termini positivi vale (A.6).

Supponiamo ora che $(a_i)_{i \in I}$ sia una famiglia arbitraria che ammette somma. Non è restrittivo assumere che $\sum_{i \in I} a_i^+ < +\infty$: il caso in cui $\sum_{i \in I} a_i^+ = +\infty$ e quindi, necessariamente, $\sum_{i \in I} a_i^- < +\infty$, si tratta in modo analogo. Anzitutto osserviamo che ogni sottofamiglia $(a_i)_{i \in I_j}$ ammette somma, dato che

$$\sum_{i \in I_j} a_i^+ \leq \sum_{i \in I} a_i^+ < +\infty.$$

Inoltre

$$s_j := \sum_{i \in I_j} a_i \leq \sum_{i \in I_j} a_i^+,$$

da cui segue $s_j^+ \leq \sum_{i \in I_j} a_i^+$; quindi, per la monotonia e quanto dimostrato sopra per somme a termini positivi,

$$\sum_{j \in J} s_j^+ \leq \sum_{j \in J} \sum_{i \in I_j} a_i^+ = \sum_{i \in I} a_i^+ < +\infty.$$

Questo mostra che la famiglia $(s_j)_{j \in J}$ ammette somma. Distinguiamo ora due casi.

Caso in cui $\sum_{i \in I} a_i^- < +\infty$. In tal caso, le famiglie $(\sum_{i \in I_j} a_i^+)_{j \in J}$ e $(\sum_{i \in I_j} a_i^-)_{j \in J}$ ammettono entrambe somma *finita*, dato che, per quanto sopra dimostrato,

$$\sum_{j \in J} \sum_{i \in I_j} a_i^+ = \sum_{i \in I} a_i^+ < +\infty,$$

e analogamente per le parti negative. Usando la linearità, possiamo dunque scrivere

$$\begin{aligned}\sum_{j \in J} s_j &= \sum_{j \in J} \left(\sum_{i \in I_j} a_i^+ - \sum_{i \in I_j} a_i^- \right) = \sum_{j \in J} \sum_{i \in I_j} a_i^+ - \sum_{j \in J} \sum_{i \in I_j} a_i^- \\ &= \sum_{i \in I} a_i^+ - \sum_{i \in I} a_i^- = \sum_{i \in I} a_i,\end{aligned}$$

che conclude la dimostrazione per questo caso.

Caso in cui $\sum_{i \in I} a_i^- = +\infty$. In questo caso $\sum_{i \in I} a_i = -\infty$, quindi si tratta di dimostrare che anche

$$\sum_{j \in J} s_j = -\infty, \quad (\text{A.8})$$

cioè $\sum_{j \in J} s_j^- = +\infty$. Dal fatto che $(x-y)^- = (y-x)^+ \geq y-x$, per ogni $x, y \geq 0$, segue che

$$s_j^- = \left(\sum_{i \in I_j} a_i^+ - \sum_{i \in I_j} a_i^- \right)^- \geq \sum_{i \in I_j} a_i^- - \sum_{i \in I_j} a_i^+,$$

e dunque, per ogni sottoinsieme finito $A \subseteq I$,

$$\begin{aligned}\sum_{j \in J} s_j^- &\geq \sum_{j \in A} s_j = \sum_{j \in A} \left(\sum_{i \in I_j} a_i^- - \sum_{i \in I_j} a_i^+ \right) \geq \sum_{j \in A} \left(\sum_{i \in I_j} a_i^- \right) - \sum_{j \in J} \left(\sum_{i \in I_j} a_i^+ \right) \\ &= \sum_{j \in A} \left(\sum_{i \in I_j} a_i^- \right) - \sum_{i \in I} a_i^+.\end{aligned}$$

Prendendo l'estremo superiore su $A \subseteq I$, con $|A| < \infty$, si ottiene

$$\begin{aligned}\sum_{j \in J} s_j^- &\geq \sup_{A \subseteq I, |A| < \infty} \left\{ \sum_{j \in A} \left(\sum_{i \in I_j} a_i^- \right) \right\} - \sum_{i \in I} a_i^+ = \sum_{j \in J} \left(\sum_{i \in I_j} a_i^- \right) - \sum_{i \in I} a_i^+ \\ &= \sum_{i \in I} a_i^- - \sum_{i \in I} a_i^+ = +\infty,\end{aligned}$$

e la dimostrazione è completa. \square

A.2 Una misura finitamente additiva (ma non σ -additiva) su \mathbb{N}

In questo paragrafo dimostriamo l'esistenza di una funzione $P : \mathcal{P}(\mathbb{N}) \rightarrow [0, 1]$ tale che $P(\mathbb{N}) = 1$, $P(A \cup B) = P(A) + P(B)$ per ogni coppia di sottoinsiemi disgiunti A e B di \mathbb{N} , ma P non è σ -additiva. Sia $l^\infty(\mathbb{N})$ l'insieme delle successioni limitate a valori in \mathbb{R} . Tale insieme ha una naturale struttura di spazio vettoriale reale: se $x = (x_n)$, $y = (y_n)$ sono elementi di $l^\infty(\mathbb{N})$, $\alpha, \beta \in \mathbb{R}$, definiamo $\alpha x + \beta y$ tramite $(\alpha x + \beta y)_n = \alpha x_n + \beta y_n$. Sia

$$V_0 = \{x \in l^\infty(\mathbb{N}) : x_n = 0 \text{ tranne che per un insieme finito di } n\}.$$

Inoltre sia $\mathbf{1}$ la successione con tutti gli elementi uguali ad 1. Poniamo

$$V_1 = \{\alpha \mathbf{1} + x : \alpha \in \mathbb{R}, x \in V_0\},$$

cioè V_1 è lo spazio vettoriale generato da V_0 e $\mathbf{1}$. Definiamo su V_1 il funzionale lineare

$$\lambda(\alpha \mathbf{1} + x) := \alpha.$$

Teorema A.5. *Esiste un funzionale lineare $\Lambda : l^\infty(\mathbb{N}) \rightarrow \mathbb{R}$ che coincide con λ su V_1 , e tale che*

$$\inf_n x_n \leq \Lambda(x) \leq \sup_n x_n, \quad \forall x \in l^\infty(\mathbb{N}). \quad (\text{A.9})$$

Prima di dimostrare questo risultato (non banale!), vediamo come usarlo per il nostro scopo. Per $A \subseteq \mathbb{N}$, sia $x^{(A)}$ l'elemento di $l^\infty(\mathbb{N})$ definito da

$$x_n^{(A)} = \begin{cases} 1 & \text{se } n \in A \\ 0 & \text{altrimenti.} \end{cases}$$

Definiamo dunque $P(A) = \Lambda(x^{(A)})$. Grazie alla relazione (A.9) si ha $P(A) \in [0, 1]$. Inoltre, $P(\mathbb{N}) = \Lambda(x^{(\mathbb{N})}) = \Lambda(\mathbf{1}) = \lambda(\mathbf{1}) = 1$, mentre $P(A) = \Lambda(x^{(A)}) = 0$ se A è un sottoinsieme finito di \mathbb{N} . Infine, se $A \cap B = \emptyset$, allora $x^{(A)} + x^{(B)} = x^{(A \cup B)}$, da cui, per la linearità di Λ , segue che $P(A \cup B) = P(A) + P(B)$. Dunque tale P è additiva. Non è però σ -additiva, in quanto

$$1 = P(\mathbb{N}) = P\left(\bigcup_{n \in \mathbb{N}} \{n\}\right) \neq \sum_{n \in \mathbb{N}} P(\{n\}) = 0.$$

Si tratta dunque di dimostrare il Teorema A.5. La dimostrazione, come il lettore più esperto riconoscerà, è un adattamento della classica dimostrazione del teorema di Hahn-Banach in analisi funzionale.

Ricordiamo che una relazione \preceq su un insieme non vuoto Θ si dice *relazione d'ordine parziale* se è riflessiva, antisimmetrica e transitiva, ossia se per ogni $a, b, c \in \Theta$ vale che

$$a \preceq a; \quad \text{se } a \preceq b \text{ e } b \preceq a, \text{ allora } a = b; \quad \text{se } a \preceq b \text{ e } b \preceq c, \text{ allora } a \preceq c.$$

L'insieme Θ munito di una tale relazione si dice *insieme parzialmente ordinato*. Si noti che, dati $a, b \in \Theta$, non è detto che a e b siano confrontabili, ossia che si abbia $a \preceq b$ oppure $b \preceq a$. Un sottoinsieme $\Sigma \subseteq \Theta$ tale che, per ogni $a, b \in \Sigma$, si ha necessariamente $a \preceq b$ oppure $b \preceq a$, si dice *totalmente ordinato*. Un elemento $a \in \Theta$ si dice *maggiorante di un sottoinsieme* $\Sigma \subseteq \Theta$ se si ha $c \preceq a$ per ogni $c \in \Sigma$. Infine, un elemento $a \in \Theta$ si dice *massimale* se non esiste nessun $b \in \Theta$ tale che $a \preceq b$, ad esclusione di $b = a$. Ricordiamo allora l'importante Lemma di Zorn.

Teorema A.6 (Lemma di Zorn). *Sia Θ un insieme parzialmente ordinato. Se per ogni sottoinsieme $\Sigma \subseteq \Theta$ totalmente ordinato esiste un elemento di Θ che è maggiorante di Σ , allora esiste un elemento di Θ massimale.*

Definiamo Θ come l'insieme delle coppie (W, Λ_W) , dove W è un sottospazio vettoriale di $l^\infty(\mathbb{N})$ che contiene V_1 e $\Lambda_W : W \rightarrow \mathbb{R}$ è un operatore lineare che coincide con λ su V_1 e che soddisfa le disuguaglianze in (A.9) per ogni $x \in W$. Si noti che tale insieme Θ è non vuoto, in quanto $(V_1, \lambda) \in \Theta$.

Definiamo ora su Θ la seguente relazione: $(W, \Lambda_W) \preceq (W', \Lambda_{W'})$ se W è un sottospazio vettoriale di W' e se $\Lambda_{W'}(x) = \Lambda_W(x)$ per ogni $x \in W$. Si verifica facilmente che \preceq è una relazione d'ordine parziale.

Se $\Sigma \subseteq \Theta$ è un sottoinsieme totalmente ordinato, definiamo

$$\overline{W} := \bigcup_{(W, \Lambda_W) \in \Sigma} W,$$

e consideriamo l'operatore $\overline{\Lambda} : \overline{W} \rightarrow \mathbb{R}$ definito da

$$\overline{\Lambda}(x) := \Lambda_W(x), \quad \text{se } x \in W \text{ con } (W, \Lambda_W) \in \Sigma.$$

Dal fatto che Σ è totalmente ordinato si deduce che \overline{W} è un sottospazio vettoriale di $l^\infty(\mathbb{N})$, e che la definizione di $\overline{\Lambda}$ è ben posta. Inoltre $\overline{\Lambda}$ è lineare, coincide con λ su V_1 e soddisfa (A.9) per ogni $x \in \overline{W}$. Questo mostra che $(\overline{W}, \overline{\Lambda}) \in \Theta$ e, per costruzione, $(\overline{W}, \overline{\Lambda}) \succeq (W, \Lambda_W)$ per ogni $(W, \Lambda_W) \in \Sigma$. Abbiamo perciò dimostrato che ogni sottoinsieme totalmente ordinato di Θ ammette un maggiorante.

Possiamo dunque applicare il Lemma di Zorn, concludendo che esiste un elemento massimale $(V, \Lambda) \in \Theta$. Si noti che, se $V = l^\infty(\mathbb{N})$, allora il Teorema A.5 è dimostrato. Mostriamo ora che, se $V \subsetneq l^\infty(\mathbb{N})$, l'elemento $(V, \Lambda) \in \Theta$ non può essere massimale. Questo completa la dimostrazione.

Supponiamo dunque che $V \subsetneq l^\infty(\mathbb{N})$ e sia $y \in l^\infty(\mathbb{N}) \setminus V$. Indichiamo con V' lo spazio vettoriale generato da V e y , ossia

$$V' = \{x + \alpha y : x \in V, \alpha \in \mathbb{R}\},$$

e mostriamo che il funzionale Λ si può estendere ad un funzionale $\Lambda' : V' \rightarrow \mathbb{R}$ che soddisfa (A.9) su tutto V' . In particolare $(V, \Lambda) \preceq (V', \Lambda')$ e dato che $(V, \Lambda) \neq (V', \Lambda')$, perché $V \subsetneq V'$, segue che (V, Λ) non è un elemento massimale.

Ogni funzionale lineare $\Lambda' : V' \rightarrow \mathbb{R}$ che estende Λ è della forma

$$\Lambda'(x + \alpha y) = \Lambda(x) + \alpha c,$$

dove $c := \Lambda'(y) \in \mathbb{R}$. Scrivendo $\sup(x)$ e $\inf(x)$ in luogo di $\sup_n x_n$ e $\inf_n x_n$, dobbiamo verificare che è possibile scegliere c in modo tale che

$$\inf(x + \alpha y) \leq \Lambda(x) + \alpha c \leq \sup(x + \alpha y), \quad \forall x \in V, \alpha \in \mathbb{R}. \quad (\text{A.10})$$

Procediamo con alcune semplificazioni:

- Per $\alpha = 0$ la relazione (A.10) è sempre verificata.
- Se $\alpha > 0$, dato che $\inf(\alpha z) = \alpha \inf(z)$ e $\sup(\alpha z) = \alpha \sup(z)$, moltiplicando per α^{-1} si vede che la relazione (A.10) è equivalente a

$$\inf(x' + y) \leq \Lambda(x') + c \leq \sup(x' + y), \quad \forall x' \in V, \quad (\text{A.11})$$

avendo posto $x' := \alpha^{-1}x$.

- Analogamente, se $\alpha < 0$, essendo $\inf(\alpha z) = \alpha \sup(z)$ e $\sup(\alpha z) = \alpha \inf(z)$, moltiplicando per $-\alpha^{-1}$ si vede che la relazione (A.10) è equivalente a

$$\inf(x' - y) \leq \Lambda(x') - c \leq \sup(x' - y), \quad \forall x' \in V. \quad (\text{A.12})$$

Ricavando c in (A.11) e (A.12) e rinominando per semplicità $x' \rightarrow x$, segue che l'estensione Λ' desiderata esiste se esiste $c \in \mathbb{R}$ tale che, per ogni $x, z \in V$,

$$\begin{aligned} \inf(x + y) - \Lambda(x) &\leq c \leq \sup(x + y) - \Lambda(x) \\ \Lambda(z) - \sup(z - y) &\leq c \leq \Lambda(z) - \inf(z - y). \end{aligned}$$

Osserviamo che una famiglia di intervalli chiusi di \mathbb{R} ha intersezione non vuota se e solo se ogni estremo superiore è maggiore o uguale di ogni estremo inferiore. Quindi l'esistenza di un $c \in \mathbb{R}$ soddisfacente alle disuguaglianze precedenti è equivalente alla validità delle seguenti disuguaglianze, per ogni $x, z \in V$:

- (i) $\inf(x + y) - \Lambda(x) \leq \sup(z + y) - \Lambda(z)$;
- (ii) $\inf(x + y) - \Lambda(x) \leq \Lambda(z) - \inf(z - y)$;
- (iii) $\Lambda(z) - \sup(z - y) \leq \Lambda(x) - \inf(x - y)$;
- (iv) $\Lambda(z) - \sup(z - y) \leq \sup(x + y) - \Lambda(x)$.

Per la linearità di Λ , la relazione (i) equivale a

$$\Lambda(x - z) \geq \inf(x + y) - \sup(z + y). \quad (\text{A.13})$$

Per ipotesi Λ soddisfa (A.9), perciò $\Lambda(x - z) \geq \inf(x - z)$. Dunque, per dimostrare (A.13), basta mostrare che

$$\inf(x - z) \geq \inf(x + y) - \sup(z + y) \quad (\text{A.14})$$

per ogni $x, z \in \Lambda$. Questo non è difficile: per $m \in \mathbb{N}$

$$x_m - z_m = x_m + y_m - (z_m + y_m) \geq \inf(x + y) - \sup(z + y),$$

da cui (A.14) segue facilmente. La verifica delle altre disuguaglianze (ii), (iii) e (iv) è del tutto analoga ed è lasciata al lettore.

A.3 Il principio fondamentale del calcolo combinatorio

In questo paragrafo formuliamo il principio fondamentale del calcolo combinatorio, descritto informalmente nel Paragrafo 1.2.3 (si ricordi il Teorema 1.26), in modo matematicamente più preciso, ma decisamente più tecnico.

Occorre innanzitutto esprimere astrattamente il concetto di “scelta”. Dato un insieme E , definiamo *scelta su E* una *partizione* di E , ossia una famiglia ordinata di sottoinsiemi (E_1, \dots, E_m) tali che $E = \bigcup_{i=1}^m E_i$ e $E_i \cap E_j = \emptyset$ per $i \neq j$. Gli insiemi E_i sono detti *esiti della scelta* e il loro numero m è detto *numero di esiti di una scelta*.

Per esempio, sull'insieme E delle funzioni iniettive da $\{1, \dots, k\}$ in un insieme finito $A = \{a_1, \dots, a_n\}$, la “scelta della prima componente” corrisponde alla partizione (E_1, \dots, E_n) definita da $E_i := \{f \in E : f(1) = a_i\}$ e ha dunque n esiti possibili.

Estendiamo la definizione: *due scelte successive su un insieme E* sono il dato di:

- (“*prima scelta*”) una scelta (E_1, \dots, E_m) sull'insieme E ;
- (“*seconda scelta*”) per ogni E_i fissato, una scelta $(E_{i,1}, \dots, E_{i,k_i})$ sull'insieme E_i , il cui numero di esiti k_i può in generale dipendere da i .

Sottolineiamo che la “seconda scelta” non è una scelta su E , ma una famiglia di scelte sugli esiti E_i della prima scelta. Nel caso in cui $k_i = k$ per ogni $i = 1, \dots, m$, diremo che la *seconda scelta ha un numero fissato k di esiti possibili*.

Ritornando all'insieme E delle funzioni iniettive da $\{1, \dots, k\}$ in $A = \{a_1, \dots, a_n\}$, un esempio di due scelte successive su E è dato dalla “scelta delle prime due componenti della funzione”: infatti, per ogni esito $E_i = \{f \in E : f(1) = a_i\}$ della prima scelta, la seconda scelta corrisponde alla partizione $(E_{i,j})_{j \in \{1, \dots, n\} \setminus \{i\}}$ di E_i definita da $E_{i,j} := \{f \in E : f(1) = a_i, f(2) = a_j\}$.[†] In particolare, la seconda scelta ha un numero fissato $n - 1$ di esiti possibili.

Il passaggio da due a k scelte successive è solo notazionalmente più complicato. Per definizione, *k scelte successive su un insieme E* sono il dato di:

- (“*prima scelta*”) una scelta (E_1, \dots, E_m) sull'insieme E ;
- (“ *j -esima scelta*”) per ogni $2 \leq j \leq k$ e per ogni esito $E_{i_1, \dots, i_{j-1}}$ delle prime $j - 1$ scelte, una scelta $(E_{i_1, \dots, i_{j-1}, \ell})_{\ell \in \{1, \dots, n_j^*\}}$ su $E_{i_1, \dots, i_{j-1}}$, il cui numero di esiti $n_j^* = n_j^*(i_1, \dots, i_{j-1})$ può dipendere dall'esito delle scelte precedenti.

Nel caso in cui $n_j^*(i_1, \dots, i_{j-1}) = n_j$ non dipenda da i_1, \dots, i_{j-1} , diremo che la *j -esima scelta ha un numero fissato n_j di esiti possibili*.

Possiamo finalmente riformulare il Teorema 1.26 nel modo seguente.

Teorema A.7. *Siano date k scelte successive su un insieme E , in cui ogni scelta abbia un numero fissato di esiti possibili: la prima scelta ha n_1 esiti possibili, la seconda scelta ne ha n_2 , ..., la k -esima scelta ne ha n_k . Supponiamo inoltre che*

[†] Si noti che è risultato conveniente indicizzare gli esiti della seconda scelta usando l'insieme $\{1, \dots, n\} \setminus \{i\}$ piuttosto che $\{1, \dots, n - 1\}$.

gli elementi di E siano determinati univocamente dalle k scelte, cioè

$$|E_{i_1, \dots, i_k}| = 1, \quad \text{per ogni } i_1, \dots, i_k.$$

Allora la cardinalità di E è pari a $n_1 \cdot n_2 \cdots n_k$.

Dimostrazione. Per definizione di scelte successive, $E = \bigcup_{i=1}^{n_1} E_i$; a sua volta $E_i = \bigcup_{j=1}^{n_2} E_{i,j}$, eccetera: di conseguenza vale la relazione

$$E = \bigcup_{i_1=1}^{n_1} \cdots \bigcup_{i_k=1}^{n_k} E_{i_1, \dots, i_k}. \quad (\text{A.15})$$

Se mostriamo che questa unione è disgiunta, ricordando che per ipotesi $|E_{i_1, \dots, i_k}| = 1$, si ottiene la relazione cercata:

$$|E| = \sum_{i_1=1}^{n_1} \cdots \sum_{i_k=1}^{n_k} |E_{i_1, \dots, i_k}| = n_1 \cdot n_2 \cdots n_k.$$

Siano dunque $(i_1, \dots, i_k) \neq (i'_1, \dots, i'_k)$ e mostriamo che $E_{i_1, \dots, i_k} \cap E_{i'_1, \dots, i'_k} = \emptyset$. Si ha $i_1 = i'_1, \dots, i_{j-1} = i'_{j-1}$ mentre $i_j \neq i'_j$, per un opportuno $j \in \{1, \dots, k\}$. Per definizione di scelte successive,

$$E_{i_1, \dots, i_k} \subseteq E_{i_1, \dots, i_{j-1}, i_j}, \quad E_{i'_1, \dots, i'_k} \subseteq E_{i'_1, \dots, i'_{j-1}, i'_j} = E_{i_1, \dots, i_{j-1}, i'_j}.$$

Ma gli insiemi $E_{i_1, \dots, i_{j-1}, \ell}$ al variare di ℓ sono disgiunti, perché formano una partizione di $E_{i_1, \dots, i_{j-1}}$ (o di E , se $j = 1$). Pertanto $E_{i_1, \dots, i_{j-1}, i_j} \cap E_{i_1, \dots, i_{j-1}, i'_j} = \emptyset$ e dunque $E_{i_1, \dots, i_k} \cap E_{i'_1, \dots, i'_k} = \emptyset$, come dovevasi dimostrare. \square

Ritornando per un'ultima volta all'insieme E delle funzioni iniettive da $\{1, \dots, k\}$ in $A = \{a_1, \dots, a_n\}$, è facile verificare che valgono le ipotesi del Teorema A.7: le scelte di $f(1), f(2), \dots, f(k)$ possono essere descritte attraverso “ k scelte successive su E ” e gli elementi di E sono univocamente determinati da queste scelte. Dato che la scelta di $f(i)$ ha $n - i + 1$ esiti possibili, segue dal Teorema 1.26 che vale la formula $|E| = n(n-1) \cdots (n-k+1)$.

Concludiamo fornendo una dimostrazione della formula (1.27), sulla cardinalità dell'insieme delle combinazioni, ossia

$$C_{n,k} = \frac{D_{n,k}}{k!} = \binom{n}{k}. \quad (\text{A.16})$$

Cominciamo con un piccolo risultato preparatorio. Siano D, E due insiemi finiti e sia $g : D \rightarrow E$ un'applicazione suriettiva. Per ogni $y \in E$, introduciamo il sottoinsieme $g^{-1}(y) := \{x \in D : g(x) = y\}$ costituito dagli elementi di D che vengono mandati da g in y . Supponiamo che esista $k \in \mathbb{N}$ tale che $|g^{-1}(y)| = k$, per ogni $y \in E$; in altri termini, per ogni $y \in E$ esistono esattamente k elementi $x \in D$ che vengono man-

dati da g in y . Segue allora che $|D| = k|E|$. La dimostrazione è semplice: possiamo sempre scrivere $D = \bigcup_{y \in E} g^{-1}(y)$ e inoltre l'unione è disgiunta (esercizio), quindi $|D| = \sum_{y \in E} |g^{-1}(y)| = \sum_{y \in E} k = k|E|$.

Fissiamo ora un insieme A con $|A| = n$ e $k \in \{1, \dots, n\}$. Indichiamo con $\mathcal{D}_{n,k}$ (risp. $\mathcal{C}_{n,k}$) l'insieme delle disposizioni semplici (risp. combinazioni) di k elementi estratti da A . Definiamo quindi una applicazione $g : \mathcal{D}_{n,k} \rightarrow \mathcal{C}_{n,k}$ nel modo seguente: data $f \in \mathcal{D}_{n,k}$, definiamo $g(f) := \text{Im}(f)$, dove $\text{Im}(f)$ indica l'immagine di f (ricordiamo che f è una funzione iniettiva da $\{1, \dots, k\}$ in A). È immediato verificare che g è ben definita, cioè effettivamente $g(f) \in \mathcal{C}_{n,k}$ per ogni $f \in \mathcal{D}_{n,k}$, e che g è suriettiva. Se mostriamo che $|g^{-1}(B)| = k!$, per ogni $B \in \mathcal{C}_{n,k}$, si ottiene $|\mathcal{D}_{n,k}| = k! |\mathcal{C}_{n,k}|$. Dato che $|\mathcal{D}_{n,k}| = D_{n,k} = n!/(n-k)!$, la formula (A.16) è dimostrata.

Indichiamo con S_B l'insieme delle permutazioni di B , cioè le applicazioni $\pi : B \rightarrow B$ biunivoche, e fissiamo un elemento arbitrario $f_0 \in g^{-1}(B)$. È molto facile convincersi che, per ogni $\pi \in S_B$, si ha $\pi \circ f_0 \in g^{-1}(B)$: infatti l'applicazione $\pi \circ f_0$ è iniettiva, perché lo sono sia f_0 sia π , e $\text{Im}(\pi \circ f_0) = \text{Im}(f_0) = B$, perché π è una permutazione di B . Risulta dunque ben posta l'applicazione $H : S_B \rightarrow g^{-1}(B)$ definita da $H(\pi) := \pi \circ f_0$. Se mostriamo che H è biunivoca, segue che gli insiemi S_B e $g^{-1}(B)$ sono in corrispondenza biunivoca e dunque $|g^{-1}(B)| = |S_B| = k!$, che è quanto resta da dimostrare.

Siano $\pi_1, \pi_2 \in S_B$ tali che $H(\pi_1) = H(\pi_2)$. Per ogni $b \in B$, se $i \in \{1, \dots, k\}$ è tale che $f_0(i) = b$ (si noti che tale i esiste, perché $\text{Im}(f_0) = B$), otteniamo $(\pi_1 \circ f_0)(i) = (\pi_2 \circ f_0)(i)$, cioè $\pi_1(b) = \pi_2(b)$; dato che $b \in B$ è arbitrario, segue che $\pi_1 = \pi_2$, dunque l'applicazione H è iniettiva. Se ora consideriamo un arbitrario $f \in g^{-1}(B)$, è facile costruire $\pi \in S_B$ tale che $\pi \circ f_0 = f$, cioè $H(\pi) = f$, quindi l'applicazione H è suriettiva e la dimostrazione di (A.16) è conclusa.

Tavola della distribuzione normale

La tavola seguente riporta i valori della funzione di ripartizione $\Phi(z)$ della distribuzione normale standard $N(0, 1)$, per $0 \leq z \leq 3.5$. Ricordiamo che

$$\Phi(z) := \int_{-\infty}^z \frac{e^{-\frac{1}{2}x^2}}{\sqrt{2\pi}} \mathrm{d}x.$$

I valori di $\Phi(z)$ per $z < 0$ possono essere ricavati grazie alla formula

$$\Phi(z) = 1 - \Phi(-z).$$

[illegible]

Principali distribuzioni notevoli su \mathbb{R}

Distribuzioni notevoli discrete				
	Densità discreta $p_X(k)$	Media $E(X)$	Varianza $\text{Var}(X)$	Funzione generatrice dei momenti $M_X(t) = E(e^{tX})$
Binomiale $\text{Bin}(n, p)$ $n \in \{1, 2, \dots\}$ $p \in [0, 1]$	$\binom{n}{k} p^k (1-p)^{n-k}$ $k \in \{0, 1, \dots, n\}$	np	$np(1-p)$	$(pe^t + (1-p))^n$
Bernoulli $\text{Be}(p)$ $\text{Bin}(1, p)$ $p \in [0, 1]$	$\begin{cases} p & \text{se } k = 1 \\ 1-p & \text{se } k = 0 \end{cases}$ $k \in \{0, 1\}$	p	$p(1-p)$	$pe^t + (1-p)$
Poisson $\text{Pois}(\lambda)$ $\lambda \in (0, \infty)$	$e^{-\lambda} \frac{\lambda^k}{k!}$ $k \in \{0, 1, \dots\}$	λ	λ	$e^{\lambda(e^t-1)}$
Geometrica $\text{Geo}(p)$ $p \in (0, 1]$	$p(1-p)^{k-1}$ $k \in \{1, 2, \dots\}$	$\frac{1}{p}$	$\frac{1-p}{p^2}$	$\begin{cases} \frac{p}{e^{-t} - (1-p)} & \text{se } t < \log \frac{1}{1-p} \\ +\infty & \text{se } t \geq \log \frac{1}{1-p} \end{cases}$
Distribuzioni notevoli assolutamente continue				
	Densità $f_X(x)$	Media $E(X)$	Varianza $\text{Var}(X)$	Funzione generatrice dei momenti $M_X(t) = E(e^{tX})$
Uniforme continua $U(a, b)$ $a, b \in \mathbb{R}, a < b$	$\frac{1}{b-a}$ $x \in (a, b)$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$	$\frac{e^{tb} - e^{ta}}{t(b-a)}$
Gamma $\text{Gamma}(\alpha, \lambda)$ $\alpha, \lambda \in (0, \infty)$	$\frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}$ $x \in (0, \infty)$	$\frac{\alpha}{\lambda}$	$\frac{\alpha}{\lambda^2}$	$\begin{cases} \left(\frac{\lambda}{\lambda-t}\right)^\alpha & \text{se } t < \lambda \\ +\infty & \text{se } t \geq \lambda \end{cases}$
Esponenziale $\text{Exp}(\lambda)$ $\text{Gamma}(1, \lambda)$ $\lambda \in (0, \infty)$	$\lambda e^{-\lambda x}$ $x \in (0, \infty)$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$	$\begin{cases} \frac{\lambda}{\lambda-t} & \text{se } t < \lambda \\ +\infty & \text{se } t \geq \lambda \end{cases}$
Normale $N(\mu, \sigma^2)$ $\mu \in \mathbb{R}, \sigma \in (0, \infty)$	$\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ $x \in \mathbb{R}$	μ	σ^2	$e^{\mu t + \frac{\sigma^2}{2} t^2}$

Riferimenti bibliografici

1. Aigner, M., Ziegler, G.M.: Proofs from The Book. Edizione italiana a cura di Alfio Quarteroni. Springer-Verlag Italia, Milano (2006)
2. Aldous, D., Diaconis, P.: Shuffling cards and stopping times. *Amer. Math. Monthly* **93**(5), 333–348 (1986)
3. Baxter, R.J.: Exactly solved models in statistical mechanics. Academic Press, London (1989). Ristampa dell'originale edito nel 1982
4. Bernoulli, J.: *Ars conjectandi*. Impensis Thurnisiorum, Fratrum (1713)
5. Berry, A.C.: The accuracy of the Gaussian approximation to the sum of independent variates. *Trans. Amer. Math. Soc.* **49**, 122–136 (1941)
6. Billingsley, P.: Probability and measure. 3rd ed. John Wiley & Sons, New York (1995)
7. Black, F., Scholes, M.: The pricing of options and corporate liabilities. *Journal of Political Economy* **81**(3), 637–654 (1973)
8. Borovkov, A.A.: Mathematical Statistics. Gordon and Breach Science Publishers, Amsterdam (1998)
9. von Bortkiewicz, L.: *Das Gesetz der Kleinen Zahlen*. B.G. Teubner, Leipzig (1898)
10. Bose, S.: Plancks gesetz und lichtquantenhypothese. *Z. phys* **26**(3), 178 (1924)
11. Brémaud, P.: Point processes and queues. Springer-Verlag, New York (1981)
12. Brémaud, P.: Markov chains. Springer-Verlag, New York (1999)
13. Clarke, R.: An application of the Poisson distribution. *Journal of the Institute of the Actuaries* **72**, 48 (1946)
14. Cox, J., Ross, S., Rubinstein, M.: Option pricing: A simplified approach. *Journal of financial Economics* **7**(3), 229–263 (1979)
15. Dalang, R.C.: Une démonstration élémentaire du théorème central limite. *Elem. Math.* **61**(2), 65–73 (2006)
16. Daley, D.J., Vere-Jones, D.: An introduction to the theory of point processes. Vol. I, 2nd ed. Springer-Verlag, New York (2003)
17. Daley, D.J., Vere-Jones, D.: An introduction to the theory of point processes. Vol. II, 2nd ed. Springer-Verlag, New York (2008)
18. De Moivre, A.: The doctrine of chances (1738)
19. Einstein, A.: Quantentheorie des einatomigen idealen Gases. *Sitzungsber. Kgl. Preuss. Akad. Wiss.* **261** (1924)
20. Esseen, C.G.: On the Liapounoff limit of error in the theory of probability. *Ark. Mat. Astr. Fys.* **28A**(9), 19 (1942)

21. Esseen, C.G.: A moment inequality with an application to the central limit theorem. *Skand. Aktuarietidskr.* **39**, 160–170 (1956)
22. Ewens, W.J.: *Mathematical population genetics. I, Theoretical Introduction*, 2nd ed. Springer-Verlag, New York (2004)
23. Feller, W.: *An introduction to probability theory and its applications*. Vol. I, 3rd ed. John Wiley & Sons, New York (1968)
24. Fischer, H.: *A history of the central limit theorem*. Springer-Verlag, New York (2011)
25. Folland, G.B.: *Real Analysis*. 2nd ed. John Wiley & Sons, New York (1999)
26. Fortuin, C.M., Kasteleyn, P.W., Ginibre, J.: Correlation Inequalities on Some Partially Ordered Sets. *Commun. Math. Phys.* **22**(2), 89–103 (1971)
27. Gallavotti, G.: *Statistical mechanics*. Springer-Verlag, Berlin (1999)
28. Grimmett, G.R., Stirzaker, D.R.: *Probability and random processes*. 3rd ed. Oxford University Press, New York (2001)
29. Hannabuss, K.: *An introduction to quantum theory*. Oxford University Press, New York (1997)
30. Hodges Jr., J.L., Le Cam, L.: The Poisson approximation to the Poisson binomial distribution. *Ann. Math. Statist.* **31**, 737–740 (1960)
31. Hogg, R.V., Craig, A.: *Introduction to mathematical statistics*. 4th ed. Macmillan Publishing Co., New York (1978)
32. Kolmogorov, A.: *Grundbegriffe der Wahrscheinlichkeitsrechnung*. Springer-Verlag, Berlin (1977). Ristampa dell'originale edito nel 1933
33. Korolev, V., Shevtsova, I.: An improvement of the Berry-Esseen inequality with applications to Poisson and mixed Poisson random sums. *Scand. Actuar. J.* **2012**(2), 81–105 (2012)
34. Laplace, P.: *Théorie analytique des probabilités*. V. Courcier (1812)
35. Lawler, G.F., Limic, V.: *Random walk: a modern introduction*. Cambridge University Press, Cambridge (2010)
36. Lebesgue, H.L.: *Leçons sur l'intégration et la recherche des fonctions primitives professées au Collège de France*. Cambridge University Press, Cambridge (2009). Ristampa dell'originale edito nel 1904
37. Lévy, P.: Sur la détermination des lois de probabilité par leurs fonctions caractéristiques. *CR Acad. Sci. Paris* **175**, 854–856 (1922)
38. Lévy, P.: *Calcul des probabilités*. Gauthier-Villars, Paris (1925)
39. Lifshits, M.: *Lectures on Gaussian processes*. Springer-Verlag, Berlin (2012)
40. Lindeberg, J.W.: Eine neue Herleitung des Exponentialgesetzes in der Wahrscheinlichkeitsrechnung. *Math. Z.* **15**(1), 211–225 (1922)
41. Metropolis, N.: The beginning of the Monte Carlo method. *Los Alamos Science* (15, Special Issue), 125–130 (1987)
42. Onsager, L.: Crystal statistics. I. A two-dimensional model with an order-disorder transition. *Phys. Rev.* **65**, 117–149 (1944)
43. Pascucci, A., Runggaldier, W.J.: *Finanza matematica*. Springer-Verlag Italia, Milano (2009)
44. Paulos, J.A.: *Innumeracy*. Hill and Wang, New York (2001)
45. Peierls, R.: On Ising's model of ferromagnetism. *Proc. Camb. Phil. Soc* **32**(3), 477–481 (1936)
46. Poisson, S.: *Recherches sur la probabilité des jugements en matière criminelle et en matière civile, précédées des règles générales du calcul des probabilités*. Bachelier, Paris (1837)
47. Rice, J.: *Mathematical statistics and data analysis*. Duxbury Press, Belmont (2007)

48. Riemann, B.: Ueber die Darstellbarkeit einer Function durch eine trigonometrische Reihe. Dieterich, Göttingen (1867)
49. vos Savant, M.: “Ask Marylin” column. Parade Magazine p. 16 (9 September 1990)
50. Selvin, S.: A problem in probability (letter to the editor). American Statistician **29**(1), 67 (1975)
51. Spitzer, F.: Principles of random walk. 2nd ed. Springer-Verlag, New York (1976)
52. Warshauer, M., Curtin, E.: The locker puzzle. The Mathematical Intelligencer **28**(1), 28–31 (2006)

Indice analitico

- F_X (funzione di ripartizione di una variabile aleatoria reale X), 155, 237
 $\text{Cov}(X, Y)$ (covarianza di due variabili aleatorie reali X, Y), 143, 240
 $E(X)$ (valor medio di una variabile aleatoria reale X), 131, 239
 M_X (funzione generatrice dei momenti di una variabile aleatoria reale o vettoriale X), 159, 296
 Φ (funzione di ripartizione della distribuzione normale standard), 327
 $\text{Var}(X)$ (varianza di una variabile aleatoria reale X), 143, 240
 μ_X (distribuzione di una variabile aleatoria X), 107, 236
 p_X (densità discreta di una variabile aleatoria discreta X), 107, 237
 σ -algebra, 233
– di Borel su \mathbb{R}^n , 235
– generata da una famiglia di insiemi, 234
 f_X (densità di una variabile aleatoria reale o vettoriale X), 252, 274
- approssimazione normale, 327
assenza di memoria
– della distribuzione esponenziale, 267
– della distribuzione geometrica, 179
- Bayes, formula di, 38
Bernoulli (distribuzione notevole), 104, 112, 166
Berry-Esseen, teorema di, 329
Beta (distribuzione notevole), 285
- binomiale
– coefficiente, 28
– distribuzione notevole, 168
- calcolo combinatorio, 20
– principio fondamentale del, 23, 382
campione, 345
– normale, 346, 351
Cauchy (distribuzione notevole), 259, 306
Cauchy-Schwarz, disuguaglianza di, 149
Chebyshev, disuguaglianza di, 147
chi-quadro (distribuzione notevole), 283
coefficiente
– binomiale, 28
– di correlazione (tra variabili aleatorie), 151
– multinomiale, 85
combinazioni, 28
convoluzione
– di densità, 260
– di densità discrete, 154
correzione di continuità, 332
covarianza, 143, 240
– matrice di, 295
– proprietà della, 144, 240
Cramér-Rao, limite inferiore di, 348
- densità
– congiunta e marginali, 276
– di un vettore aleatorio, 274
– di una variabile aleatoria reale, 252
densità discreta
– congiunta e marginali, 117

- di una variabile aleatoria discreta, 107, 237

- su un insieme, 12

deviazione standard, 144

disposizioni

- con ripetizione, 21

- semplici (senza ripetizione), 24

distribuzione

- congiunta e marginali, 117

- di una variabile aleatoria, 107, 236

distribuzioni notevoli

- assolutamente continue, 261

- discrete, 164

disuguaglianza

- di Cauchy-Schwarz, 149

- di Chebyshev, 147

- di Jensen, 148

- di Markov, 147

esponenziale (distribuzione notevole), 266

Eulero

- funzione Gamma di, 263

- prodotto di, 46

eventi, 10, 234

- generati da una variabile aleatoria, 105

formula

- delle probabilità totali, 37

- di Bayes, 38

- di inclusione-esclusione, 16

- di Stirling, 25

funzione di ripartizione, 155, 237

funzione Gamma di Eulero, 263

funzione generatrice dei momenti, 159, 296

funzione misurabile, 236

Gamma

- funzione di Eulero, 263

- distribuzione notevole, 263

gaussiana

- distribuzione notevole (reale), 267

- distribuzione notevole (vettoriale), 294

geometrica

- distribuzione notevole, 176

- serie, 3

i.i.d. (indipendenti e identicamente distribuite), 242, 314

inclusione-esclusione (formula di), 16

indipendenza

- di due eventi, 41

- di una famiglia di eventi, 44

- di variabili aleatorie discrete, 121

- di variabili aleatorie generali, 238

- e scorrelazione, 146

- per vettori aleatori normali, 300

intervallo di confidenza, 350

- asintotico, 355

- per la media di un campione normale, 353

- per la varianza di un campione normale, 353

ipergeometrica (distribuzione notevole), 171

Jensen, disuguaglianza di, 148

legge, *vedi* distribuzione

legge dei grandi numeri, 313

limite inferiore di Cramér-Rao, 348

linearità del valor medio, 136, 239

Markov, disuguaglianza di, 147

matrice di covarianza, 295

matrice semidefinita positiva, 295

misurabile

- funzione, 236

- spazio, 233

modello statistico parametrico, 345

momenti di una variabile aleatoria, 141

monotonia del valor medio, 136, 239

MV (massima verosimiglianza), stimatore di, 360

normale

- distribuzione notevole (reale), 267

- distribuzione notevole (vettoriale), 294

- campione, 346, 351

paradosso

- dei compleanni, 25

- dei figli, 55

- dei tre prigionieri, 60

- della scimmia di Borel, 48

- delle tre carte, 61

- di Bertrand, 292

- di Monty Hall, 52, 67

passegiate aleatorie, 77, 211

permutazioni aleatorie, 69, 166

Poisson

- distribuzione notevole, 172
- processo di, 286, 289
- probabilità
 - condizionale, 34
 - discreta su un insieme generico, 106
 - su un insieme numerabile, 10
 - su uno spazio misurabile, 233
 - totali, formula delle, 37
 - uniforme, 13
- processo di Poisson, 286, 289
- prodotto di Eulero, 46
- prove ripetute e indipendenti, 47, 105, 112, 124
- quasi certamente
 - costante (variabile aleatoria), 112, 144
 - uguali (variabili aleatorie), 110
- retta di regressione lineare, 152
- scorrelazione, 143
 - e indipendenza, 146
- semidefinita positiva (matrice), 295
- serie geometrica, 3
- somma a blocchi, 5, 376
- somme infinite, 3, 373
- spazi di probabilità
 - discreti, 10
 - generali, 233
- spazio campionario, 10, 234
- spazio misurabile, 233
- statistica campionaria, 346
- stimatore, 346
 - asintoticamente corretto, 354
 - asintoticamente normale, 356
 - consistente, 355
 - corretto, 346
 - di massima verosimiglianza, 360

- efficiente, 348
- Stirling, formula di, 25
- t di Student (distribuzione notevole), 352
- teorema
 - di approssimazione di Weierstrass, 319
 - di Berry-Esseen, 329
 - limite centrale, 324
- uniforme continua
 - distribuzione notevole (reale), 261
 - distribuzione notevole (vettoriale), 275
- uniforme discreta (distribuzione notevole), 13, 164
- valor medio
 - linearità, 136, 239
 - monotonia, 136, 239
 - per variabili aleatorie discrete, 131
 - per variabili aleatorie generali, 239
- variabili aleatorie
 - assolutamente continue (reali), 252
 - assolutamente continue (vettoriali), 274
 - costruzione canonica, 115
 - discrete, 104, 237
 - generali, 235
 - notevoli assolutamente continue, 261
 - notevoli discrete, 164
 - reali e vettoriali, 104, 235
- varianza, 143, 240
 - proprietà della, 145, 240
- vettori aleatori, 104
 - assolutamente continui, 274
 - normali, 294
- Weierstrass, teorema di approssimazione di, 319

Collana Unitext – La Matematica per il 3+2

A cura di:

A. Quarteroni (Editor-in-Chief)

L. Ambrosio

P. Biscari

C. Ciliberto

G. van der Geer

G. Rinaldi

W.J. Runggaldier

Editor in Springer:

F. Bonadei

francesca.bonadei@springer.com

Volumi pubblicati. A partire dal 2004, i volumi della serie sono contrassegnati da un numero di identificazione. I volumi indicati in grigio si riferiscono a edizioni non più in commercio.

A. Bernasconi, B. Codenotti

Introduzione alla complessità computazionale

1998, X+260 pp, ISBN 88-470-0020-3

A. Bernasconi, B. Codenotti, G. Resta

Metodi matematici in complessità computazionale

1999, X+364 pp, ISBN 88-470-0060-2

E. Salinelli, F. Tomarelli

Modelli dinamici discreti

2002, XII+354 pp, ISBN 88-470-0187-0

S. Bosch

Algebra

2003, VIII+380 pp, ISBN 88-470-0221-4

S. Graffi, M. Degli Esposti

Fisica matematica discreta

2003, X+248 pp, ISBN 88-470-0212-5

S. Margarita, E. Salinelli
MultiMath – Matematica Multimediale per l'Università
2004, XX+270 pp, ISBN 88-470-0228-1

A. Quarteroni, R. Sacco, F. Saleri
Matematica numerica (2a Ed.)
2000, XIV+448 pp, ISBN 88-470-0077-7
2002, 2004 ristampa riveduta e corretta
(1a edizione 1998, ISBN 88-470-0010-6)

13. A. Quarteroni, F. Saleri
Introduzione al Calcolo Scientifico (2a Ed.)
2004, X+262 pp, ISBN 88-470-0256-7
(1a edizione 2002, ISBN 88-470-0149-8)

14. S. Salsa
Equazioni a derivate parziali - Metodi, modelli e applicazioni
2004, XII+426 pp, ISBN 88-470-0259-1

15. G. Riccardi
Calcolo differenziale ed integrale
2004, XII+314 pp, ISBN 88-470-0285-0

16. M. Impedovo
Matematica generale con il calcolatore
2005, X+526 pp, ISBN 88-470-0258-3

17. L. Formaggia, F. Saleri, A. Veneziani
Applicazioni ed esercizi di modellistica numerica
per problemi differenziali
2005, VIII+396 pp, ISBN 88-470-0257-5

18. S. Salsa, G. Verzini
Equazioni a derivate parziali – Complementi ed esercizi
2005, VIII+406 pp, ISBN 88-470-0260-5
2007, ristampa con modifiche

19. C. Canuto, A. Tabacco
Analisi Matematica I (2a Ed.)
2005, XII+448 pp, ISBN 88-470-0337-7
(1a edizione, 2003, XII+376 pp, ISBN 88-470-0220-6)

20. F. Biagini, M. Campanino
Elementi di Probabilità e Statistica
2006, XII+236 pp, ISBN 88-470-0330-X
21. S. Leonesi, C. Toffalori
Numeri e Crittografia
2006, VIII+178 pp, ISBN 88-470-0331-8
22. A. Quarteroni, F. Saleri
Introduzione al Calcolo Scientifico (3a Ed.)
2006, X+306 pp, ISBN 88-470-0480-2
23. S. Leonesi, C. Toffalori
Un invito all'Algebra
2006, XVII+432 pp, ISBN 88-470-0313-X
24. W.M. Baldoni, C. Ciliberto, G.M. Piacentini Cattaneo
Aritmetica, Crittografia e Codici
2006, XVI+518 pp, ISBN 88-470-0455-1
25. A. Quarteroni
Modellistica numerica per problemi differenziali (3a Ed.)
2006, XIV+452 pp, ISBN 88-470-0493-4
(1a edizione 2000, ISBN 88-470-0108-0)
(2a edizione 2003, ISBN 88-470-0203-6)
26. M. Abate, F. Tovena
Curve e superfici
2006, XIV+394 pp, ISBN 88-470-0535-3
27. L. Giuzzi
Codici correttori
2006, XVI+402 pp, ISBN 88-470-0539-6
28. L. Robbiano
Algebra lineare
2007, XVI+210 pp, ISBN 88-470-0446-2
29. E. Rosazza Gianin, C. Sgarra
Esercizi di finanza matematica
2007, X+184 pp, ISBN 978-88-470-0610-2

30. A. Machì
Gruppi – Una introduzione a idee e metodi della Teoria dei Gruppi
2007, XII+350 pp, ISBN 978-88-470-0622-5
2010, ristampa con modifiche
31. Y. Biollay, A. Chaabouni, J. Stubbe
Matematica si parte!
A cura di A. Quarteroni
2007, XII+196 pp, ISBN 978-88-470-0675-1
32. M. Manetti
Topologia
2008, XII+298 pp, ISBN 978-88-470-0756-7
33. A. Pascucci
Calcolo stocastico per la finanza
2008, XVI+518 pp, ISBN 978-88-470-0600-3
34. A. Quarteroni, R. Sacco, F. Saleri
Matematica numerica (3a Ed.)
2008, XVI+510 pp, ISBN 978-88-470-0782-6
35. P. Cannarsa, T. D'Aprile
Introduzione alla teoria della misura e all'analisi funzionale
2008, XII+268 pp, ISBN 978-88-470-0701-7
36. A. Quarteroni, F. Saleri
Calcolo scientifico (4a Ed.)
2008, XIV+358 pp, ISBN 978-88-470-0837-3
37. C. Canuto, A. Tabacco
Analisi Matematica I (3a Ed.)
2008, XIV+452 pp, ISBN 978-88-470-0871-3
38. S. Gabelli
Teoria delle Equazioni e Teoria di Galois
2008, XVI+410 pp, ISBN 978-88-470-0618-8
39. A. Quarteroni
Modellistica numerica per problemi differenziali (4a Ed.)
2008, XVI+560 pp, ISBN 978-88-470-0841-0
40. C. Canuto, A. Tabacco
Analisi Matematica II
2008, XVI+536 pp, ISBN 978-88-470-0873-1
2010, ristampa con modifiche

41. E. Salinelli, F. Tomarelli
Modelli Dinamici Discreti (2a Ed.)
2009, XIV+382 pp, ISBN 978-88-470-1075-8
42. S. Salsa, F.M.G. Vegni, A. Zaretti, P. Zunino
Invito alle equazioni a derivate parziali
2009, XIV+440 pp, ISBN 978-88-470-1179-3
43. S. Dulli, S. Furini, E. Peron
Data mining
2009, XIV+178 pp, ISBN 978-88-470-1162-5
44. A. Pascucci, W.J. Runggaldier
Finanza Matematica
2009, X+264 pp, ISBN 978-88-470-1441-1
45. S. Salsa
Equazioni a derivate parziali – Metodi, modelli e applicazioni (2a Ed.)
2010, XVI+614 pp, ISBN 978-88-470-1645-3
46. C. D'Angelo, A. Quarteroni
Matematica Numerica – Esercizi, Laboratori e Progetti
2010, VIII+374 pp, ISBN 978-88-470-1639-2
47. V. Moretti
Teoria Spettrale e Meccanica Quantistica – Operatori in spazi di Hilbert
2010, XVI+704 pp, ISBN 978-88-470-1610-1
48. C. Parenti, A. Parmeggiani
Algebra lineare ed equazioni differenziali ordinarie
2010, VIII+208 pp, ISBN 978-88-470-1787-0
49. B. Korte, J. Vygen
Ottimizzazione Combinatoria. Teoria e Algoritmi
2010, XVI+662 pp, ISBN 978-88-470-1522-7
50. D. Mundici
Logica: Metodo Breve
2011, XII+126 pp, ISBN 978-88-470-1883-9
51. E. Fortuna, R. Frigerio, R. Pardini
Geometria proiettiva. Problemi risolti e richiami di teoria
2011, VIII+274 pp, ISBN 978-88-470-1746-7

52. C. Presilla
Elementi di Analisi Complessa. Funzioni di una variabile
2011, XII+324 pp, ISBN 978-88-470-1829-7
53. L. Grippo, M. Sciandrone
Metodi di ottimizzazione non vincolata
2011, XIV+614 pp, ISBN 978-88-470-1793-1
54. M. Abate, F. Tovena
Geometria Differenziale
2011, XIV+466 pp, ISBN 978-88-470-1919-5
55. M. Abate, F. Tovena
Curves and Surfaces
2011, XIV+390 pp, ISBN 978-88-470-1940-9
56. A. Ambrosetti
Appunti sulle equazioni differenziali ordinarie
2011, X+114 pp, ISBN 978-88-470-2393-2
57. L. Formaggia, F. Saleri, A. Veneziani
Solving Numerical PDEs: Problems, Applications, Exercises
2011, X+434 pp, ISBN 978-88-470-2411-3
58. A. Machì
Groups. An Introduction to Ideas and Methods of the Theory of Groups
2011, XIV+372 pp, ISBN 978-88-470-2420-5
59. A. Pascucci, W.J. Runggaldier
Financial Mathematics. Theory and Problems for Multi-period Models
2011, X+288 pp, ISBN 978-88-470-2537-0
60. D. Mundici
Logic: a Brief Course
2012, XII+124 pp, ISBN 978-88-470-2360-4
61. A. Machì
Algebra for Symbolic Computation
2012, VIII+174 pp, ISBN 978-88-470-2396-3
62. A. Quarteroni, F. Saleri, P. Gervasio
Calcolo Scientifico (5a ed.)
2012, XVIII+450 pp, ISBN 978-88-470-2744-2

- 63. A. Quarteroni
Modellistica Numerica per Problemi Differenziali (5a ed.)
2012, XVIII+628 pp, ISBN 978-88-470-2747-3
- 64. V. Moretti
Spectral Theory and Quantum Mechanics
With an Introduction to the Algebraic Formulation
2013, XVI+728 pp, ISBN 978-88-470-2834-0
- 65. S. Salsa, F.M.G. Vegni, A. Zaretti, P. Zunino
A Primer on PDEs. Models, Methods, Simulations
2013, XIV+482 pp, ISBN 978-88-470-2861-6
- 66. V.I. Arnold
Real Algebraic Geometry
2013, X+110 pp, ISBN 978-3-642-36242-2
- 67. F. Caravenna, P. Dai Pra
Probabilità. Un'introduzione attraverso modelli e applicazioni
2013, X+396 pp, ISBN 978-88-470-2594-3

La versione online dei libri pubblicati nella serie è disponibile su SpringerLink. Per ulteriori informazioni, visitare il sito:
<http://www.springer.com/series/5418>