
Giacinto Gelli

Probabilità e informazione

Manuale per il corso di Teoria dei Fenomeni Aleatori



NAPOLI 2002

© Giacinto Gelli gelli@unina.it

L'autore consente la riproduzione anche parziale del testo agli studenti del corso. Non è consentito modificare il testo, diffonderlo, pubblicarlo anche con mezzi telematici senza il consenso scritto dell'autore.

Prima versione (1.0): settembre 2001.

Seconda versione (2.0): febbraio 2002.

Terza versione (3.0): ottobre 2002.

Quarta versione (3.1): marzo 2003.

Quinta versione (3.2): settembre 2003.

Dedicato ad Annalisa, Andrea, ed Alice.

Prefazione

*Poiché non è dal lavoro che nasce la civiltà:
essa nasce dal tempo libero e dal giuoco.*
Alexandre Koyré, "I filosofi e la macchina"

Questo libro costituisce un tentativo di fornire un'introduzione snella, ma rigorosa, ai concetti fondamentali di *probabilità* ed *informazione* per gli allievi dei corsi di laurea dell'Ingegneria dell'Informazione.

Il libro è organizzato in 10 capitoli ed alcune appendici; nei capitoli 1 e 2 si espongono le basi della teoria della probabilità; i capitoli 3, 4 e 5 sono dedicati allo studio della teoria di una variabile aleatoria; i capitoli 6 e 7 si occupano della teoria di due variabili aleatorie; il capitolo 8 generalizza molti dei concetti esposti nei capitoli precedenti al caso di $n > 2$ variabili aleatorie e discute brevemente i teoremi limite (legge dei grandi numeri e teorema limite fondamentale); nel capitolo 9 sono introdotte le distribuzioni condizionali; infine, il capitolo 10 è dedicato all'introduzione dei concetti fondamentali della teoria dell'informazione (entropia, codifica di sorgente, primo teorema di Shannon, codici di Huffmann). Gli argomenti marcati con il simbolo ★ possono essere saltati ad una prima lettura, senza pregiudicare la comprensione del resto. Il libro è corredato da numerosi esempi svolti e da oltre 200 esercizi proposti, suddivisi per capitolo; gli esercizi contrassegnati con il simbolo ★ sono di maggiore difficoltà.

Per la comprensione del testo, sono richieste conoscenze di base di calcolo combinatorio, di analisi reale (teoria delle funzioni di una e più variabili, integrazione delle funzioni di una e più variabili, derivazione delle funzioni di una e più variabili, successioni e serie) e di algebra lineare e geometria (vettori, matrici, determinanti). È necessaria anche una conoscenza *operativa* dell'impulso di Dirac (le proprietà fondamentali sono richiamate nell'appendice D).

Il libro è disponibile su Internet in formato pdf alla seguente URL:

<http://www.die.unina.it/GruppoTLC/gelli/didattica/CorsoFAlaurea/materiale>

ed è stato composto dall'autore utilizzando $\text{\LaTeX}2\epsilon$. Commenti, segnalazioni di errori e suggerimenti possono essere indirizzati a gelli@unina.it.

Si ringraziano gli studenti della Facoltà di Ingegneria dell'Università di Napoli per il loro incoraggiamento, la loro inesauribile curiosità, e particolarmente per le osservazioni che hanno consentito di correggere molti degli errori presenti nelle precedenti versioni.

Giacinto Gelli, ottobre 2002

Principali notazioni

A, B, C	insiemi
$\mathcal{A}, \mathcal{B}, \mathcal{C}$	classi (collezioni di insiemi)
\emptyset	insieme vuoto
$\omega \in A$	ω appartiene ad A
$\omega \notin A$	ω non appartiene ad A
$A \subseteq B$	A è un sottoinsieme di B
$A \subset B$	A è un sottoinsieme proprio di B
$A \cup B, A + B$	unione di A e B
$A \cap B, AB$	intersezione di A e B
$A - B$	differenza tra A e B
\overline{A}	complemento di A
$A \times B$	prodotto cartesiano di A e B
\triangleq	uguale per definizione
\mathbb{N}	insieme dei numeri naturali $\{1, 2, \dots\}$
$\mathbb{N}_0 = \mathbb{N} \cup \{0\}$	insieme dei numeri naturali, zero incluso $\{0, 1, 2, \dots\}$
\mathbb{Z}	insieme dei numeri interi relativi $\{\dots, -2, -1, 0, 1, 2, \dots\}$
\mathbb{R}	insieme dei numeri reali
$\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, \infty\}$	insieme ampliato dei numeri reali
$[a, b]$	intervallo $a \leq x \leq b$
$[a, b[$	intervallo $a \leq x < b$
$]a, b]$	intervallo $a < x \leq b$
$]a, b[$	intervallo $a < x < b$
$] -\infty, b[$	intervallo $x < b$
$] -\infty, b]$	intervallo $x \leq b$
$]a, \infty[$	intervallo $x > a$
$]a, \infty]$	intervallo $x \geq a$
(a, b)	indica indifferentemente un qualunque intervallo di estremi a e b
Ω	spazio campione
\mathcal{S}	σ -campo costruito su uno spazio campione Ω
$\mathcal{P}(\Omega)$	collezione delle parti di Ω
$P(A)$	probabilità dell'evento A
$P(A B)$	probabilità condizionata dell'evento A dato l'evento B
X, Y, Z	variabili aleatorie
$\mathbf{x}, \mathbf{y}, \mathbf{z}$	vettori
$\mathbf{A}, \mathbf{B}, \mathbf{C}$	matrici
$\det(\mathbf{A})$	determinante della matrice \mathbf{A}
\mathbf{A}^{-1}	inversa della matrice \mathbf{A}
\mathbf{A}^T	trasposta della matrice \mathbf{A}

Indice

1	Probabilità elementare	1
1.1	Introduzione	1
1.2	Richiami di teoria degli insiemi	3
1.3	Probabilità: definizioni preliminari	6
1.4	Probabilità assiomatica	8
1.4.1	Campi e σ -campi	8
1.4.2	Assiomi di Kolmogorov	9
1.4.3	Proprietà elementari della probabilità	10
1.4.4	Spazi di probabilità	12
1.4.5	Proprietà di continuità della probabilità★	12
1.5	Altri approcci alla teoria della probabilità	13
1.5.1	Approccio frequentista	14
1.5.2	Approccio classico	14
1.5.3	Vantaggi (e svantaggi) dell'approccio assiomatico	15
1.6	Esempi di costruzione di spazi di probabilità	16
1.6.1	Spazi di probabilità discreti	16
1.6.2	Spazi di probabilità continui★	18
1.7	Esercizi proposti	24
2	Probabilità condizionale e indipendenza	27
2.1	Introduzione	27
2.2	Probabilità condizionale	28
2.2.1	Interpretazioni della probabilità condizionale	29
2.2.2	Legge della probabilità composta	30
2.2.3	Regola della catena	31
2.2.4	Teorema della probabilità totale e teorema di Bayes	32
2.3	Indipendenza tra eventi	35
2.3.1	Indipendenza di tre o più eventi	36
2.3.2	Indipendenza condizionale tra eventi	37
2.4	Esperimenti combinati	37
2.4.1	Esperimenti indipendenti	39
2.5	Elementi di un sistema di comunicazione	41
2.5.1	Sorgente di informazione	42
2.5.2	Canale di comunicazione e canale binario simmetrico (BSC)	42
2.5.3	Sorgenti e canali senza memoria★	46
2.6	Esercizi proposti	48

3	Variabili aleatorie	51
3.1	Introduzione	51
3.1.1	Definizione formale di variabile aleatoria	54
3.2	Funzione di distribuzione cumulativa (CDF)	54
3.2.1	Proprietà della CDF	56
3.2.2	Variabili aleatorie discrete, continue, miste	58
3.2.3	Percentile e mediana ★	59
3.3	Funzione densità di probabilità (pdf)	61
3.3.1	Proprietà della pdf	62
3.4	Funzione distribuzione di probabilità (DF)	64
3.4.1	Proprietà della DF	65
3.5	Variabili aleatorie notevoli	66
3.5.1	Variabile aleatoria di Bernoulli	67
3.5.2	Variabile aleatoria binomiale e problema delle prove ripetute	67
3.5.3	Variabile aleatoria binomiale negativa	70
3.5.4	Variabile aleatoria geometrica	71
3.5.5	Variabile aleatoria di Poisson	72
3.5.6	Variabile aleatoria uniforme	72
3.5.7	Variabile aleatoria gaussiana o normale	73
3.5.8	Variabile aleatoria esponenziale	75
3.5.9	Variabile aleatoria di Laplace (esponenziale bilatera)	76
3.5.10	Variabile aleatoria di Rayleigh	76
3.5.11	Variabile aleatoria di tipo “mixture”	77
3.5.12	Relazioni tra variabile aleatoria binomiale e gaussiana: i teoremi di de Moivre-Laplace★	78
3.6	Esercizi proposti	81
4	Trasformazioni di una variabile aleatoria	85
4.1	Introduzione	85
4.1.1	Condizioni da imporre alla funzione $g(x)$ ★	86
4.2	Caratterizzazione statistica di $Y = g(X)$	87
4.2.1	Calcolo della CDF di $Y = g(X)$	87
4.2.2	Calcolo della DF di $Y = g(X)$	92
4.2.3	Calcolo della pdf di $Y = g(X)$	93
4.3	Problema inverso: determinazione di $g(x)$	96
4.3.1	Generazione di una variabile aleatoria con CDF assegnata	98
4.3.2	Generazione automatica di numeri casuali	102
4.3.3	Algoritmo “middle-square” (Von Neumann)	102
4.3.4	Algoritmo lineare congruente	103
4.3.5	Test statistici sui generatori★	105
4.4	Esercizi proposti	107
5	Caratterizzazione sintetica di una variabile aleatoria	109
5.1	Introduzione	109
5.2	Media di una variabile aleatoria	110
5.2.1	Teorema fondamentale della media	113
5.2.2	Proprietà della media	113
5.3	Varianza e valor quadratico medio di una variabile aleatoria	114
5.3.1	Proprietà della varianza	116
5.4	Momenti di una variabile aleatoria	118
5.4.1	Relazione tra momenti e momenti centrali	119
5.5	Disuguaglianze notevoli	121
5.6	Esercizi proposti	124

6	Copie di variabili aleatorie	127
6.1	Introduzione	127
6.2	Funzione di distribuzione cumulativa (CDF) congiunta	128
6.2.1	Proprietà della CDF congiunta	129
6.3	Funzione densità di probabilità (pdf) congiunta	130
6.3.1	Proprietà della pdf congiunta	131
6.4	Funzione di distribuzione di probabilità (DF) congiunta	133
6.5	Statistiche congiunte e marginali	134
6.6	Copie di variabili aleatorie indipendenti	137
6.6.1	Proprietà delle variabili aleatorie indipendenti	138
6.7	Trasformazioni di coppie di variabili aleatorie	139
6.7.1	Trasformazione $2 \rightarrow 1$	139
6.7.2	Trasformazione $2 \rightarrow 2$	141
6.7.3	Metodo della variabile ausiliaria	144
6.8	Variabili aleatorie complesse [★]	145
6.9	Esercizi proposti	148
7	Caratterizzazione sintetica di una coppia di variabili aleatorie	151
7.1	Introduzione	151
7.2	Teorema fondamentale della media per una coppia di variabili aleatorie	152
7.3	Momenti congiunti di una coppia di variabili aleatorie	153
7.4	Misure di correlazione di una coppia di variabili aleatorie	154
7.4.1	Correlazione	154
7.4.2	Spazio vettoriale di variabili aleatorie	154
7.4.3	Covarianza	156
7.4.4	Coefficiente di correlazione	157
7.4.5	Incorrelazione tra due variabili aleatorie	158
7.5	Stima lineare a minimo errore quadratico medio [★]	159
7.5.1	Principio di ortogonalità	161
7.6	Esercizi proposti	163
8	Vettori di variabili aleatorie	165
8.1	Introduzione	165
8.2	Caratterizzazione statistica di n variabili aleatorie	166
8.2.1	Funzione di distribuzione cumulativa (CDF)	166
8.2.2	Funzione densità di probabilità (pdf)	166
8.2.3	Funzione di distribuzione di probabilità (DF)	167
8.2.4	Proprietà delle distribuzioni congiunte di n variabili aleatorie	167
8.3	Trasformazioni di n variabili aleatorie	168
8.4	Variabili aleatorie indipendenti	170
8.5	Momenti di n variabili aleatorie	172
8.5.1	Vettore delle medie	173
8.5.2	Matrice di correlazione	173
8.5.3	Matrice di covarianza	175
8.5.4	Incorrelazione	176
8.6	Teoremi limite e convergenza di una sequenza di variabili aleatorie [★]	179
8.6.1	Legge dei grandi numeri	180
8.6.2	Teorema limite fondamentale	183
8.7	Esercizi proposti	186
9	Distribuzioni e medie condizionali	189
9.1	Introduzione	189
9.2	Distribuzioni condizionali per una variabile aleatoria	189
9.2.1	Funzione di distribuzione cumulativa (CDF) condizionale	190
9.2.2	Funzione densità di probabilità (pdf) condizionale	191
9.2.3	Funzione distribuzione di probabilità (DF) condizionale	191
9.2.4	Teorema della probabilità totale per CDF, pdf, DF	193
9.2.5	Probabilità a posteriori di un evento [★]	195
9.2.6	Probabilità a posteriori dato $X = x$ [★]	195

9.2.7	Teorema della probabilità totale (versione continua)★	198
9.2.8	Teorema di Bayes per le pdf★	199
9.3	Distribuzioni condizionali per coppie di variabili aleatorie	199
9.3.1	Distribuzioni condizionali dato $X = x$ ed $Y = y$	201
9.4	Distribuzioni condizionali per vettori di variabili aleatorie	203
9.4.1	Indipendenza condizionale e regola della catena per le pdf	204
9.5	Media condizionale e momenti condizionali	205
9.5.1	Teorema della media condizionale	206
9.5.2	Generalizzazione al caso di coppie di variabili aleatorie★	207
9.6	Esercizi proposti	212
10	Elementi di teoria dell'informazione	215
10.1	Introduzione	215
10.2	Misura dell'informazione ed entropia	217
10.2.1	Autoinformazione	218
10.2.2	Entropia	219
10.2.3	Proprietà dell'entropia	220
10.2.4	Entropia congiunta	221
10.3	Sorgenti di informazione	222
10.3.1	Entropia di sorgente	223
10.3.2	Tasso d'informazione di una sorgente	224
10.3.3	Sorgenti discrete senza memoria (DMS)	225
10.3.4	Codifica di sorgente	226
10.4	Codici per la compattazione dati	227
10.4.1	Codici a lunghezza fissa	227
10.4.2	Codici a lunghezza variabile	227
10.4.3	Codici univocamente decifrabili	228
10.4.4	Codici a prefisso	228
10.4.5	Condizioni per l'univoca decifrabilità	230
10.5	Efficienza dei codici per la compattazione dati	231
10.5.1	Codici di Shannon	232
10.5.2	Codifica a blocchi e primo teorema di Shannon	234
10.5.3	Efficienza dei codici a lunghezza fissa	236
10.5.4	Codici di Huffmann	236
10.6	Esercizi proposti	240
A	Fattoriale e coefficiente binomiale	243
A.1	Fattoriale	243
A.2	Coefficiente binomiale	243
A.3	Espansioni binomiali	244
B	Elementi di calcolo combinatorio	245
B.1	Introduzione	245
B.2	Schema fondamentale del conteggio	246
B.3	Applicazione al calcolo delle probabilità nel gioco del poker	249
C	La funzione $\mathbb{G}(x)$	255
C.1	La funzione $\mathbb{G}(x)$	255
D	L'impulso di Dirac	259
D.1	Impulso di Dirac	259
E	Richiami di algebra lineare	263
E.1	Definizioni ed operazioni fondamentali	263
E.1.1	Matrici e vettori	263
E.1.2	Somma di due matrici e prodotto per uno scalare	264
E.1.3	Prodotto di due matrici (righe per colonne)	264
E.1.4	Trasposizione	264
E.2	Operazioni e proprietà delle matrici quadrate	265

E.2.1	Determinante	265
E.2.2	Inversa	265
E.2.3	Matrici diagonali	266
E.2.4	Matrici simmetriche e forme quadratiche	266
F	Identità matematiche notevoli	269
F.1	Sommatorie e serie	269
F.1.1	Sommatorie di potenze di interi	269
F.1.2	Somma dei primi n termini di una serie geometrica	269
F.1.3	Serie geometrica	270
F.2	Formula di Leibnitz	270
	Bibliografia	271

Probabilità elementare

In questo capitolo si introducono i concetti basilari della teoria della probabilità. Dopo aver fornito le definizioni preliminari di esperimento, spazio campione, ed evento, si mostra come costruire in modo rigoroso una legge di probabilità utilizzando l'approccio assiomatico di Kolmogorov e si presentano le proprietà elementari della probabilità. Si accenna poi ad alcuni approcci alternativi allo studio della probabilità (classico e frequentista), discutendo i vantaggi e gli svantaggi dell'approccio assiomatico e motivando la scelta di quest'ultimo. I concetti introdotti vengono infine applicati per costruire leggi di probabilità su spazi campione di tipo discreto oppure continuo.

1.1 Introduzione

La teoria della probabilità è uno strumento matematico utile per lo studio dei cosiddetti *fenomeni aleatori*, che sono fenomeni complessi o di difficile modellizzazione, il cui esito non è prevedibile a priori con certezza, ma che tuttavia presentano una qualche forma di *regolarità*; per questo motivo, il comportamento di tali fenomeni può essere descritto solo attraverso opportune grandezze globali o medie.

Per esempio, il lancio di una moneta su un tavolo è un fenomeno fisico che può essere certamente descritto in termini delle equazioni matematiche tipiche della cinematica e della dinamica; tuttavia è estremamente difficile, se non praticamente impossibile, pur supponendo di conoscere esattamente la forma, la massa, la velocità iniziale della moneta, le caratteristiche del tavolo, e ogni altro parametro del problema, prevedere quale faccia della moneta si manifesterà in un singolo lancio. Nonostante ciò, la nostra intuizione ci dice che se lanciamo la moneta (supposta non truccata) un numero sufficientemente elevato di volte, la percentuale di volte che si presenterà la faccia *testa* o la faccia *croce* sarà prossima al 50%. Quindi, pur non essendo possibile prevedere il risultato di un *singolo* lancio, riconosciamo che il fenomeno aleatorio presenta una qualche forma

di *regolarità* se si considera un numero *elevato* di lanci o di ripetizioni dell'esperimento. La teoria della probabilità si occupa proprio di individuare, studiare e modellare tali regolarità.¹

Un altro esempio di fenomeno aleatorio è un fluido gassoso, composto da un numero elevato di particelle in moto casuale. È praticamente impossibile descrivere il comportamento del gas descrivendo il comportamento di ogni particella che lo compone; tuttavia l'aggregato delle particelle tende ad esibire proprietà regolari: ad esempio, la *pressione* del gas stesso è una quantità perfettamente definita e misurabile. In questo caso, la regolarità del fenomeno si manifesta in quanto esso, a livello macroscopico, è composto da un numero *elevato* di particelle microscopiche, ciascuna delle quali presenta un comportamento aleatorio. La disciplina che studia il comportamento dei gas con un approccio basato sulla teoria della probabilità prende il nome di *meccanica statistica*.

Altri fenomeni aleatori che possono essere convenientemente modellati attraverso la teoria della probabilità sono, ad esempio, l'arrivo di utenti ad uno sportello di una banca, nel quale è impossibile prevedere con esattezza l'istante di arrivo di ciascun utente, ma il comportamento globale dell'insieme degli utenti (ad esempio, la lunghezza media della coda allo sportello) può essere modellato con esattezza. In un ambito completamente differente, gli "arrivi" possono essere le telefonate che giungono ad una centrale telefonica, e la teoria della probabilità può servire a dimensionare opportunamente il numero di linee di tale centrale. L'applicazione della teoria della probabilità a tali problemi ha determinato la nascita della disciplina denominata *teoria delle code*, ampiamente utilizzata nell'analisi e nel progetto delle reti di telecomunicazioni.

In ambito economico e finanziario, la teoria della probabilità è stata utilizzata con successo per modellare aggregati composti da un gran numero di soggetti economici, quali ad esempio i mercati nei quali avvengono le transazioni di borsa. Se infatti è impossibile prevedere con esattezza il comportamento del singolo investitore, tuttavia il comportamento globale di un gran numero di investitori tende ad esibire regolarità che rendono possibile una descrizione basata sui modelli della teoria della probabilità.

Un altro campo nel quale la teoria della probabilità trova un'importante applicazione è l'elaborazione e la trasmissione dell'*informazione*; bisogna infatti osservare che, per sua natura, il concetto di informazione è intrinsecamente legato a quello di *impredicibilità*. Ad esempio, l'affermazione "stanotte farà buio" non convoglia nessuna informazione, semplicemente perché è una affermazione certa, perfettamente predicibile. Viceversa, una affermazione poco probabile, quale "domani il pianeta Terra sarà invaso dai marziani" convoglia una grande quantità di informazione, perché poco probabile, e quindi non predicibile. La disciplina che studia i problemi associati all'informazione con approccio probabilistico prende il nome di *teoria dell'informazione*; alcuni aspetti basilari di tale disciplina saranno introdotti e discussi nel capitolo 10.

Abbiamo fornito alcuni esempi, certamente non esaustivi, di applicazione della teoria della probabilità, che dovrebbero evidenziare l'ampia portata e la rilevanza di tale disciplina. Siamo adesso pronti a porre le basi di tale teoria, che ha un forte contenuto matematico, ma che cercheremo di trattare in modo semplice, e con continuo ricorso ad esempi. In particolare, prima di addentrarci nel vivo dello studio della teoria della probabilità, richiamiamo brevemente nel paragrafo successivo gli elementi fondamentali della *teoria degli insiemi*. Il lettore in possesso di

¹L'esempio del lancio di una moneta non è scelto a caso: la nascita stessa della teoria della probabilità è attribuita da molti storici alla necessità di calcolare le percentuali di vittoria o di sconfitta per i più comuni giochi d'azzardo (lancio di dadi, roulette, poker, etc.). Un episodio storicamente documentato, cui spesso si fa risalire la nascita della moderna teoria della probabilità, è la corrispondenza (1654) tra il matematico B. Pascal ed il giocatore cavalier de Meré su una particolare scommessa relativa al gioco dei dadi (nota come il "paradosso di de Meré", vedi esercizio 2.13).

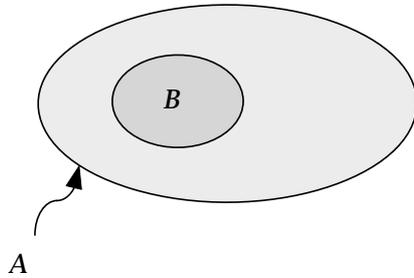


Fig. 1.1. L'insieme B è sottoinsieme dell'insieme A ($B \subseteq A$).

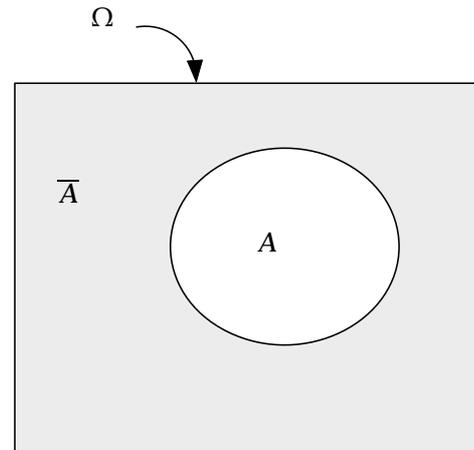


Fig. 1.2. Il complemento $\bar{A} = \Omega - A$ di un insieme A (in grigio).

sufficiente familiarità con tali concetti può scorrere rapidamente il paragrafo 1.2 per familiarizzare con la notazione utilizzata, oppure saltare direttamente al paragrafo 1.3, dove si introducono i primi elementi di teoria della probabilità.

1.2 Richiami di teoria degli insiemi

Un *insieme* A è una collezione di oggetti, chiamati *elementi* dell'insieme. Un insieme può essere definito per enumerazione, vale a dire specificando in dettaglio i suoi elementi, per esempio $A = \{\omega_1, \omega_2, \dots, \omega_n\}$ o $A = \{\text{bianco, rosso, verde}\}$, oppure descrivendo quali proprietà devono possedere tali elementi, ad esempio² $A = \{\omega \in \mathbb{R} \text{ tali che } \omega \geq 0\}$. Per indicare che ω è un elemento di A , si usa la notazione $\omega \in A$. L'insieme vuoto \emptyset è l'insieme che non contiene elementi. Due insiemi A e B si dicono coincidenti, e si scrive $A = B$, se essi contengono gli stessi elementi.

Per agevolare la comprensione delle relazioni che coinvolgono gli insiemi, è utile ricorrere ad un particolare tipo di rappresentazione grafica, denominata *diagramma di Venn*, nel quale gli insiemi sono rappresentati come porzioni del piano, come ad esempio in Fig. 1.1 oppure in Fig. 1.2.

Un sottoinsieme B di A è un insieme i cui elementi sono anche elementi di A (Fig. 1.1). Per indicare che B è un sottoinsieme di A (ovvero è incluso in A) si usa la notazione $B \subseteq A$; se esiste almeno un elemento di A che non appartiene a B , B si dice sottoinsieme *proprio* di A , e si indica $B \subset A$ (relazione di inclusione stretta). Si assume che l'insieme vuoto sia sottoinsieme di un qualunque insieme. Nella logica formale, la relazione di inclusione corrisponde all'implicazione logica. Notiamo che risulta $A = B$ se e solo se $A \subseteq B$ e $B \subseteq A$.

Dato un insieme Ω , si dice *classe* una collezione \mathcal{C} di sottoinsiemi di Ω . In particolare, la classe di *tutti* i sottoinsiemi di Ω (ivi incluso Ω e l'insieme vuoto \emptyset) prende il nome di *collezione delle parti di Ω* , e si denota con $\mathcal{P}(\Omega)$.

²Qui e nel seguito denotiamo con \mathbb{R} l'insieme dei numeri reali, con \mathbb{N} l'insieme dei numeri naturali (interi positivi escluso lo zero), con \mathbb{Z} l'insieme dei numeri relativi (interi positivi e negativi, zero incluso). In generale, il significato delle principali notazioni utilizzate è richiamato all'inizio del libro.

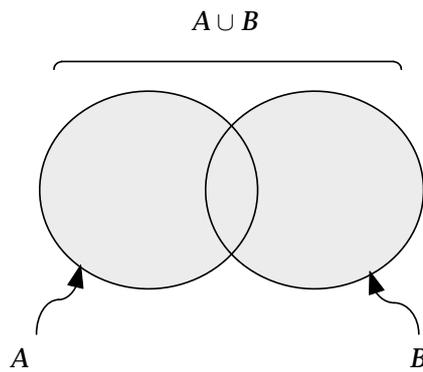


Fig. 1.3. L'unione $A \cup B$ di due insiemi (in grigio).

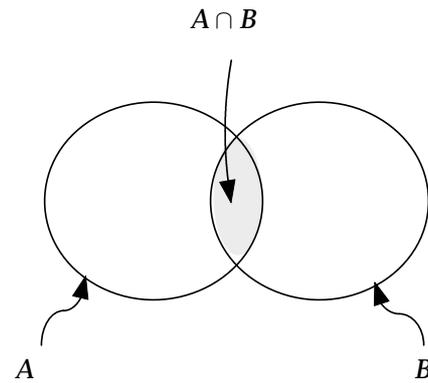


Fig. 1.4. L'intersezione $A \cap B$ di due insiemi (in grigio).

La differenza $A - B$ tra due insiemi è l'insieme che contiene gli elementi di A che non appartengono a B .

Sia A un sottoinsieme di Ω . Il *complemento* \bar{A} di A (rispetto ad Ω) è l'insieme contenente tutti gli elementi di Ω che non appartengono ad A (Fig. 1.2), ovvero $\bar{A} = \Omega - A$. Nella logica formale, il complemento corrisponde all'operazione di NOT.

L'*unione* o somma di due insiemi A , B è l'insieme che contiene tutti gli elementi di A , di B , o di entrambi (Fig. 1.3). L'unione di due insiemi si denota con $A \cup B$ oppure $A + B$, e gode della proprietà commutativa:

$$A \cup B = B \cup A.$$

L'operazione di unione, inoltre, si può estendere a più di due insiemi in maniera naturale, in quanto essa gode della proprietà associativa:

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

il che giustifica la scrittura $A \cup B \cup C$ oppure $A + B + C$ senza parentesi. Nella logica formale, l'unione corrisponde all'operazione di OR (non esclusivo).

L'*intersezione* o prodotto di due insiemi A , B è l'insieme che contiene tutti gli elementi comuni ad A e B (Fig. 1.4). L'intersezione di due insiemi si denota con $A \cap B$ oppure AB , e gode della proprietà commutativa:

$$A \cap B = B \cap A.$$

L'operazione di intersezione, inoltre, si può estendere a più di due insiemi in maniera naturale, in quanto essa gode della proprietà associativa:

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

il che giustifica la scrittura $A \cap B \cap C$ oppure ABC senza parentesi. Inoltre l'intersezione gode della proprietà distributiva rispetto all'unione:

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

che ha un'interpretazione più immediata se scritta con il simbolismo algebrico:

$$A(B + C) = AB + AC.$$

Nella logica formale, l'intersezione corrisponde all'operazione di AND.

Il *prodotto cartesiano* di due insiemi A, B è l'insieme i cui elementi sono le coppie *ordinate* (ω_1, ω_2) , con $\omega_1 \in A$ e $\omega_2 \in B$. Il prodotto cartesiano tra due insiemi si denota con $A \times B$; notiamo che, poiché le coppie sono ordinate, il prodotto cartesiano non è in generale commutativo, nel senso che $A \times B \neq B \times A$; un caso particolare è quello in cui $A = B$, per il quale vale la proprietà commutativa e il prodotto cartesiano $A \times A$ si indica semplicemente come A^2 . Ad esempio, il prodotto cartesiano dell'insieme \mathbb{R} dei numeri reali con sé stesso è il piano "cartesiano" \mathbb{R}^2 . L'operazione di prodotto cartesiano, infine, si può estendere a più di due insiemi in maniera naturale, in quanto essa gode della proprietà associativa:

$$(A \times B) \times C = A \times (B \times C),$$

il che giustifica la scrittura $A \times B \times C$ senza parentesi.

Notiamo che ragionando ricorsivamente le operazioni di unione, intersezione e prodotto cartesiano possono essere estese anche al caso di *infiniti* insiemi.

Due insiemi A e B si dicono *mutuamente esclusivi* o *disgiunti* o *incompatibili* se $A \cap B = \emptyset$. Dati n insiemi A_1, A_2, \dots, A_n , essi si dicono mutuamente esclusivi o disgiunti o incompatibili se $A_i \cap A_j = \emptyset$ per ogni $i \neq j$. Dati n insiemi A_1, A_2, \dots, A_n mutuamente esclusivi, si dice che essi costituiscono una *partizione* di Ω se $\bigcup_{k=1}^n A_k = \Omega$. I concetti di insiemi mutuamente esclusivi e di partizione si possono estendere al caso di infiniti insiemi: ad esempio, gli intervalli $[k, k+1[$, $k \in \mathbb{Z}$, sono mutuamente esclusivi e costituiscono una partizione infinita dell'insieme \mathbb{R} .

La *cardinalità* $\text{card}(A)$ di un insieme A è il numero degli elementi di A . Se A contiene infiniti elementi, $\text{card}(A) = \infty$. La cardinalità di un insieme infinito A si dice *infinita numerabile* se gli infiniti elementi di A si possono porre in corrispondenza biunivoca con l'insieme \mathbb{N} dei numeri naturali; se ciò non è possibile, la cardinalità di A si dirà *infinita continua*. Ad esempio, l'insieme A dei numeri non negativi e pari è un insieme con cardinalità infinita numerabile; l'insieme dei numeri razionali è un insieme con cardinalità infinita numerabile; viceversa l'insieme $A = [0, 1] = \{x \in \mathbb{R} \text{ tali che } 0 \leq x \leq 1\}$ ha cardinalità infinita continua.³

E' facile verificare che la cardinalità gode delle seguenti proprietà:

1. se A, B sono mutuamente esclusivi, allora $\text{card}(A \cup B) = \text{card}(A) + \text{card}(B)$;
2. in generale, se A, B non sono mutuamente esclusivi, si ha $\text{card}(A \cup B) = \text{card}(A) + \text{card}(B) - \text{card}(A \cap B)$;
3. se $A \subseteq B$, allora $\text{card}(A) \leq \text{card}(B)$;
4. $\text{card}(\emptyset) = 0$;
5. $\text{card}(A \times B) = \text{card}(A) \text{card}(B)$.

Le *leggi di de Morgan*, utilizzate anche nella logica formale, mettono in relazione tra loro le operazioni di unione, intersezione e complementazione:

$$\overline{A \cup B} = \overline{A} \cap \overline{B}, \quad (1.1)$$

$$\overline{A \cap B} = \overline{A} \cup \overline{B}. \quad (1.2)$$

Tali leggi possono essere estese anche all'unione e all'intersezione di più di due insiemi, e finanche al caso di infiniti insiemi.

³Questo risultato fu dimostrato da G. Cantor (1845–1918) con un procedimento ora noto come *procedimento diagonale di Cantor* e suscitò notevole scalpore e perplessità tra i matematici suoi contemporanei.

1.3 Probabilità: definizioni preliminari

Alla base della teoria della probabilità sono i concetti “primitivi” di *esperimento*, *spazio campione*, ed *evento*.

Definizione (esperimento). Un esperimento (aleatorio) è una procedura sperimentale con un ben definito insieme di possibili risultati, il cui esito non è prevedibile a priori.

► *Esempio 1.1.* Un possibile esperimento è il lancio di una moneta, con risultati convenzionalmente denominati “testa” (T) e “croce” (C); oppure il lancio di un dado, con possibili risultati una faccia marcata con un numero intero tra uno e sei; oppure ancora l'estrazione di un numero al gioco del lotto, con possibili risultati un numero intero tra 1 e 90. ◀

Definizione (spazio campione). Lo spazio campione Ω (finito o infinito) associato ad un esperimento è l'insieme di tutti i possibili risultati ω dell'esperimento.

► *Esempio 1.2.* Nel lancio di una moneta lo spazio campione è $\Omega = \{T, C\}$; nel lancio di un dado, lo spazio campione è $\Omega = \{1, 2, 3, 4, 5, 6\}$; nell'estrazione di un numero al gioco del lotto, lo spazio campione è $\Omega = \{1, 2, \dots, 89, 90\}$. ◀

Definizione (evento). Dato uno spazio campione Ω , si dice evento un sottoinsieme A di Ω .

► *Esempio 1.3.* Nel lancio di una moneta un possibile evento è $A = \{T\}$ (evento elementare, costituito da un solo elemento); nel lancio di un dado, un possibile evento è $A = \{\text{pari}\} = \{2, 4, 6\}$; nell'estrazione di un numero al gioco del lotto, un possibile evento è $A = \{\text{minore di } 10\} = \{1, 2, 3, \dots, 9\}$. ◀

Si definisce *prova* una singola ripetizione di un esperimento. Supponiamo allora di effettuare una prova e di ottenere il risultato $\omega \in \Omega$: diremo allora che, nella prova considerata, si è verificato l'evento A , se $\omega \in A$. Allo stesso modo, diremo che:

- non si è verificato l'evento A , se $\omega \notin A$ o, equivalentemente, se $\omega \in \bar{A}$;
- si sono verificati gli eventi A e B , se $\omega \in A \cap B$;
- si è verificato l'evento A oppure B , se $\omega \in A \cup B$ (gli eventi A e B potrebbero verificarsi anche entrambi, ovvero l'OR non è esclusivo).

Ad esempio, poichè $\omega \in \Omega$ sempre, l'evento Ω (evento *certo*) si verifica ad ogni prova, mentre l'evento \emptyset (evento *impossibile*) non si verifica in nessuna prova. Tra i possibili eventi, i più semplici sono quelli del tipo $A = \{\omega\}$, costituiti cioè da un singolo elemento di Ω ; tali eventi “atomici” (in quanto non ulteriormente decomponibili in eventi più semplici) si dicono *eventi elementari*. Notiamo la distinzione tra *risultato* ω ed *evento elementare* $\{\omega\}$ (evidenziato dall'uso delle parentesi graffe): il risultato ω è il generico elemento dello spazio campione Ω (non è un evento), mentre l'evento elementare $\{\omega\}$ è l'insieme costituito da un solo elemento (è un evento).

► *Esempio 1.4.* Nel lancio di un dado, consideriamo gli eventi $A = \{\text{pari}\}$, $B = \{\text{maggiore o uguale a } 3\}$, $C = \{\text{minore di } 2\}$. Se il risultato dell'esperimento è il numero 4, diremo allora che:

- si è verificato l'evento A , ovvero "è uscito un numero pari";
- si è verificato l'evento B , ovvero "è uscito un numero maggiore o uguale a 3";
- non si è verificato l'evento C , ovvero "non è uscito un numero minore di 2".

Analogamente, diremo che si sono verificati gli eventi A e B , e si sono verificati gli eventi A oppure C . ◀

Possiamo adesso introdurre i concetti di *spazio degli eventi* ed una prima definizione di *probabilità*. Per *spazio degli eventi* intendiamo la classe \mathcal{S} di tutti gli eventi di interesse (poiché gli eventi sono sottoinsiemi di Ω , si tratta di una classe, cioè di una collezione di insiemi). La *probabilità* è una funzione P definita⁴ sullo spazio degli eventi \mathcal{S} e a valori in $[0, 1]$:

$$P: A \in \mathcal{S} \rightarrow P(A) \in [0, 1].$$

In altri termini, una legge di probabilità consiste nell'assegnare ad ogni evento A un numero compreso tra 0 ed 1 che in qualche modo *misura* il grado di incertezza associato al verificarsi dell'evento stesso.

A questo punto sorge un problema fondamentale: dato un qualsiasi esperimento, è abbastanza semplice identificare in maniera non ambigua lo spazio campione Ω , gli eventi A , lo spazio dei possibili eventi \mathcal{S} . Ad esempio, sembra naturale scegliere come spazio degli eventi \mathcal{S} la classe $\mathcal{P}(\Omega)$ di *tutti* i sottoinsiemi di Ω (vedremo poi che questa scelta non è sempre possibile). Ma come è possibile specificare la legge di probabilità? Vediamo un semplice esempio.

► *Esempio 1.5.* Consideriamo il lancio di una moneta, il cui spazio campione denotiamo con $\Omega = \{T, C\}$. Come spazio degli eventi, consideriamo la collezione $\mathcal{P}(\Omega)$ delle parti di Ω , ovvero la classe di tutti i sottoinsiemi di Ω , incluso Ω e \emptyset . In generale, la collezione delle parti, per un insieme con N elementi, contiene 2^N sottoinsiemi;⁵ nel caso in esame, poniamo $\mathcal{S} = \mathcal{P}(\Omega) = \{\{T\}, \{C\}, \{T, C\}, \{\emptyset\}\}$. Possiamo assegnare la probabilità a *tutti* gli eventi di \mathcal{S} nel seguente modo:

$$\begin{aligned} P(\{T\}) &= P(\{C\}) = 1/2, && \text{per simmetria;} \\ P(\{T, C\}) &= 1, && \text{evento certo;} \\ P(\{\emptyset\}) &= 0, && \text{evento impossibile.} \end{aligned}$$

In questo caso, allora, abbiamo assegnato un valore numerico di probabilità ad un qualunque evento dello spazio degli eventi, e quindi abbiamo costruito una legge di probabilità. ◀

Nel semplice esempio precedente una plausibile legge di probabilità si è ottenuta sulla base di considerazioni intuitive e per motivi di simmetria. Tuttavia, per trattare casi più complicati è necessario ricorrere ad un approccio sistematico. In particolare, è necessario introdurre degli *assiomi* o dei *postulati*⁶ a cui deve soddisfare una legge di probabilità; questa strada è quella seguita dall'*approccio assiomatico*, introdotto nel 1933 dal matematico russo A. N. Kolmogorov (1903–1987),⁷ ed è quella ritenuta più soddisfacente dal punto di vista matematico. Tuttavia,

⁴Notiamo che la probabilità è una funzione che opera, anziché su numeri, su insiemi (eventi): una tale funzione è denominata *funzione di insieme*.

⁵Tale risultato si può facilmente motivare, se pensiamo che individuare un particolare sottoinsieme di un insieme Ω con N elementi equivale a costruire una stringa di N bit, nella quale ai simboli "0" si associa la mancanza nel sottoinsieme dell'elemento di Ω corrispondente, mentre ai simboli "1" si associa la sua presenza. Poiché è possibile costruire 2^N distinte stringhe di N bit, tale sarà il numero dei distinti sottoinsiemi di Ω .

⁶Ricordiamo che, in una teoria formale, un assioma o un postulato è un'asserzione che non dev'essere dimostrata. Ad esempio, l'assioma fondamentale della geometria euclidea è il cosiddetto *assioma delle rette parallele*: "in un piano, per un punto non appartenente ad una retta, passa una ed una sola retta parallela alla retta data".

⁷Il contributo di Kolmogorov apparve per la prima volta con il titolo "Grundebegriffe der Wahrscheinlichkeitrechnung" (Fondamenti del calcolo delle probabilità) nella rivista tedesca *Ergebnisse Der Mathematik* nel 1933; una traduzione in inglese (curata da N. Morrison) di non difficile reperibilità è Kolmogorov, A. N. "Foundations of the theory of probability", Chelsea Publishing Co., New York, 1956 (ristampata da American Mathematical Society, 2000).

l'approccio assiomatico soffre di una limitazione fondamentale: esso è un approccio *incompleto* (nel senso che non consente di determinare univocamente i *valori* delle probabilità da attribuire agli eventi), come discuteremo più approfonditamente nel seguito.

► *Esempio 1.6.* L'approccio assiomatico ci consentirà di costruire leggi di probabilità su esperimenti più complessi, quali quelli ad esempio che hanno un numero *infinito* di possibili risultati. Si pensi, ad esempio, all'esperimento che consiste nel contare il numero di automobili che passano ad un casello autostradale in un determinato intervallo di tempo; sebbene in pratica tale numero sarà limitato superiormente, in mancanza di informazioni su tale limite superiore possiamo assumere come spazio campione $\Omega = \{0, 1, 2, \dots\}$, ovvero l'insieme \mathbb{N}_0 dei numeri interi non negativi, avente cardinalità *infinita numerabile*. Un altro esempio è l'esperimento consistente nel misurare la durata (il "tempo di vita") di un dispositivo (si pensi, ad esempio, ad una lampadina appena montata). In questo caso potremmo assumere come spazio campione Ω un opportuno intervallo $[0, a]$ di numeri reali positivi, anche se, non conoscendo il valore di a (il massimo tempo di vita) risulta più semplice assumere $\Omega = [0, \infty[$; in questo caso abbiamo a che fare con uno spazio campione Ω di cardinalità *infinita continua*. La costruzione di leggi di probabilità su spazi campione aventi cardinalità infinita (in particolare, continua) non può essere affrontata soltanto con considerazioni intuitive, ma richiede una formulazione più rigorosa dei principi della probabilità. ◀

1.4 Probabilità assiomatica

Per costruire una legge di probabilità secondo l'approccio assiomatico dobbiamo richiedere qualche proprietà particolare allo spazio \mathcal{S} degli eventi di interesse. In particolare, dobbiamo richiedere che \mathcal{S} possieda la struttura di *campo* o, più precisamente di σ -*campo*.

1.4.1 Campi e σ -campi

Iniziamo col definire il concetto di *campo*:

Definizione (campo). Una classe \mathcal{S} non vuota di eventi si dice *campo* se soddisfa le seguenti proprietà:

1. $A \in \mathcal{S} \Rightarrow \bar{A} \in \mathcal{S}$ (*chiusura rispetto al complemento*);
2. $A, B \in \mathcal{S} \Rightarrow A \cup B \in \mathcal{S}$ (*chiusura rispetto all'unione*).

Sulla base delle proprietà 1-2, è facile dimostrare che, se \mathcal{S} è un campo, si ha anche:

$$1'. \quad \Omega, \emptyset \in \mathcal{S}.$$

Prova. Infatti, poichè \mathcal{S} è non vuoto, contiene almeno un elemento $A \Rightarrow \bar{A} \in \mathcal{S}$ (per la proprietà 1) $\Rightarrow A \cup \bar{A} = \Omega \in \mathcal{S}$ (per la proprietà 2) $\Rightarrow \bar{\Omega} = \emptyset \in \mathcal{S}$ (per la proprietà 1). ◻

$$2'. \quad A, B \in \mathcal{S} \Rightarrow A \cap B \in \mathcal{S} \quad (\textit{chiusura rispetto all'intersezione}).$$

Prova. Se $A, B \in \mathcal{S} \Rightarrow \bar{A}, \bar{B} \in \mathcal{S}$ (per la proprietà 1) $\Rightarrow \bar{A} \cup \bar{B} \in \mathcal{S}$ (per la proprietà 2) $\Rightarrow \overline{\bar{A} \cup \bar{B}} \in \mathcal{S}$ (per la proprietà 1). Ma $\overline{\bar{A} \cup \bar{B}} = A \cap B$ per le leggi di de Morgan. ◻

L'applicazione ripetuta delle proprietà 2 e 2' mostra che ogni insieme che possa essere espresso come unione e/o intersezione di un numero *finito* di elementi di \mathcal{S} appartiene anch'esso ad \mathcal{S} . Tale proprietà non rimane valida, tuttavia, se si considera un numero *infinito* di insiemi, che è un caso di interesse nella teoria della probabilità. È allora necessario estendere il concetto di campo al caso di infiniti insiemi, definendo il cosiddetto σ -*campo*:

Definizione (σ -campo). Un σ -campo \mathcal{S} di eventi è un campo che soddisfa, oltre alle proprietà 1 e 2, anche la seguente:

$$3. \{A_n\}_{n=1}^{\infty} \in \mathcal{S} \Rightarrow \bigcup_{n=1}^{\infty} A_n \in \mathcal{S} \quad (\text{chiusura rispetto all'unione numerabile}).$$

Applicando le leggi di de Morgan e la chiusura rispetto al complemento, è facile verificare che anche $\bigcap_{n=1}^{\infty} A_n$ appartiene a \mathcal{S} (proprietà di *chiusura rispetto all'intersezione numerabile*).

Poiché Ω e \emptyset devono necessariamente appartenere ad \mathcal{S} , ne segue che $\mathcal{S} = \{\emptyset, \Omega\}$ è il *più piccolo* σ -campo che è possibile costruire: esso prende il nome di *σ -campo banale*. D'altra parte, la classe $\mathcal{P}(\Omega)$ delle parti, poiché contiene *tutti* i sottoinsiemi di Ω , conterrà senz'altro il complemento, l'unione e l'intersezione numerabile di qualunque insieme; dunque $\mathcal{P}(\Omega)$ è il *più grande* σ -campo che è possibile costruire.

Osserviamo in conclusione che la distinzione tra campo e σ -campo è significativa se il numero di eventi possibili è *infinito*, il che può accadere solo se lo spazio campione Ω ha infiniti elementi. Se lo spazio campione Ω ha un numero N finito di elementi, la classe delle parti $\mathcal{P}(\Omega)$ contiene un numero finito (2^N) di sottoinsiemi, e quindi è un campo ed anche un σ -campo. Vedremo che in questo caso è effettivamente possibile scegliere come σ -campo $\mathcal{S} = \mathcal{P}(\Omega)$ e costruire senza problemi valide leggi di probabilità su Ω (cfr. § 1.6.1). La scelta $\mathcal{S} = \mathcal{P}(\Omega)$ è lecita (§ 1.6.1) anche nel caso in cui Ω risulti di cardinalità infinita numerabile. Viceversa, vedremo nel § 1.6.2 che la scelta $\mathcal{S} = \mathcal{P}(\Omega)$ non è lecita nel caso in cui Ω ha cardinalità infinita *continua*, in quanto tale σ -campo (che, ricordiamo, è il più grande σ -campo) è in genere *troppo grande* per definire una valida legge di probabilità su di esso.

1.4.2 Assiomi di Kolmogorov

Dopo l'introduzione delle definizioni preliminari, siamo in grado di fornire una definizione rigorosa della probabilità:

Definizione (probabilità). Assegnato uno spazio campione Ω ed un σ -campo \mathcal{S} di eventi di Ω , si definisce *probabilità* una funzione P definita in \mathcal{S} , a valori reali non negativi, tale da soddisfare i seguenti tre assiomi (*assiomi di Kolmogorov*):

I. $P(A) \geq 0$ per ogni $A \in \mathcal{S}$ (*assioma di non negatività*);

II. $P(\Omega) = 1$ (*assioma di normalizzazione*);

III. Se $\{A_n\}_{n=1}^{\infty}$ è una successione di eventi mutuamente esclusivi ($A_i \cap A_j = \emptyset, \forall i \neq j$) di \mathcal{S} , allora $P(\bigcup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} P(A_n)$ (*assioma di numerabile additività*).

L'intera teoria della probabilità discende dai precedenti assiomi in maniera *deduttiva*.⁸ Abbiamo già osservato che assegnare i valori di probabilità agli eventi equivale a *misurare* il livello di incertezza associato agli stessi. In effetti, bisogna osservare che una funzione definita su un insieme Ω , che soddisfa assiomi analoghi a quelli di Kolmogorov, viene proprio definita dai matematici una *misura* (casi elementari di misura sono la *lunghezza*, l'*area*, ed il *volume*); pertanto, il contributo più significativo di Kolmogorov è stato in sostanza quello di riconoscere che, per definire una

⁸Una teoria si dice *deduttiva* se ricava i casi particolari a partire da principi generali; viceversa, si dice *induttiva* se ricava i principi generali a partire da casi particolari. Il principio di induzione è stato spesso severamente messo in discussione da scienziati e filosofi; per una interessante discussione critica sui due approcci si veda K. Popper, "Logica della ricerca scientifica", Einaudi, 1970.

corretta teoria della probabilità, quest'ultima va inquadrata come un caso particolare della *teoria della misura*. Notiamo, in particolare, che l'assioma di normalizzazione impone che la misura di Ω sia unitaria, e per questo motivo si parla anche della probabilità come di una misura *normalizzata*. Va osservato che nel seguito, per mantenere la trattazione ad un livello elementare, *non* faremo uso di tale analogia in maniera estesa; tuttavia, sfrutteremo l'analogia tra probabilità e misura per giustificare intuitivamente alcune proprietà della probabilità, quali quelle presentate nel paragrafo seguente.

1.4.3 Proprietà elementari della probabilità

A partire dagli assiomi di Kolmogorov, applicando semplici concetti di teoria degli insiemi, è possibile ricavare le proprietà elementari della probabilità riportate in questo paragrafo. Per ciascuna di queste proprietà, è fornita una dimostrazione formale rigorosa; tuttavia, una giustificazione più intuitiva si può dare sfruttando l'analogia tra probabilità e misura e ragionando sui diagrammi di Venn; in tal caso, possiamo identificare la probabilità di un insieme A con l'area della superficie che occupa sul diagramma di Venn. In particolare, per l'assioma di normalizzazione, l'analogia richiede che lo spazio campione Ω abbia "area" unitaria. Per brevità, tutti gli insiemi di cui si calcolano le probabilità nelle proprietà che seguono sono sempre assunti appartenenti al σ -campo \mathcal{S} .

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

Prova. Scegliendo $A_1 = \Omega$ e $A_n = \emptyset, \forall n > 1$ (tali A_n risultano chiaramente mutuamente esclusivi), risulta $\cup_{n=1}^{\infty} A_n = \Omega + \emptyset = \Omega$. Per l'assioma III allora si ha:

$$P(\Omega) = P(\cup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} P(A_n) = P(\Omega) + \sum_{n=2}^{\infty} P(\emptyset)$$

da cui risulta necessariamente $P(\emptyset) = 0$. □

2. $A \cap B = \emptyset \Rightarrow P(A \cup B) = P(A) + P(B)$ (finita additività).

Prova. Segue dall'assioma III e dalla proprietà 1, scegliendo $A_1 = A, A_2 = B, A_n = \emptyset, \forall n > 2$. □

3. $P(\bar{A}) = 1 - P(A).$

Prova. Poichè $A \cup \bar{A} = \Omega$ e $A \cap \bar{A} = \emptyset$, per la proprietà 2 e per l'assioma II si ha:

$$P(A \cup \bar{A}) = P(A) + P(\bar{A}) = P(\Omega) = 1 \Rightarrow P(\bar{A}) = 1 - P(A).$$

□

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

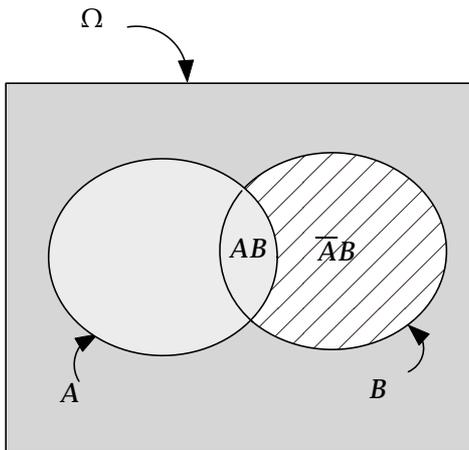


Fig. 1.5. Diagramma di Venn delle relazioni $A \cup B = A \cup \bar{A}B$ e $B = AB \cup \bar{A}B$.

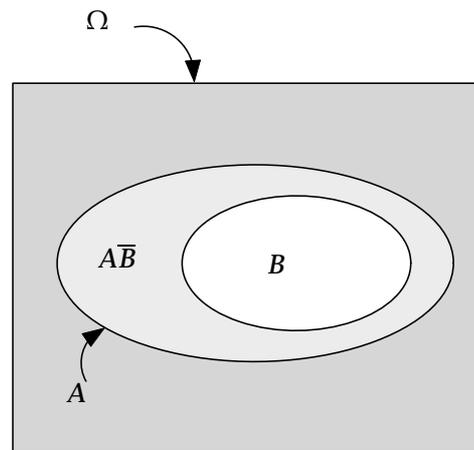


Fig. 1.6. Diagramma di Venn della relazione $A = B \cup A\bar{B}$ (valida se $B \subseteq A$).

Prova. Utilizzando i diagrammi di Venn (Fig. 1.5) è facile verificare che:

$$A \cup B = A \cup \bar{A}B$$

con A e $\bar{A}B$ mutuamente esclusivi. Allo stesso modo (Fig. 1.5), si ha:

$$B = \Omega \cap B = (A + \bar{A}) \cap B = AB \cup \bar{A}B$$

con AB e $\bar{A}B$ mutuamente esclusivi. Applicando la proprietà 2 si ha:

$$P(A \cup B) = P(A) + P(\bar{A}B),$$

$$P(B) = P(AB) + P(\bar{A}B).$$

Eliminando $P(\bar{A}B)$ tra le due equazioni si ottiene il risultato. □

Poiché $P(A \cap B) \geq 0$, risulta $P(A \cup B) \leq P(A) + P(B)$ (*disuguaglianza di Boole*). Si ha uguaglianza se e solo se $P(A \cap B) = 0$, ovvero se gli eventi A e B sono mutuamente esclusivi.

5. $B \subseteq A \Rightarrow P(B) \leq P(A).$

Prova. Utilizzando i diagrammi di Venn (Fig. 1.6) è facile verificare che, se $B \subseteq A$, si ha:

$$A = B \cup A\bar{B}$$

con B e $A\bar{B}$ mutuamente esclusivi. Per la proprietà 2 si ha:

$$P(A) = P(B \cup A\bar{B}) = P(B) + P(A\bar{B}) \Rightarrow P(B) \leq P(A)$$

perché $P(A\bar{B}) \geq 0$. □

6. $P(B) \leq 1.$

Prova. Segue direttamente dalla proprietà precedente e dall'assioma II, scegliendo $A = \Omega$. □

1.4.4 Spazi di probabilità

In sostanza, per definire una legge di probabilità, occorre specificare: 1) uno spazio campione Ω ; 2) un σ -campo \mathcal{S} di eventi di Ω ; 3) una funzione P definita su \mathcal{S} e a valori in $[0, 1]$ che soddisfi gli assiomi I-III di Kolmogorov (vedi § 1.4.2). La terna (Ω, \mathcal{S}, P) prende il nome di *spazio di probabilità*. Si noti che, nell'approccio assiomatico, l'intera teoria della probabilità viene costruita in maniera *deduttiva* a partire dagli assiomi di Kolmogorov. Questo significa che a partire dai principi generali (gli assiomi) e dalle probabilità di eventi semplici, si ricavano le probabilità di eventi complessi applicando le proprietà formali del calcolo delle probabilità, tra cui quelle ricavate nel § 1.4.3.

► *Esempio 1.7.* Riprendiamo l'esempio del lancio di una moneta. Abbiamo definito lo spazio campione $\Omega = \{T, C\}$ ed il σ -campo $\mathcal{S} = \{\{T\}, \{C\}, \{T, C\}, \{\emptyset\}\}$. Per definire una legge di probabilità bisogna allora assegnare le probabilità agli eventi. A tale scopo è sufficiente assegnare le probabilità ai cosiddetti *eventi elementari* $\{T\}$ e $\{C\}$. Una scelta ragionevole è:

$$P(\{T\}) = P(\{C\}) = 1/2,$$

tuttavia se assegniamo le probabilità come:

$$P(\{T\}) = 1/3, \quad P(\{C\}) = 2/3;$$

è facile vedere che anche tale assegnazione soddisfa gli assiomi di Kolmogorov. Allora qual è la legge di probabilità "corretta"? ◀

L'esempio precedente mette in luce la principale limitazione dell'approccio assiomatico di Kolmogorov, ovvero il fatto che esso è un sistema di assiomi *incompleto*, non consente cioè di determinare univocamente *quali* debbano essere le probabilità degli eventi. Come si fa allora a capire quale sia la legge di probabilità "corretta"? In pratica una volta definita una legge di probabilità che soddisfa all'approccio assiomatico, si utilizza tale legge per effettuare *previsioni* sull'esperimento (ad esempio, per calcolare probabilità di eventi più complessi a partire da probabilità di eventi semplici). Se le previsioni sono accurate (validazione sperimentale) le probabilità ipotizzate sono corrette, altrimenti è necessario modificare la legge (i valori) di probabilità. Il processo si può iterare fino ad avere un accordo soddisfacente tra valori teorici e valori sperimentali. La disciplina che si occupa di validare sperimentalmente le previsioni probabilistiche e/o di ricavare i valori di probabilità a partire dai dati sperimentali va sotto il nome di *statistica*.

1.4.5 Proprietà di continuità della probabilità★

Introduciamo in questa sezione⁹ una proprietà che, sebbene non frequentemente utilizzata nel calcolo delle probabilità, è estremamente importante per alcune derivazioni teoriche. Ricordiamo che la probabilità è una funzione P avente per insieme di definizione il σ -campo \mathcal{S} degli eventi. Mostriamo ora che tale funzione P è *continua*, nel senso che se $\{A_n\}_{n=1}^{\infty}$ è una successione di eventi di \mathcal{S} , tali che $\lim_n A_n = A$, allora:

$$\lim_n P(A_n) = P(\lim_n A_n) = P(A). \quad (1.3)$$

Tale continuità sembra simile a quella comunemente introdotta per le funzioni reali di una variabile reale, ma va interpretata con cautela: poiché infatti \mathcal{S} non è un insieme numerico, non

⁹Le sezioni contrassegnate dal simbolo ★ possono essere saltate ad una prima lettura.

è chiaro che in che senso vada intesa la convergenza della successione di *insiemi* A_n all'*insieme* A . Una trattazione rigorosa richiederebbe l'introduzione e l'uso di concetti matematici avanzati, quali la teoria degli *spazi metrici* e/o degli *spazi topologici*. Qui considereremo un caso più semplice, nel quale definiremo il concetto di limite solo per particolari successioni di insiemi: in particolare, diremo che A_n è una successione *decescente* (rispetto alla relazione di inclusione) se $A_n \supseteq A_{n+1}, \forall n \in \mathbb{N}$; viceversa, diremo che A_n è una successione *crescente* (rispetto alla relazione di inclusione) se $A_n \subseteq A_{n+1}, \forall n \in \mathbb{N}$. Porremo allora le seguenti definizioni di limite:

$$\lim_n A_n \triangleq \bigcap_{n=1}^{\infty} A_n, \quad \text{se } \{A_n\}_{n=1}^{\infty} \text{ è decrescente;} \quad (1.4)$$

$$\lim_n A_n \triangleq \bigcup_{n=1}^{\infty} A_n, \quad \text{se } \{A_n\}_{n=1}^{\infty} \text{ è crescente.} \quad (1.5)$$

Sulla base di queste definizioni, è possibile enunciare il seguente teorema:

Teorema 1.1 (continuità della probabilità). Sia (Ω, \mathcal{S}, P) uno spazio di probabilità.

i) Se $\{A_n\}_{n=1}^{\infty}$ è una successione decrescente di eventi, posto $A = \lim A_n \triangleq \bigcap_{n=1}^{\infty} A_n$, si ha:

$$\lim_n P(A_n) = P(\lim_n A_n) = P(A). \quad (1.6)$$

ii) Se $\{A_n\}_{n=1}^{\infty}$ è una successione crescente di eventi, posto $A = \lim A_n \triangleq \bigcup_{n=1}^{\infty} A_n$, si ha:

$$\lim_n P(A_n) = P(\lim_n A_n) = P(A). \quad (1.7)$$

Prova. La dimostrazione non è complicata, ma viene omessa per brevità, rimandando il lettore interessato a [7]. Limitiamoci ad osservare che poiché \mathcal{S} è un σ -campo, allora $A \in \mathcal{S}$, essendo ottenuto come intersezione o unione numerabile di eventi di \mathcal{S} ; pertanto ha senso calcolare $P(A)$ in entrambi i casi. \square

Notiamo che è possibile costruire sequenze $\{A_n\}_{n=1}^{\infty}$ decrescenti tali che $\bigcap_{n=1}^{\infty} A_n = \emptyset$. In tal caso, l'applicazione del risultato i) del teorema precedente consente di affermare che, per sequenze siffatte, risulta

$$\lim_n P(A_n) = P(\emptyset) = 0. \quad (1.8)$$

Si può mostrare (si veda [2] oppure [4]) che la (1.8) è logicamente equivalente all'assioma III di Kolmogorov (numerabile additività), e quindi potrebbe sostituirlo in una diversa assiomatizzazione della teoria della probabilità. Per tale motivo, la relazione (1.8) viene talvolta chiamata *assioma di continuità*.¹⁰

1.5 Altri approcci alla teoria della probabilità

L'approccio assiomatico è quello più recentemente (1933) proposto per la teoria della probabilità. Storicamente, nel corso degli anni, oltre all'approccio assiomatico si sono sviluppati almeno altri due importanti approcci: l'approccio *frequentista* e l'approccio *classico*.¹¹

¹⁰D'altra parte, si può anche mostrare che ciascuno dei risultati (i) e (ii) del teorema 1.1 è logicamente equivalente all'assioma di numerabile additività.

¹¹Nell'ambito delle scienze fisiche ed economiche è abbastanza diffuso anche l'approccio "soggettivista", dovuto principalmente a Bruno de Finetti (si veda B. de Finetti, "Theory of Probability", Wiley, New York, 1974), secondo il quale non è possibile assegnare alla probabilità un significato ed un valore "oggettivo" (come avviene nell'approccio classico

1.5.1 Approccio frequentista

L'approccio *frequentista*, dovuto a R. E. von Mises (1883–1953), definisce la probabilità di un evento nel seguente modo: se un esperimento è ripetuto n volte e se l'evento A si verifica n_A volte, si definisce probabilità dell'evento A il limite della *frequenza di successo*:

$$P(A) \triangleq \lim_{n \rightarrow \infty} \frac{n_A}{n}. \quad (1.9)$$

L'approccio frequentista è un approccio *induttivo*, cioè un approccio che si basa (o vorrebbe basarsi) sull'esperienza, e presenta il vantaggio innegabile di essere vicino al nostro concetto intuitivo di probabilità; tuttavia non è del tutto soddisfacente per fornire una definizione operativa di probabilità, perché non è possibile ripetere un esperimento un numero *infinito* di volte. Inoltre, dal punto di vista matematico, l'esistenza stessa del limite nella (1.9) può essere messa in discussione.

È interessante tuttavia interpretare gli assiomi di Kolmogorov in senso frequentista, visto che in molti casi la nostra interpretazione intuitiva della probabilità è vicina a quella frequentista. Infatti, se interpretiamo la probabilità come *frequenza di successo* dell'evento A su n prove, cioè trascuriamo l'operazione di limite nella (1.9), ponendo

$$P(A) \triangleq \frac{n_A}{n},$$

si ha:

- I. $P(A) \geq 0$, banalmente perché $n_A \geq 0$ ed $n > 0$;
- II. $P(\Omega) = 1$, perché $n_\Omega = n$ (l'evento certo si verifica ad ogni prova);
- III. se $A \cap B = \emptyset$, allora $n_{A \cup B} = n_A + n_B$ perché non possono verificarsi entrambi simultaneamente. Allora:

$$P(A \cup B) = \frac{n_{A \cup B}}{n} = \frac{n_A}{n} + \frac{n_B}{n} = P(A) + P(B).$$

Si noti che abbiamo scritto il terzo assioma per semplicità nella forma finita, per evitare l'astrazione insita nel considerare infiniti eventi.

1.5.2 Approccio classico

Nell'approccio classico o *laplaciano*, dovuto per l'appunto a P. S. Laplace (1749–1827), la probabilità di un evento si definisce a priori come il rapporto

$$P(A) \triangleq \frac{N_A}{N}, \quad (1.10)$$

dove N è il numero (supposto finito) dei *possibili* risultati dell'esperimento, ed N_A è il numero dei risultati *favorevoli* all'evento A . In pratica, utilizzando la simbologia dell'approccio assiomatico, risulta $N = \text{card}(\Omega)$ e $N_A = \text{card}(A)$, per cui

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

e frequentista, e parzialmente in quello assiomatico), ma la probabilità stessa esprime piuttosto il grado di aspettativa soggettiva di un individuo relativamente al verificarsi di un evento. Tale approccio, sebbene ulteriormente sviluppato da de Finetti e dai suoi discepoli in modo da garantire un'assegnazione "coerente" delle probabilità, è stato guardato spesso con sospetto dalla comunità scientifica per la sua natura, appunto, soggettiva.

► **Esempio 1.8.** Consideriamo il lancio di un dado, per il quale $\Omega = \{1, 2, 3, 4, 5, 6\}$. Sia poi $A = \{\text{pari}\} = \{2, 4, 6\}$. Sulla base dell'approccio classico, risulta $N_A = \text{card}(A) = 3$, $N = \text{card}(\Omega) = 6$, per cui $P(A) = \frac{\text{card}(A)}{\text{card}(\Omega)} = \frac{1}{2}$. ◀

L'approccio classico è anch'esso, come quello assiomatico, di tipo *deduttivo*, cioè si fonda su una definizione a priori di probabilità, data dalla (1.10). Inoltre è facile verificare che le leggi di probabilità costruite a partire dalla definizione classica soddisfano gli assiomi di Kolmogorov. A prima vista, allora, l'approccio classico pare più soddisfacente dell'approccio assiomatico, in quanto mediante esso è possibile assegnare dei precisi *valori* alle probabilità, sulla base della (1.10). Tuttavia, i limiti insiti nell'uso di tale approccio appaiono chiari se ragioniamo più approfonditamente sull'esempio precedente. Infatti, il valore di probabilità dell'esempio precedente è "corretto" a patto che si assuma che il dado non sia truccato. E se viceversa assumessi il dado truccato? Secondo l'approccio classico, otterrei esattamente lo stesso valore di probabilità, il che ovviamente non è il risultato corretto, ed evidenzia la più seria limitazione di tale approccio. Potrei modificare la definizione classica richiedendo che i risultati da considerare nella (1.10) siano *equiprobabili*, ma in questo modo userei il concetto di "equiprobabilità" per definire il concetto di "probabilità", cioè ricadrei in un circolo vizioso o tautologico. Infine, non è chiaro come estendere la (1.10) al caso di un esperimento con *infiniti* risultati.

1.5.3 Vantaggi (e svantaggi) dell'approccio assiomatico

Tra i tre approcci considerati (assiomatico, frequentista, classico), l'approccio assiomatico è senz'altro il più astratto, basandosi su concetti della teoria degli insiemi e sull'introduzione di una serie di assiomi cui deve soddisfare la definizione di probabilità. Nonostante il suo carattere poco intuitivo, esso è riconosciuto come l'unico approccio che consente di definire matematicamente la teoria della probabilità in maniera soddisfacente ed elegante, evitando una serie di incongruenze ed inconsistenze tipiche dell'approccio frequentista e di quello classico. In particolare, l'intera teoria viene ad assumere un carattere puramente *deduttivo*, discende cioè in maniera logica e rigorosa dagli assiomi della probabilità così come, ad esempio, la geometria euclidea discende dal postulato sul numero di rette parallele ad una retta data passanti per un punto. Per questo motivo, quello assiomatico è stato riconosciuto dai matematici come l'approccio più soddisfacente alla teoria della probabilità, e sarà quello considerato nella trattazione che segue.¹² Tuttavia anche gli approcci frequentista e classico presentano alcuni vantaggi. In sintesi, è possibile affermare che:

- l'approccio frequentista è il più vicino al nostro concetto intuitivo di probabilità, e spesso è d'aiuto per interpretare intuitivamente i risultati ottenuti;
- l'approccio classico può servire ad assegnare i valori di probabilità in molti casi pratici (es. giochi, scommesse, etc.), in cui i risultati possibili si possono ritenere *equiprobabili*;
- l'approccio assiomatico è il più soddisfacente dal punto di vista formale (matematico), ma non consente di fissare univocamente i valori numerici delle probabilità da assegnare agli eventi (incompletezza).

¹²Va osservato, peraltro, che i tre approcci considerati (ed anche quello soggettivista) differiscono "soltanto" nel modo in cui si definisce la probabilità e nella sua interpretazione, mentre le regole formali del calcolo delle probabilità restano esattamente le stesse.

1.6 Esempi di costruzione di spazi di probabilità

Per concludere questo primo capitolo, consideriamo alcuni esempi di spazi di probabilità; per semplicità di trattazione, considereremo prima il caso più semplice di *spazi di probabilità discreti* (ovvero con un numero finito o infinito numerabile di possibili risultati), e successivamente quello più astratto di *spazi di probabilità continui* (ovvero con un numero infinito non numerabile di risultati).

1.6.1 Spazi di probabilità discreti

Sia $\Omega = \{\omega_1, \omega_2, \dots, \omega_n, \dots\}$ un insieme discreto, vale a dire di cardinalità finita o infinita numerabile. In tal caso, è possibile scegliere come σ -campo la collezione delle parti di Ω :

$$\mathcal{S} = \mathcal{P}(\Omega) = \{\text{tutti i sottoinsiemi di } \Omega, \Omega \text{ e } \emptyset \text{ inclusi}\}.$$

Osserviamo che, poiché Ω è finito o numerabile, qualunque evento A appartenente ad \mathcal{S} si può esprimere come *unione al più numerabile di eventi elementari* $\{\omega_i\}$, cioè

$$A = \bigcup_{i \in I_A} \{\omega_i\}.$$

dove $I_A \subseteq \mathbb{N}$ è l'insieme degli indici che identificano gli elementi appartenenti ad A . Poiché gli eventi elementari $\{\omega_i\}$ sono mutuamente esclusivi, allora si ha, per l'assioma III (numerabile additività):

$$P(A) = \sum_{i \in I_A} P(\{\omega_i\}).$$

Pertanto, per assegnare la probabilità di un qualunque evento A , è sufficiente assegnare le probabilità degli eventi elementari $p_i \triangleq P(\{\omega_i\})$, $\forall \omega_i \in \Omega$, garantendo che, per l'assioma II (normalizzazione), si abbia

$$P(\Omega) = \sum_{i=1}^{\infty} P(\{\omega_i\}) = \sum_{i=1}^{\infty} p_i = 1. \quad (1.11)$$

Consideriamo il caso di un insieme Ω di cardinalità finita ($\text{card}(\Omega) = N$): se è possibile assumere gli eventi elementari *equiprobabili* (per motivi di simmetria o applicando il cosiddetto *principio di ragione insufficiente*¹³) risulta necessariamente, per la (1.11),

$$p_i = \frac{1}{N} = \frac{1}{\text{card}(\Omega)}$$

ed inoltre si ha, evidentemente,

$$P(A) = \sum_{i \in I_A} \frac{1}{N} = \frac{\text{card}(A)}{\text{card}(\Omega)}. \quad (1.12)$$

Tale risultato è esattamente equivalente alla definizione (1.10) di probabilità secondo l'approccio classico, che quindi può riguardarsi come l'applicazione dell'approccio assiomatico a spazi campione Ω finiti con eventi elementari equiprobabili, un caso tipico della teoria dei giochi e delle scommesse. Osserviamo inoltre esplicitamente che determinare la probabilità di un evento A secondo la (1.12) equivale a *contare* gli elementi di A e quelli di Ω . Evidentemente, se $\text{card}(\Omega) = \infty$

¹³Tale principio, noto anche come "rasoio di Occam", dal nome del filosofo inglese William of Ockham (1280-1349) che lo formulò, stabilisce che, se si deve scegliere tra diverse ipotesi riguardanti un fenomeno, bisogna scegliere *la più semplice*.

non è possibile assumere gli eventi equiprobabili, in quanto avrei $P(\Omega) = \infty$ dalla (1.11) in tal caso!

In definitiva, la (1.12) mostra che in molti casi il calcolo delle probabilità di eventi si riduce ad un problema puramente *combinatorio*, consistente cioè nel *contare* gli elementi di un insieme, problema semplice in linea di principio, ma la cui applicazione a casi reali può giungere a notevoli livelli di complessità. I principali risultati del calcolo combinatorio sono riportati in Appendice B; in particolare, le formule di conteggio più frequentemente utilizzate sono raccolte in Tab. B.1.

► *Esempio 1.9.* Consideriamo il lancio di una moneta. In tal caso lo spazio campione è $\Omega = \{T, C\}$, e come σ -campo \mathcal{S} è possibile scegliere la classe $\mathcal{P}(\Omega)$ di tutti i possibili sottoinsiemi di Ω (in numero pari a $2^2 = 4$). Per assegnare la legge di probabilità, basta assegnare la probabilità degli eventi elementari $\{T\}$ e $\{C\}$. Per simmetria, poniamo:

$$P(\{T\}) = P(\{C\}) = 1/2$$

e le probabilità di tutti gli altri eventi in \mathcal{S} si ricavano da queste. ◀

► *Esempio 1.10.* Consideriamo il lancio di un dado. Lo spazio campione è $\Omega = \{1, 2, 3, 4, 5, 6\}$, e come σ -campo \mathcal{S} è possibile scegliere la classe $\mathcal{P}(\Omega)$ di tutti i possibili sottoinsiemi di Ω (in numero pari a $2^6 = 64$). Per assegnare la legge di probabilità, basta assegnare la probabilità degli eventi elementari $\{1\}, \{2\}, \{3\}, \{4\}, \{5\}, \{6\}$. Per simmetria, poniamo:

$$P(\{1\}) = P(\{2\}) = \dots = P(\{6\}) = 1/6$$

e le probabilità di tutti gli altri eventi in \mathcal{S} si ricavano da queste. ◀

► *Esempio 1.11.* Consideriamo il lancio di due monete uguali, o di una moneta due volte. In tal caso, lo spazio campione è $\Omega = \{TT, TC, CT, CC\}$, e come σ -campo \mathcal{S} è possibile scegliere la classe $\mathcal{P}(\Omega)$ di tutti i possibili sottoinsiemi di Ω (in numero pari a $2^4 = 16$). Osserviamo che l'evento

$$A = \{\text{esce testa al primo lancio}\}$$

non è un evento elementare. Infatti:

$$A = \{TT, TC\} = \{TT\} \cup \{TC\}.$$

Per assegnare la legge di probabilità, basta associare un valore di probabilità a ciascuno degli eventi elementari $\{TT\}, \{TC\}, \{CT\}, \{CC\}$. Per simmetria, poniamo:

$$P(\{TT\}) = P(\{TC\}) = P(\{CT\}) = P(\{CC\}) = 1/4$$

e le probabilità di tutti gli altri eventi in \mathcal{S} si ricavano da queste. Ad esempio, per l'evento A definito precedentemente, si ha:

$$P(A) = P(\{TT\}) + P(\{TC\}) = 1/4 + 1/4 = 1/2$$

perché $\{TT\} \cap \{TC\} = \emptyset$ (gli eventi elementari sono sempre mutuamente esclusivi) e per l'assioma III di Kolmogorov. ◀

In sintesi, se Ω è uno spazio discreto (finito o infinito numerabile) è possibile scegliere come σ -campo la classe $\mathcal{P}(\Omega)$ delle parti di Ω , ed assegnare la legge di probabilità definendo le probabilità p_i degli eventi elementari $\{\omega_i\}$; in particolare, se Ω è finito con N elementi, è possibile assumere i risultati equiprobabili e quindi $p_i = 1/N$; tale scelta non è legittima se Ω è infinito.

► *Esempio 1.12.* Sebbene nella maggior parte dei problemi riguardanti spazi discreti si consideri $\mathcal{S} = \mathcal{P}(\Omega)$, non bisogna pensare che questa sia l'unica scelta possibile. Ad esempio, con riferimento a $\Omega = \{1, 2, 3, 4, 5, 6\}$, se un giocatore intende scommettere solo su $A = \{\text{pari}\}$ o $\bar{A} = \{\text{dispari}\}$, allora una scelta più opportuna sarà $\mathcal{S} = \{\emptyset, A, \bar{A}, \Omega\}$; si può verificare che questo è un σ -campo, anzi è il più piccolo σ -campo contenente A , e prende il nome di σ -campo *generato* da A . In questo caso si ottiene una semplificazione notevole nella descrizione probabilistica dell'esperimento. ◀

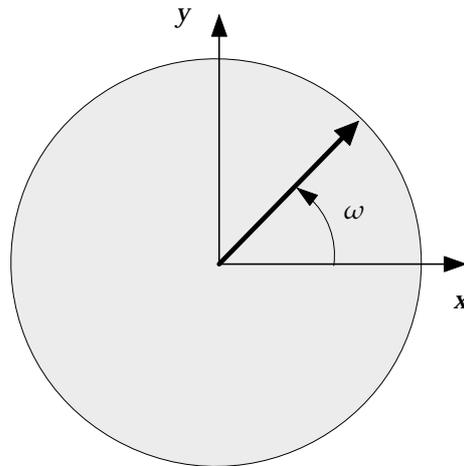


Fig. 1.7. Lancetta ruotante degli esempi 1.13 e 1.14.

1.6.2 Spazi di probabilità continui★

Lo spazio campione Ω si dice *continuo* se ha una cardinalità infinita non numerabile, ovvero se ha infiniti elementi, che però non si possono mettere in relazione biunivoca con l'insieme \mathbb{N} dei numeri naturali. Esempi di spazi campione continui sono $\Omega = \mathbb{R}$, $\Omega = (a, b) \subseteq \mathbb{R}$, $\Omega = \mathbb{R}^2$, $\Omega = \mathbb{R}^3$.

► *Esempio 1.13 (lancetta ruotante).* Un semplice esempio di esperimento aleatorio i cui risultati si possano considerare, con un piccolo sforzo di astrazione, appartenenti ad uno spazio campione continuo è raffigurato in Fig. 1.7. Una lancetta sottile (idealmente filiforme) è messa in rotazione su un piano orizzontale e si ferma in una posizione arbitraria. Tale posizione può essere individuata univocamente introducendo un sistema di riferimento cartesiano con origine nell'estremità fissa della lancetta e misurando la posizione della lancetta con l'angolo ω formato da quest'ultima con l'asse delle ascisse. Pertanto, lo spazio campione associato ad un tale esperimento sarà $\Omega = [0, 2\pi[$. Eventi di interesse potranno essere allora sottoinsiemi di Ω , del tipo:

$$A_1 = [0, \pi/2] = \{\text{la lancetta si ferma nel primo quadrante}\}$$

$$A_2 = [\pi, 2\pi[= \{\text{la lancetta si ferma nel terzo o nel quarto quadrante}\}$$

$$A_3 = \{\pi/4\} = \{\text{la lancetta si ferma con un angolo di } 45^\circ \text{ rispetto all'asse delle ascisse}\}$$

Dovremo poi assegnare una legge di probabilità che consenta di definire la probabilità di tali eventi e di tutti gli eventi di interesse. Per fare ciò, tuttavia, dovremo prima individuare la classe di tutti gli eventi di interesse, ovvero il σ -campo \mathcal{S} . ◀

Se Ω è continuo,¹⁴ non è possibile scegliere come σ -campo \mathcal{S} la classe $\mathcal{P}(\Omega)$ delle parti di Ω , cioè la classe di tutti i possibili sottoinsiemi di Ω . Abbiamo già osservato che $\mathcal{P}(\Omega)$ è senz'altro un σ -campo, anzi è il σ -campo più grande che è possibile concepire, ma si può dimostrare che è impossibile costruire una valida legge di probabilità (che soddisfi gli assiomi di Kolmogorov) su di esso. L'approccio corretto è invece scegliere \mathcal{S} come il *più piccolo* σ -campo che contiene tutti gli

¹⁴Osserviamo che per definire rigorosamente leggi di probabilità su spazi continui sono necessari concetti di *teoria degli spazi con misura* e nel caso di $\Omega \subseteq \mathbb{R}^k$ i concetti della *misura secondo Lebesgue* in \mathbb{R}^k . Per una trattazione rigorosa di tali concetti si veda [7].

insiemi *aperti* di Ω .¹⁵ Gli insiemi che appartengono a tale σ -campo si dicono gli *insiemi di Borel* (o *borelliani*) di Ω .

In pratica considereremo solo spazi continui che sono sottoinsiemi dello spazio euclideo \mathbb{R}^k ; in particolare, se $\Omega \subseteq \mathbb{R}$, denoteremo con x il generico elemento ω di Ω ; se $\Omega \subseteq \mathbb{R}^2$, denoteremo con (x, y) il generico elemento ω di Ω , e così via.

Se $\Omega \subseteq \mathbb{R}$, allora, è possibile definire \mathcal{S} come il più piccolo σ -campo che contiene tutti gli *intervalli aperti* $]a, b[$ di Ω . Si può facilmente verificare che complementando, unendo ed intersecando un'infinità *numerabile* di intervalli di questo tipo, si ottengono tutti i tipi di intervalli $[a, b]$, $[a, b[$, $]a, b]$, $]a, b[$, così come i punti isolati $\{a\}$, e tutti i loro complementi, unioni e intersezioni (tutti questi insiemi costituiscono la classe degli insiemi di Borel in \mathbb{R}). Tuttavia è possibile (anche se non immediato) costruire sottoinsiemi di Ω che non stanno in \mathcal{S} , e quindi \mathcal{S} *non* contiene tutti i sottoinsiemi di Ω , ovvero $\mathcal{S} \subset \mathcal{P}(\Omega)$. Senza essere eccessivamente formali, tuttavia, potremo assumere che tutti i sottoinsiemi di \mathbb{R} che si utilizzano nella pratica appartengano a \mathcal{S} , siano cioè insiemi di Borel.

Una volta determinato il σ -campo, ci rendiamo conto che non è possibile procedere come abbiamo fatto nel caso discreto, ovvero assegnando le probabilità degli eventi elementari $\{x\}$. In questo caso, infatti, utilizzando l'assioma di numerabile additività, riusciremmo a definire la probabilità *solo* di sottoinsiemi numerabili di Ω ; invece, non potremmo mai definire in questo modo la probabilità di eventi del tipo (a, b) .

Dobbiamo allora procedere in maniera alternativa. Una possibile strada è quella di considerare una funzione reale $f(x) \geq 0$ tale che

$$\int_{\Omega} f(x) dx = 1 \quad (1.13)$$

e porre, per ogni $A \in \mathcal{S}$,

$$P(A) = P(\{x \in A\}) \triangleq \int_A f(x) dx, \quad (1.14)$$

dove si assume che l'integrale esista finito per ogni $A \in \mathcal{S}$. Si può facilmente osservare che la (1.14) definisce una funzione da \mathcal{S} a \mathbb{R} che rispetta gli assiomi di Kolmogorov, ed è quindi una valida legge di probabilità. Infatti, $P(A) \geq 0$ perché $f(x) \geq 0$ (assioma I); $P(\Omega) = \int_{\Omega} f(x) dx = 1$ per la (1.13) (assioma II); infine, se A e B sono insiemi disgiunti, si ha $P(A \cup B) = P(A) + P(B)$ per l'additività dell'integrale (assioma III nella forma finita).¹⁶

► *Esempio 1.14.* Ad esempio, per la lancetta rotante dell'esempio 1.13, potremo scegliere una funzione $f(x)$ così definita:

$$f(x) = \begin{cases} \frac{1}{2\pi}, & \text{se } x \in [0, 2\pi]; \\ 0, & \text{altrimenti.} \end{cases}$$

Come si vede, tale funzione è non negativa e soddisfa alla condizione di normalizzazione (1.13): tale legge di probabilità si dice *uniforme* nell'intervallo $[0, 2\pi[$. A questo punto, la probabilità che la lancetta si fermi in qualunque intervallo angolare $A = [\theta_1, \theta_2] \subseteq [0, 2\pi[= \Omega$ è:

$$P(A) = \frac{1}{2\pi} \int_{\theta_1}^{\theta_2} dx = \frac{\theta_2 - \theta_1}{2\pi}.$$

¹⁵Nello spazio $\Omega \subseteq \mathbb{R}$, un insieme A si dice aperto se per un qualunque $x \in A$ esiste un intervallo aperto $A_x =]a, b[$ tale che $x \in A_x \subset A$. In uno spazio astratto Ω qualsiasi, per definire un insieme aperto occorre definire una *topologia* su Ω .

¹⁶A voler essere precisi, bisogna dire che non tutte le leggi di probabilità su $\Omega \subseteq \mathbb{R}$ possono essere espresse nella forma (1.14), a meno di non ricorrere a funzioni $f(x)$ particolari (distribuzioni).

Utilizzando tale formulazione, è semplice calcolare la probabilità degli eventi A_1 , A_2 e A_3 definiti nell'esempio 1.13. Si ha:

$$\begin{aligned} P(A_1) &= \frac{\pi/2}{2\pi} = \frac{1}{4} \\ P(A_2) &= \frac{\pi}{2\pi} = \frac{1}{2} \\ P(A_3) &= \frac{0}{2\pi} = 0 \end{aligned}$$

I primi due risultati sono in accordo con la nostra intuizione, mentre l'ultimo risultato appare sorprendente: *la probabilità che la lancetta si fermi in una precisa posizione angolare è zero!* ◀

Come osservato nell'esempio precedente, definire la legge di probabilità mediante la (1.14) ha delle conseguenze apparentemente sorprendenti per la probabilità degli eventi elementari. Infatti, nell'ipotesi in cui $f(x)$ è limitata, si trova $P(\{x\}) = 0$, e quindi *tutti gli eventi elementari hanno probabilità nulla*.

Prova. La dimostrazione rigorosa sfrutta la proprietà di continuità della probabilità (cfr. § 1.4.5). Per calcolare la probabilità dell'evento $A = \{x\}$, possiamo costruire una successione decrescente di eventi $A_n = \{x \leq u \leq x + 1/n\}$ tale che, evidentemente, $\bigcap_{n=1}^{\infty} A_n = A$. Per la continuità della probabilità, si ha allora:

$$P(\{x\}) = P(A) = \lim_n P(A_n) = \lim_n \left(\int_{A_n} f(u) du \right) = \lim_n \left(\int_x^{x+1/n} f(u) du \right).$$

Ma se $f(x)$ è una funzione limitata ($|f(x)| \leq M, \forall x \in \overline{\mathbb{R}}$), si ha:

$$\left| \int_x^{x+1/n} f(u) du \right| \leq \int_x^{x+1/n} |f(u)| du \leq \frac{M}{n},$$

per cui

$$\lim_n \left(\int_x^{x+1/n} f(u) du \right) = 0$$

da cui l'asserto. ◻

Il risultato che gli eventi elementari $\{x\}$ abbiano probabilità nulla, sebbene possa apparire a prima vista sorprendente, non è in contrasto con l'assioma di normalizzazione ($P(\Omega) = 1$), nè con quello di numerabile additività. Infatti, nel caso continuo risulta $\Omega = \bigcup_{x \in \Omega} \{x\}$, ovvero Ω è esprimibile come l'unione degli eventi elementari disgiunti, ma tale unione *non* è numerabile, e quindi non è applicabile il terzo assioma (che restituirebbe un paradossale $P(\Omega) = 0$). In questo caso, allora, può evidentemente risultare $P(\Omega) = 1$ anche se gli eventi elementari hanno probabilità zero. D'altra parte, l'apparente paradosso nasce dal fatto che specificare l'evento elementare $\{x\}$ significa idealmente assegnare un numero reale x con *tutte* le cifre significative; nella pratica questo è impossibile, e ci limitiamo a fornire la rappresentazione di x solo fino alla K -esima cifra significativa, per cui quello che consideriamo un "numero reale approssimato" è in realtà l'insieme (continuo) dei numeri reali la cui rappresentazione fino alla K -esima cifra significativa coincide con quella assegnata. Ad esempio, l'approssimazione $x = 3.14$ di π rappresenta in realtà qualunque numero reale compreso tra $3.140000 \dots$ e $3.149999 \dots$, ovvero l'intervallo di valori $[3.14, 3.15[$. Pertanto, nella pratica non è possibile considerare veri e propri eventi elementari, ma solo intervalli di \mathbb{R} , la cui probabilità, calcolata sulla base della (1.14), è generalmente diversa da zero.

Per completare il nostro ragionamento, resta da approfondire l'interpretazione da dare alla funzione $f(x)$. Se $f(x)$ è continua, consideriamo l'evento $A = \{x \leq u \leq x + \Delta x\}$ ed applichiamo

il teorema della media del calcolo integrale:

$$P(A) = \int_x^{x+\Delta x} f(u) du = f(x + \theta \Delta x) \Delta x \approx f(x) \Delta x$$

con $\theta \in [0, 1]$, da cui, dividendo per Δx e passando al limite per $\Delta x \rightarrow 0$, si ha:

$$f(x) = \lim_{\Delta x \rightarrow 0} \frac{P(A)}{\Delta x} = \lim_{\Delta x \rightarrow 0} \frac{P(\{x \leq u \leq x + \Delta x\})}{\Delta x},$$

e quindi la funzione $f(x)$ si può interpretare come una *densità di probabilità*. Notiamo che tale densità di probabilità è in genere diversa da zero, anche se la probabilità dell'evento elementare $\{x\}$ è nulla.

Nelle precedenti considerazioni, un punto non completamente chiaro è come scegliere la funzione $f(x)$. Anche qui emerge l'incompletezza dell'approccio assiomatico, ovvero ogni funzione $f(x) \geq 0$ che soddisfi la (1.13) definisce una valida legge di probabilità. Ma, se vogliamo invocare il principio di ragione insufficiente, qual è la scelta più "semplice" da fare? A prima vista, sembrerebbe che, in mancanza di altre informazioni, la scelta di una funzione $f(x)$ *costante* ovvero di una legge di probabilità *uniforme* (vedi esempio 1.14) sia la più naturale. Tuttavia, tale scelta non è lecita se Ω non è limitato, perché una funzione costante e positiva avrebbe integrale infinito su un insieme non limitato, e quindi non potrebbe soddisfare la condizione di normalizzazione (1.13). La scelta di una funzione costante è viceversa perfettamente legittima se Ω è limitato, ad esempio se $\Omega = [x_1, x_2]$, come già osservato nell'esempio 1.14 ed ulteriormente discusso nel seguente esempio.

► *Esempio 1.15.* Si consideri l'esperimento consistente nell'arrivo a caso di una telefonata ad una centrale telefonica nell'intervallo $[t_1, t_2]$. In tal caso, il risultato dell'esperimento è un numero reale $x \in [t_1, t_2]$, che rappresenta l'istante di arrivo della telefonata, per cui lo spazio campione è $\Omega = [t_1, t_2]$. Come σ -campo, tenendo conto dell'osservazione fatta nel precedente esempio, scegliamo il più piccolo σ -campo che contiene tutti gli intervalli aperti $]a, b[\subseteq [t_1, t_2]$. Come legge di probabilità, in mancanza di altre informazioni, scegliamo una funzione $f(x)$ così definita:

$$f(x) = \begin{cases} \alpha, & \text{se } x \in [t_1, t_2]; \\ 0, & \text{altrimenti.} \end{cases}$$

Tale $f(x)$ si dice *uniforme* in $[t_1, t_2]$. Imponiamo ora che la condizione di normalizzazione (1.13) sia soddisfatta:

$$\int_{t_1}^{t_2} f(x) dx = 1 \Rightarrow \alpha = \frac{1}{t_2 - t_1}.$$

In base a questa scelta della funzione $f(x)$, la probabilità che giunga una telefonata in un intervallo $A = [a, b] \subseteq \Omega$ è:

$$P(A) = \frac{1}{t_2 - t_1} \int_a^b dx = \frac{b - a}{t_2 - t_1}.$$

Osserviamo che, poichè $b - a$ è la *misura* dell'intervallo $[a, b]$, e $t_2 - t_1$ è la misura dell'intervallo $\Omega = [t_1, t_2]$, la probabilità $P(A)$ si può interpretare come una *misura normalizzata*:

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

Tale interpretazione della probabilità mostra chiaramente i legami della teoria della probabilità con la teoria della misura, e prende anche il nome di interpretazione *geometrica* della probabilità o semplicemente *probabilità geometrica*. ◀

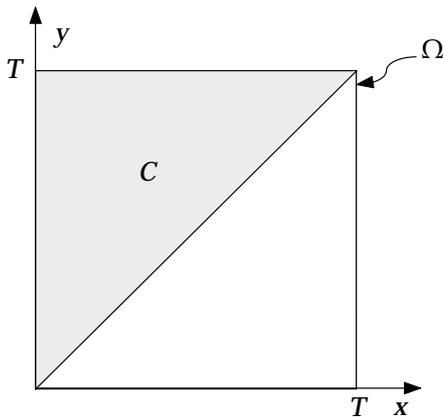


Fig. 1.8. Problema dell'incontro: $C = \{x \leq y\}$ rappresenta l'evento "Tizio arriva prima di Caio".

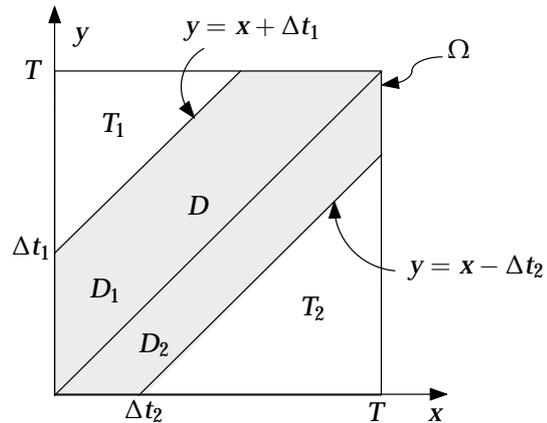


Fig. 1.9. Problema dell'incontro: $D = \{x - \Delta t_2 \leq y \leq x + \Delta t_1\}$ rappresenta l'evento "Tizio e Caio si incontrano".

I concetti introdotti per il caso $\Omega \subseteq \mathbb{R}$ possono essere estesi, senza grosse difficoltà concettuali, al caso più generale in cui $\Omega \subseteq \mathbb{R}^k$. Il caso $k = 2$ è discusso nel seguente esempio, con riferimento ad un problema concreto.

► **Esempio 1.16 (problema dell'incontro).** Un esempio di spazio di probabilità continuo su un sottoinsieme di \mathbb{R}^2 è il cosiddetto *problema dell'incontro*, una cui possibile formulazione è la seguente: due amici, Tizio e Caio, si recano, per caso e indipendentemente l'uno dall'altro, nello stesso bar nell'intervallo $[0, T]$, e ciascuno si trattiene per Δt_1 e Δt_2 secondi, rispettivamente.

Tale esperimento può essere descritto in termini probabilistici come segue. Il risultato dell'esperimento è una coppia ordinata di numeri (x, y) , con $x \in [0, T]$ e $y \in [0, T]$, dove x ed y rappresentano rispettivamente gli istanti di arrivo del primo e del secondo amico. Lo spazio campione è allora il quadrato $\Omega = [0, T] \times [0, T] \subset \mathbb{R}^2$. Come σ -campo, potremo scegliere il più piccolo σ -campo che contiene tutti i *rettangoli aperti* $A =]a, b[\times]c, d[$. Come legge di probabilità, infine, in analogia all'esempio 1.15, utilizzeremo la *misura normalizzata*, corrispondente a scegliere una densità di probabilità uniforme nel quadrato; se cioè A è un evento, ovvero è un sottoinsieme del quadrato appartenente ad \mathcal{S} , e se $\text{misura}(A)$ rappresenta la sua misura (un'area, in questo caso), allora porremo:

$$P(A) = \frac{\text{misura}(A)}{\text{misura}(\Omega)},$$

dove $\text{misura}(\Omega) = \text{misura}(\text{quadrato}) = T^2$. Ad esempio, la probabilità che $(x, y) \in A = [a, b] \times [c, d]$ è data da:

$$P(A) = \frac{(b-a)(d-c)}{T^2}.$$

Una volta individuato un corretto modello probabilistico, possiamo affrontare il calcolo della probabilità di un qualsiasi evento, e data la definizione della probabilità come misura normalizzata, il calcolo si può effettuare utilizzando semplici considerazioni geometriche.

Ad esempio, sia C il seguente evento: "Tizio arriva prima di Caio". In termini numerici, risulta evidentemente $C = \{x \leq y\}$, per cui l'evento C è il triangolo rappresentato in Fig. 1.8. Si ha allora:

$$P(C) = \frac{\text{misura}(C)}{\text{misura}(\Omega)} = \frac{T^2/2}{T^2} = \frac{1}{2}.$$

Calcoliamo adesso la probabilità dell'evento D definito come segue: "Tizio e Caio si incontrano". Evidentemente, ciò si verifica se:

- arriva prima Tizio, e risulta $y \leq x + \Delta t_1$; corrisponde al dominio $D_1 = \{x \leq y, y \leq x + \Delta t_1\}$ di Fig. 1.9; oppure:

- arriva prima Caio, e risulta $x \leq y + \Delta t_2$; corrisponde al dominio $D_2 = \{y \leq x, x \leq y + \Delta t_2\}$ di Fig. 1.9.

I domini D_1 e D_2 sono mutuamente esclusivi (se si escludono i punti sulla frontiera, che possiamo attribuire indifferentemente all'uno o all'altro, in quanto la frontiera ha misura nulla), e sono tali che $D = D_1 \cup D_2$. Pertanto si ha $P(D) = P(D_1) + P(D_2)$, e utilizzando semplici considerazioni possiamo ottenere l'area del dominio D per sottrazione, in quanto si ha:

$$\text{misura}(D) = \text{misura}(\Omega) - \text{misura}(T_1) - \text{misura}(T_2)$$

e le aree dei triangoli T_1 e T_2 sono:

$$\begin{aligned}\text{misura}(T_1) &= \frac{(T - \Delta t_1)^2}{2}, \\ \text{misura}(T_2) &= \frac{(T - \Delta t_2)^2}{2},\end{aligned}$$

da cui sostituendo si ha il risultato finale:

$$P(D) = \frac{\text{misura}(D)}{\text{misura}(\Omega)} = \frac{\Delta t_1 + \Delta t_2}{T} - \left(\frac{\Delta t_1^2 + \Delta t_2^2}{2 T^2} \right).$$

Ad esempio, se $T = 60$ (minuti) e $\Delta t_1 = \Delta t_2 = 5$ (minuti), si trova $P(D) \approx 0.139$. ◀

1.7 Esercizi proposti

Esercizio 1.1. Per ciascuno dei seguenti esperimenti, si descriva lo spazio campione:

- lanciare quattro volte una moneta bilanciata;
- individuare il numero di foglie danneggiate da un parassita su una pianta;
- misurare il tempo di vita (in ore) di una lampadina;
- misurare il peso di una cavia di laboratorio;
- controllare il numero di componenti difettosi in un lotto di componenti elettronici.

Esercizio 1.2. Sia $\Omega = \{1, 2, 3, 4, 5, 6\}$ uno spazio campione, verificare se le seguenti collezioni di insiemi sono σ -campi:

$$\begin{aligned} \mathcal{S}_1 &= \{\emptyset, \text{pari, dispari}, \Omega\}; \\ \mathcal{S}_2 &= \{\emptyset, \{1\}, \{3\}, \{1, 3\}, \Omega\}; \\ \mathcal{S}_3 &= \{\emptyset, \{1\}, \{2\}, \{2, 4\}, \Omega\}. \end{aligned}$$

Esercizio 1.3. Siano A e B due eventi tali che $P(A \cap B) = 1/4$, $P(\bar{A}) = 1/3$ e $P(B) = 1/2$. Calcolare la probabilità di $A \cup B$. [Risposta: $11/12$]

Esercizio 1.4. Siano A , B e C tre eventi di uno spazio di probabilità. Esprimere i seguenti eventi in termini di operazioni elementari sugli insiemi:

- si verificano almeno due dei tre eventi A , B , C ;
- si verificano esattamente due dei tre eventi A , B , C ;
- si verificano al più due dei tre eventi A , B , C ;
- si verifica esattamente uno dei tre eventi A , B , C .

Esercizio 1.5. Siano A e B due eventi di uno spazio di probabilità. Calcolare la probabilità dell'evento $A - B$ in termini di $P(A)$ e $P(A \cap B)$. [Risposta: $P(A - B) = P(A) - P(A \cap B)$]

Esercizio 1.6. Siano A e B due eventi di uno spazio di probabilità. Se definisce *differenza simmetrica* degli insiemi A e B l'insieme $A \Delta B$ contenente gli elementi di A oppure di B ma non di entrambi (corrisponde logicamente all'operazione di OR esclusivo). Calcolare la probabilità dell'evento $A \Delta B$ in termini di $P(A)$, $P(B)$ e $P(A \cap B)$. [Risposta: $P(A \Delta B) = P(A) + P(B) - 2P(A \cap B)$]

Esercizio 1.7. Siano A e B due eventi di uno spazio di probabilità. Esprimere i seguenti eventi in termini di operazioni elementari sugli insiemi e calcolarne le probabilità in termini di $P(A)$, $P(B)$ e $P(A \cap B)$:

- A oppure B oppure entrambi;
- almeno uno tra A e B ;
- A ma non B ;
- A oppure B ma non entrambi;
- al più uno tra A e B .

Esercizio 1.8. Siano A , B e C tre eventi di uno spazio di probabilità. Mostrare che

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

★ *Esercizio 1.9.*¹⁷ Giovanni e Maria seguono un corso di matematica, il cui esame finale prevede solo tre punteggi: A , B e C . La probabilità che Giovanni prenda B è pari a 0.3, la probabilità che Maria prenda B è pari a 0.4, la probabilità che nessuno dei due prenda A ma almeno uno dei due prenda B è pari a 0.1. Qual è la probabilità che almeno uno dei due prenda B ma nessuno prenda C ? [Risposta: 0.6]

¹⁷Gli esercizi contrassegnati con il simbolo ★ sono di maggiore difficoltà e non vanno affrontati per primi.

Esercizio 1.10. I risultati di un esperimento sono numeri interi equiprobabili tra 1 (incluso) e 12 (incluso). Si considerino i seguenti eventi:

$$\begin{aligned} A &= \{\text{il numero è dispari}\}; \\ B &= \{\text{il numero è divisibile per } 3\}; \\ C &= \{\text{il numero è divisibile per } 4\}. \end{aligned}$$

Individuare gli eventi A , B , C , $A \cap B$, $A \cap C$ e $\overline{A \cap B}$ e calcolarne le probabilità. [Risposta: $\frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \frac{1}{6}, 0, \frac{1}{6}$]

Esercizio 1.11. Si lanciano due dadi¹⁸. Calcolare la probabilità dei seguenti eventi:

$$\begin{aligned} A &= \{\text{la somma dei due dadi è maggiore o uguale a } 8\}; \\ B &= \{\text{la somma dei due dadi è esattamente uguale a } 8\}; \\ C &= \{\text{si ottiene almeno un } 6 \text{ nei due lanci}\}. \end{aligned}$$

[Risposta: $\frac{15}{36}, \frac{5}{36}, \frac{11}{36}$]

★ **Esercizio 1.12.** Nel 1600, alcuni giocatori chiesero a Galileo Galilei di spiegare perché, nel lancio di tre dadi, la somma 10 si presenti con maggior frequenza di 9, nonostante sia 10 che 9 si possano ottenere come somme di 6 terne distinte di interi tra 1 e 6.

- Formulare un appropriato modello probabilistico del problema.
- Calcolare la probabilità di ottenere 10 e 9 e verificare che effettivamente l'osservazione dei giocatori era fondata.¹⁹

Esercizio 1.13. Un dado è truccato in modo che la probabilità di ogni faccia sia proporzionale al numero di punti sulla faccia stessa (ad esempio, un "sei" è tre volte più probabile di un "due"). Calcolare la probabilità di ottenere un numero pari in un singolo lancio del dado.

Esercizio 1.14. Si lanciano due dadi. Siano A e B i seguenti eventi:

$$\begin{aligned} A &= \{\text{la somma dei due dadi è dispari}\}; \\ B &= \{\text{si ottiene almeno un } 6 \text{ nei due lanci}\}. \end{aligned}$$

Individuare gli eventi $A \cap B$, $A \cup B$, $A \cap \overline{B}$ e calcolarne le probabilità. [Risposta: $\frac{1}{6}, \frac{23}{36}, \frac{1}{3}$]

Esercizio 1.15. Si lanciano due dadi, e si denotano i risultati come d_1 ed d_2 . Qual è la probabilità che l'equazione di secondo grado $x^2 + x d_1 + d_2 = 0$ abbia radici reali? [Risposta: $\frac{19}{36}$]

Esercizio 1.16. Si considerino le cifre 1, 2, 3, 4, 5. L'esperimento è il seguente: si sceglie prima una cifra, e poi una seconda tra le restanti. Assumendo i 20 (perché?) possibili risultati dell'esperimento equiprobabili, determinare la probabilità che

- la prima volta venga scelta una cifra dispari;
- la seconda volta venga scelta una cifra dispari;
- entrambe le volte venga scelta una cifra dispari.

[Risposta: $\frac{3}{5}, \frac{3}{5}, \frac{3}{10}$]

Esercizio 1.17. Si estraggono simultaneamente due carte da un mazzo di carte francesi (senza jolly). Calcolare la probabilità di ottenere due assi. [Risposta: $\frac{1}{221} \approx 0.0045$]

Esercizio 1.18. Si estraggono simultaneamente due carte da un mazzo di carte francesi (senza jolly). Calcolare la probabilità che almeno una sia di cuori. [Risposta: $\frac{15}{34} \approx 0.441$]

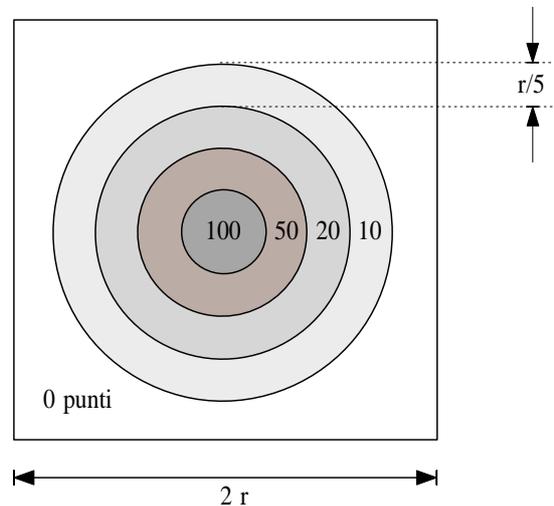
¹⁸In questo e negli esercizi che seguono, salvo avviso contrario, le monete ed i dadi sono bilanciati, i mazzi di carte sono ben mischiati, le estrazioni di numeri sono casuali.

¹⁹Galileo Galilei riportò le sue considerazioni in un trattato intitolato "Sopra le scoperte dei dadi" (Opere, vol.8).

★ *Esercizio 1.19.* Si estraggono in successione due carte da un mazzo di carte francesi (senza jolly). Calcolare la probabilità che la seconda carta estratta sia maggiore della prima. [Risposta: $\frac{16}{34} \approx 0.471$] [Suggerimento: osservare che $P(\text{uguale}) + P(\text{maggiore}) + P(\text{minore}) = 1$, che $P(\text{maggiore}) = P(\text{minore})$ per simmetria, e calcolare $P(\text{uguale})$]

Esercizio 1.20. La metropolitana arriva nella stazione di Campi Flegrei in un istante qualsiasi fra le 14.00 e le 14.30 e vi sosta T minuti. Uno studente, a sua volta, arriva nella stazione di Campi Flegrei in un istante qualsiasi dello stesso intervallo di tempo, indipendentemente dalla metropolitana. Quanto deve valere T affinché lo studente prenda la metropolitana con probabilità 0.8? [Risposta: $T \approx 11$ minuti]

Esercizio 1.21. Il gioco delle freccette consiste nel lanciare una freccetta su un bersaglio (vedi figura), ottenendo un punteggio corrispondente alla regione colpita. Il quadrato ha lato $2r$, e la distanza tra due cerchi concentrici adiacenti è pari a $r/5$. Determinare la probabilità di effettuare 100, 50, 20, 10, oppure 0 punti, lanciando una freccetta a caso (si supponga che la freccetta colpisca comunque il quadrato). [Risposta: $\frac{\pi}{100}, \frac{3\pi}{100}, \frac{5\pi}{100}, \frac{7\pi}{100}, 1 - \frac{16\pi}{100}$]



Probabilità condizionale e indipendenza

Si affrontano in questo capitolo due argomenti fondamentali della teoria della probabilità: la probabilità condizionale e l'indipendenza statistica tra eventi. Dopo aver definito il concetto di probabilità condizionale, si mostra che tutte le proprietà della probabilità possono essere applicate anche alla probabilità condizionale, e si introducono la legge della probabilità composta, il teorema della probabilità totale ed il teorema di Bayes, estremamente utili nella risoluzione di problemi pratici. Successivamente si introduce il concetto di indipendenza statistica, che viene applicato per semplificare la costruzione di spazi di probabilità su esperimenti combinati. I concetti studiati vengono infine applicati ad un modello semplificato di sistema di comunicazione, comprendente una sorgente di informazione, un canale di comunicazione ed una destinazione; in particolare, viene studiato in dettaglio il modello di canale binario simmetrico (BSC).

2.1 Introduzione

Nel precedente capitolo abbiamo introdotto i concetti basilari della teoria della probabilità, ed in particolare abbiamo visto come si definisce la probabilità di un evento A appartenente ad uno spazio di probabilità (Ω, \mathcal{S}, P) ; inoltre, a partire dalle probabilità di eventi semplici, abbiamo derivato delle semplici regole per il calcolo delle probabilità di eventi ottenuti come unione, complementazione e intersezione di più eventi semplici.

Tuttavia, anche nello studio delle relazioni che coinvolgono più eventi di uno spazio di probabilità, non abbiamo approfondito le relazioni di *dipendenza* (o di assenza di dipendenza, ovvero *indipendenza*) tra tali eventi. Lo studio di tali relazioni, affrontato nel corso di questo capitolo, consentirà di dare risposta a quesiti del tipo: se sappiamo che si è verificato l'evento B , come si modifica la probabilità dell'evento A ? Il punto di partenza delle nostre considerazioni sarà il concetto di *probabilità condizionale*.

2.2 Probabilità condizionale

Siano A e B due eventi di uno spazio di probabilità (Ω, \mathcal{S}, P) . Vogliamo mettere in relazione, in senso probabilistico, gli eventi A e B , introducendo una misura dell'incertezza residua su A sapendo che B si è verificato. Tale misura è fornita dalla cosiddetta *probabilità condizionale* di A "dato" B :

Definizione (probabilità condizionale). Sia (Ω, \mathcal{S}, P) uno spazio di probabilità, e siano $A, B \in \mathcal{S}$ due eventi, con $P(B) \neq 0$. La probabilità condizionale (o condizionata) di A dato B è:

$$P(A|B) = \frac{P(AB)}{P(B)}. \quad (2.1)$$

Ricordiamo che, nella (2.1), AB rappresenta l'intersezione $A \cap B$. Poiché $AB \subseteq B$, si ha che $P(AB) \leq P(B)$ e quindi $P(A|B) \leq 1$ (ovviamente $P(A|B)$ è sicuramente positiva perché rapporto di due quantità positive). Quindi $P(A|B) \in [0, 1]$ ed è lecito allora parlare di probabilità condizionale: vedremo che, di fatto, la (2.1) definisce una vera e propria legge di probabilità, nel senso che soddisfa agli assiomi di Kolmogorov.

► *Esempio 2.1.* In un lancio di un dado ben equilibrato, calcolare la probabilità che esca 6 sapendo che (dato che) è uscito un numero pari.

Definiamo come al solito lo spazio campione $\Omega = \{1, 2, 3, 4, 5, 6\}$, e gli eventi $A = \{6\}$, $B = \{\text{pari}\} = \{2, 4, 6\}$. Assumendo gli eventi elementari equiprobabili, si ha:

$$\begin{aligned} P(A) &= 1/6, \\ P(B) &= 3/6 = 1/2, \\ P(AB) &= P(A) = 1/6, \\ P(A|B) &= \frac{1/6}{1/2} = 1/3. \end{aligned}$$

Si noti che $P(A|B) > P(A)$, cioè è aumentata la probabilità che esca 6 se sappiamo che è uscito un numero pari, come è naturale. ◀

Il concetto espresso nel precedente esempio si può formalizzare come segue:

- se $P(A|B) > P(A)$, allora A è "attratto" da B ; sapere che B si è verificato fa aumentare la probabilità che si verifichi A ;
- se $P(A|B) < P(A)$, allora A è "respinto" da B ; sapere che B si è verificato fa diminuire la probabilità che si verifichi A ;
- se $P(A|B) = P(A)$, A è "indifferente" a B ; in questo caso vedremo nel seguito (cfr. § 2.3) che A e B sono *statisticamente indipendenti*.

► *Esempio 2.2.* Dati gli eventi $A = \{\text{oggi piove}\}$ e $B = \{\text{oggi è estate}\}$, si ha che $P(A|B) < P(A)$ cioè diminuisce la probabilità che oggi piova se so che è estate. Se viceversa $B = \{\text{oggi è inverno}\}$ ho che $P(A|B) > P(A)$ cioè aumenta la probabilità che oggi piova se so che è inverno. ◀

Dalla definizione (2.1), il lettore può facilmente dimostrare che:

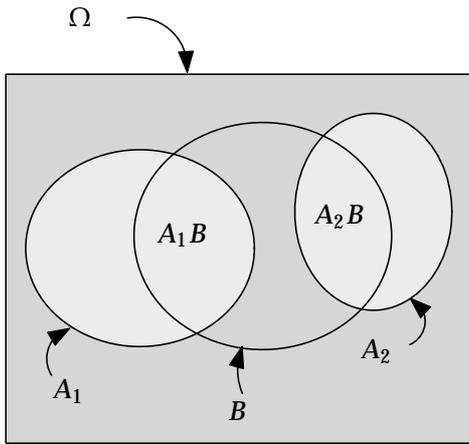


Fig. 2.1. Diagramma di Venn che mostra che se A_1 ed A_2 sono mutuamente esclusivi, anche A_1B ed A_2B sono mutuamente esclusivi.

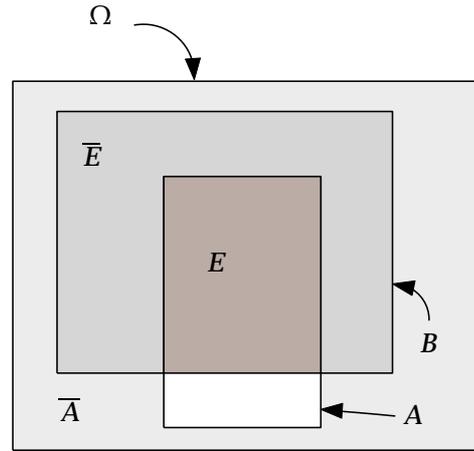


Fig. 2.2. Diagramma di Venn della relazione $\bar{E} = B - E = \bar{A} \cap B$, con $E = A \cap B$.

1. se $B \subseteq A$, allora $P(A|B) = 1$; infatti, poichè B implica A , se sappiamo che si è verificato B allora siamo certi che si sia verificato anche A ;
2. se $A \subseteq B$, allora $P(A|B) = \frac{P(A)}{P(B)} \geq P(A)$; in tal caso, è A ad implicare B (e non viceversa), e quindi se sappiamo che si è verificato B non possiamo affermare con certezza che si sia verificato anche A ; tuttavia, la probabilità $P(A|B)$ che A si sia verificato dato B è *non inferiore* alla probabilità $P(A)$ che A si sia verificato se non abbiamo nessuna informazione su B .

2.2.1 Interpretazioni della probabilità condizionale

E' interessante notare che, per un fissato B , la probabilità condizionale definisce una vera e propria legge di probabilità su \mathcal{S} , in quanto gli assiomi di Kolmogorov risultano soddisfatti. Pertanto, *tutti i risultati e le proprietà validi per le probabilità (ad esempio, le proprietà elementari del § 1.4.3) valgono ugualmente anche per le probabilità condizionali.*

Prova. Verifichiamo che, dato B con $P(B) \neq 0$, la legge $P(\cdot|B)$ soddisfa gli assiomi di Kolmogorov. Si ha:

- I. $P(A|B) \geq 0, \forall A \in \mathcal{S}$ banalmente;
- II. $P(\Omega|B) = \frac{P(\Omega \cap B)}{P(B)} = \frac{P(B)}{P(B)} = 1$;
- III. Se A_1 ed A_2 sono mutuamente esclusivi:

$$P(A_1 \cup A_2|B) = \frac{P[(A_1 \cup A_2)B]}{P(B)} = \frac{P(A_1B \cup A_2B)}{P(B)}.$$

Ma A_1B ed A_2B sono anch'essi mutuamente esclusivi (Fig. 2.1) per cui:

$$P(A_1 \cup A_2|B) = \frac{P(A_1B) + P(A_2B)}{P(B)} = P(A_1|B) + P(A_2|B).$$

Similmente vale per la numerabile additività.

Gli assiomi di Kolmogorov sono verificati e quindi la $P(\cdot|B)$ è una valida legge di probabilità. □

Quindi, fissata una legge di probabilità $P(\cdot)$ su Ω , il fatto che si sia verificato B ci induce a modificare tale legge nella $P(\cdot|B)$, per tenere conto di tale ulteriore informazione. In questo senso,

possiamo vedere la legge di probabilità condizionata come una sorta di “raffinamento” della legge di probabilità originaria.

Una interpretazione leggermente differente è quella secondo la quale la probabilità condizionale definisce una legge di probabilità P_B su *un nuovo* spazio campione $\Omega_B = B$, con eventi del tipo $E = A \cap B$ ($A \in \mathcal{S}$). Secondo questa interpretazione, se sappiamo che si è verificato B allora possiamo “restringere” il nostro spazio campione a B stesso, eliminando cioè tutti quei risultati che non appartengono a B , e che quindi certamente non si sono verificati. In questo senso, la legge di probabilità condizionata è una “restrizione” della legge di probabilità originaria al sottoinsieme B .

Prova. Consideriamo il nuovo spazio campione $\Omega_B = B$, come σ -campo l'insieme \mathcal{S}_B composto da tutti gli insiemi del tipo $E = A \cap B$, con $A \in \mathcal{S}$, e come legge di probabilità su Ω_B la P_B definita dalla seguente:

$$P_B(E) \triangleq \frac{P(AB)}{P(B)}.$$

Osserviamo preliminarmente che, nel nuovo spazio campione $\Omega_B = B$, il complemento di E va inteso rispetto a B , cioè si ha $\bar{E} = B - E$.

Il punto centrale è dimostrare che \mathcal{S}_B è effettivamente un σ -campo. Anzitutto, notiamo che \mathcal{S}_B non è vuoto, perchè $\emptyset \in \mathcal{S}_B$. Se poi $E \in \mathcal{S}_B$, allora $E = A \cap B$, $\bar{E} = B - E = B - A \cap B = \bar{A} \cap B$ (Fig. 2.2). Ma $\bar{A} \cap B \in \mathcal{S}_B$ perchè $\bar{A} \in \mathcal{S}$. Similmente, se $E_1, E_2 \in \mathcal{S}_B$, allora $E_1 = A_1 \cap B$ e $E_2 = A_2 \cap B$, per cui $E_1 \cup E_2 = (A_1 \cup A_2) \cap B \in \mathcal{S}_B$ perchè $A_1 \cup A_2 \in \mathcal{S}$. Allo stesso modo si prova anche la chiusura rispetto all'unione numerabile, per cui effettivamente \mathcal{S}_B è un σ -campo.

A questo punto è banale verificare che la P_B soddisfa gli assiomi di Kolmogorov:

- I. $P_B(E) \geq 0, \forall E \in \mathcal{S}_B$ banalmente;
- II. $P_B(\Omega_B) = \frac{P(B \cap B)}{P(B)} = \frac{P(B)}{P(B)} = 1$;
- III. Se E_1 ed E_2 sono mutuamente esclusivi:

$$P_B(E_1 \cup E_2) = \frac{P[(A_1 \cup A_2)B]}{P(B)} = \frac{P(A_1B \cup A_2B)}{P(B)}.$$

Ma A_1B ed A_2B sono mutuamente esclusivi per cui:

$$P(E_1 \cup E_2) = \frac{P(A_1B) + P(A_2B)}{P(B)} = P_B(E_1) + P_B(E_2).$$

Gli assiomi di Kolmogorov sono verificati e quindi la $P_B(\cdot)$ è una valida legge di probabilità su $\Omega_B = B$. \square

2.2.2 Legge della probabilità composta

Una conseguenza banale della definizione di probabilità condizionale è la cosiddetta *legge della probabilità composta*:

$$P(A \cap B) = P(A|B) P(B) = P(B|A) P(A). \quad (2.2)$$

A rigore, per ricavare tale legge dalla definizione (2.1), è necessario supporre che $P(A)$ e $P(B)$ siano non nulle. Tuttavia, si osservi che essa vale formalmente anche se $P(A)$ e/o $P(B)$ è zero, e quindi la probabilità condizionale non è ben definita. Infatti, in tal caso, dalla relazione $A \cap B \subseteq A$ e $A \cap B \subseteq B$ si ricava necessariamente $P(A \cap B) = 0$. Si osservi, infine, che la probabilità $P(A \cap B)$ viene comunemente denominata *probabilità congiunta* degli eventi A e B .

L'utilità della legge della probabilità composta è che essa consente di calcolare la probabilità dell'evento $A \cap B$ in tre passi: (i) si calcola prima la probabilità di A ; (ii) si calcola la probabilità di B dato A ; (iii) si moltiplicano i due valori di probabilità. Ovviamente, data la simmetria della legge, si possono scambiare i ruoli di A e B secondo convenienza.

► *Esempio 2.3.* Una scatola contiene 3 palle bianche (w_1, w_2, w_3) e 2 rosse (r_1, r_2). Si rimuovono due palle in successione. Qual è la probabilità che la prima sia bianca e la seconda rossa?

Lo spazio campionario è costituito da tutte le coppie *ordinate* di palle, che sono esattamente venti:

$$\Omega = \{w_1 w_2, w_1 w_3, w_1 r_1, w_1 r_2, w_2 w_1, w_2 w_3, \dots, r_1 r_2\}.$$

Infatti, la prima palla può essere scelta in 5 modi differenti; fissata la prima palla, la seconda può essere scelta in 4 modi differenti, per cui ho un totale di $5 \times 4 = 20$ differenti modi. Più formalmente, le disposizioni ordinate senza sostituzione di n elementi su k posti sono $n!/(n-k)! = n(n-1) \cdots (n-k+1)$ (cfr. Tab. B.1), e nel caso in questione $n = 3 + 2 = 5$ e $k = 2$, da cui il risultato.

L'evento $C = \{\text{prima palla bianca, seconda rossa}\}$ è costituito da 6 elementi:

$$C = \{w_1 r_1, w_2 r_1, w_3 r_1, w_1 r_2, w_2 r_2, w_3 r_2\}.$$

Se assumiamo gli eventi elementari equiprobabili e con probabilità pari ad $1/20$, allora $P(C) = \text{card}(C)/\text{card}(\Omega) = 6/20 = 3/10$.

Vediamo ora se possiamo applicare il concetto di probabilità condizionale per arrivare allo stesso risultato più semplicemente. Definiamo:

$$C = \{\text{prima palla bianca, seconda rossa}\} = \underbrace{\{\text{prima bianca}\}}_A \cap \underbrace{\{\text{seconda rossa}\}}_B$$

per cui, applicando la legge della probabilità composta, si ha:

$$P(C) = P(AB) = P(B|A) P(A).$$

Ora, evidentemente,

$$P(A) = P(\text{prima bianca}) = 3/5$$

e rimane da calcolare

$$P(B|A) = P(\text{seconda rossa}|\text{prima bianca}).$$

Se la prima palla estratta è bianca, rimangono nella scatola 4 palle, 2 bianche e 2 rosse, per cui $P(B|A) = 1/2$. Si ha allora:

$$P(C) = P(B|A) P(A) = \frac{3}{5} \cdot \frac{1}{2} = \frac{3}{10}.$$

Notiamo che seguendo il secondo approccio non è stato necessario determinare (contare) il numero di elementi di Ω . ◀

2.2.3 Regola della catena

È possibile estendere la definizione di probabilità condizionata anche al caso di più eventi condizionanti. Per esempio, si ha:

$$P(A|B, C) \triangleq \frac{P(ABC)}{P(BC)}, \quad P(BC) \neq 0.$$

Si noti che $P(A|B, C)$ è da intendersi come $P(A|BC)$, cioè si condiziona all'evento BC , ovvero al fatto che si sono verificati *congiuntamente* l'evento B e l'evento C . Riscrivendo la precedente, si trova allora una legge di fattorizzazione analoga alle legge della probabilità composta:

$$P(ABC) = P(A|B, C) P(BC)$$

e poichè, per la legge della probabilità composta, $P(BC) = P(B|C) P(C)$, si ottiene:

$$P(ABC) = P(A|B, C) P(B|C) P(C).$$

Applicando tale relazione iterativamente al caso di n eventi A_1, A_2, \dots, A_n , si ha la cosiddetta *regola della catena* per il calcolo della probabilità congiunta di n eventi:

$$P(A_1 A_2 \cdots A_n) = P(A_1) P(A_2|A_1) P(A_3|A_1, A_2) \cdots P(A_n|A_1, A_2, \dots, A_{n-1}).$$

La regola precedente si applica indipendentemente dall'ordine in cui si considerano gli eventi. In effetti, poichè esistono $n!$ distinte permutazioni degli eventi A_1, A_2, \dots, A_n , la fattorizzazione secondo la regola della catena può avvenire in $n!$ modi distinti.

2.2.4 Teorema della probabilità totale e teorema di Bayes

Due importanti proprietà della probabilità condizionale, che risultano estremamente utili nelle applicazioni, sono descritte dai due teoremi seguenti:

Teorema 2.1 (probabilità totale). Siano A_1, A_2, \dots, A_n eventi mutuamente esclusivi ($A_i \cap A_j = \emptyset, \forall i \neq j$) e sia $B \subseteq \bigcup_{i=1}^n A_i$. Si ha:

$$P(B) = \sum_{i=1}^n P(B|A_i)P(A_i).$$

Prova. Si faccia riferimento al diagramma di Venn in Fig. 2.3. Poichè $B \subseteq \bigcup_{i=1}^n A_i \Rightarrow B = B \cap \{\bigcup_{i=1}^n A_i\} \Rightarrow B = \bigcup_{i=1}^n (B \cap A_i)$. Ma se gli A_i sono mutuamente esclusivi, anche gli eventi $B \cap A_i$ lo sono. Allora per il III assioma si ha:

$$P(B) = \sum_{i=1}^n P(B \cap A_i).$$

Per la legge della probabilità composta, si ha:

$$P(B \cap A_i) = P(B|A_i) P(A_i)$$

valida anche se $P(A_i) = 0$. Sostituendo nella precedente si ha l'asserto. \square

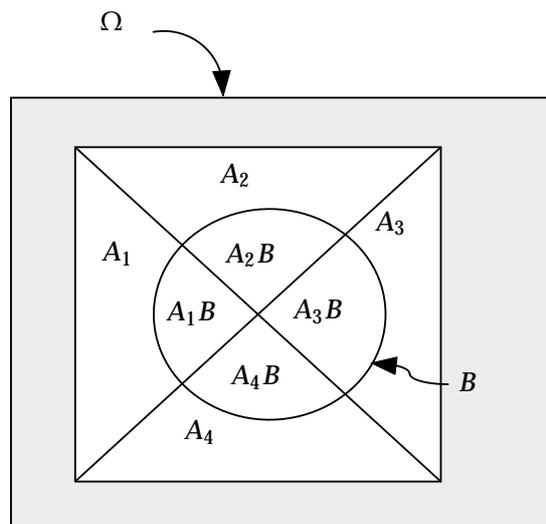


Fig. 2.3. Diagramma di Venn del teorema della probabilità totale ($n = 4$).

Nella pratica può essere complicato verificare la condizione $B \subseteq \bigcup_{i=1}^n A_i$, per cui spesso si assume che gli insiemi A_1, A_2, \dots, A_n , mutuamente esclusivi, costituiscano una *partizione* di Ω . In tal caso $\bigcup_{i=1}^n A_i = \Omega$ per cui la condizione precedente risulta senz'altro verificata.

Teorema 2.2 (Bayes). Siano A_1, A_2, \dots, A_n eventi mutuamente esclusivi ($A_i \cap A_j = \emptyset, \forall i \neq j$) e sia $B \subseteq \bigcup_{i=1}^n A_i$. Si ha:

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{\sum_{i=1}^n P(B|A_i)P(A_i)}.$$

Prova. Il teorema è una conseguenza banale della legge della probabilità composta e del teorema della probabilità totale. Infatti, per la legge della probabilità composta, si ha:

$$P(B \cap A_i) = P(B|A_i)P(A_i) = P(A_i|B)P(B)$$

per cui:

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

Sostituendo $P(B)$ come espresso dal teorema della probabilità totale nella precedente si ha l'asserto. \square

Il teorema di Bayes vale nelle stesse ipotesi del teorema della probabilità totale; inoltre entrambi i teoremi possono estendersi al caso in cui gli eventi A_i condizionanti siano un'infinità numerabile. Nel teorema di Bayes, la probabilità $P(A_i)$ è spesso definita *probabilità a priori*, mentre la $P(A_i|B)$ è definita *probabilità a posteriori*.

► *Esempio 2.4.* Si considerino 3 scatole che contengono componenti elettronici:

- nella scatola 1, ci sono 2000 componenti, di cui il 5% (100 componenti) difettosi;
- nella scatola 2, ci sono 1000 componenti, di cui il 40% (400 componenti) difettosi;
- nella scatola 3, ci sono 1000 componenti, di cui il 10% (100 componenti) difettosi;

Si seleziona una scatola a caso, e successivamente si rimuove un componente a caso dalla scatola.

1. Qual è la probabilità che il componente scelto sia difettoso?
2. Sapendo che un componente è difettoso, qual è la probabilità che provenga dalla scatola 2?

Lo spazio campione Ω associato a tale esperimento è costituito dai $2000 + 1000 + 1000 = 4000$ componenti, di cui 600 difettosi. Iniziamo con l'osservare che se i componenti fossero tutti in un'unica scatola e ne scegliessi uno a caso, la probabilità di scegliere un componente difettoso sarebbe pari a $\frac{600}{4000} = 0.15$. L'esperimento però è condotto in maniera differente: prima si sceglie la scatola, e successivamente il componente all'interno della scatola. Definiamo allora i seguenti eventi:

$$\begin{aligned} A_1 &= \{\text{il componente proviene dalla scatola 1}\}, \\ A_2 &= \{\text{il componente proviene dalla scatola 2}\}, \\ A_3 &= \{\text{il componente proviene dalla scatola 3}\}, \\ B &= \{\text{il componente è difettoso}\}. \end{aligned}$$

Poichè la scatola è scelta a caso, allora, per simmetria, si ha

$$P(A_1) = P(A_2) = P(A_3) = 1/3.$$

Inoltre, gli eventi A_1, A_2 e A_3 sono mutuamente esclusivi e $A_1 \cup A_2 \cup A_3 = \Omega$. Si ha poi, dai dati del problema,

$$\begin{aligned} P(B|A_1) &= 0.05, \\ P(B|A_2) &= 0.40, \\ P(B|A_3) &= 0.10. \end{aligned}$$

Poichè siamo nelle ipotesi del teorema della probabilità totale, possiamo scrivere:

$$P(B) = P(B|A_1) P(A_1) + P(B|A_2) P(A_2) + P(B|A_3) P(A_3) .$$

Sostituendo i valori numerici si trova $P(B) \approx 0.18$, che pertanto è la risposta al primo quesito. Per rispondere al secondo, possiamo applicare il teorema di Bayes:

$$P(A_2|B) = \frac{P(B|A_2) P(A_2)}{P(B)} \approx 0.73 .$$

Notiamo che la probabilità a posteriori $P(A_2|B)$ che il componente provenga dalla scatola 2, sapendo che è difettoso, è molto maggiore della probabilità a priori $P(A_2)$ che il componente provenga dalla stessa scatola, che è pari ad $1/3$. Questo è intuitivamente chiaro, perché la percentuale di componenti difettosi contenuti nella scatola 2 è maggiore di quella delle altre scatole. ◀

► *Esempio 2.5.* Si dispone di un test per individuare una malattia molto rara, che colpisce 1 paziente su 100 000. Il test è abbastanza affidabile: se il paziente ha la malattia, il test la individua con probabilità 0.95; se il paziente non ha la malattia, il test è falsamente positivo con probabilità 0.005. Se il test dice che la malattia è presente, qual è la probabilità che il paziente abbia effettivamente la malattia?

Lo spazio campione Ω associato a tale esperimento è costituito da tutti i pazienti sottoposti al test. Definiamo i tre eventi:

$$\begin{aligned} A_1 &= \{\text{il paziente ha la malattia}\} , \\ A_2 &= \{\text{il paziente non ha la malattia}\} , \\ B &= \{\text{il paziente è positivo al test}\} . \end{aligned}$$

Dobbiamo allora calcolare la probabilità:

$$P(A_1|B) = P(\text{il paziente ha la malattia}|\text{il test è positivo}) .$$

Poichè gli eventi A_1 ed A_2 sono mutuamente esclusivi, e $A_1 \cup A_2 = \Omega$, possiamo adoperare il teorema di Bayes, e scrivere:

$$P(A_1|B) = \frac{P(B|A_1)P(A_1)}{P(B|A_1)P(A_1) + P(B|A_2)P(A_2)} .$$

Per semplificare i calcoli, e fare alcune considerazioni sull'ordine di grandezza del risultato, possiamo riscrivere la precedente nella forma:

$$P(A_1|B) = \frac{1}{1 + \frac{P(B|A_2)P(A_2)}{P(B|A_1)P(A_1)}} .$$

Ora, con la notazione introdotta, si ha:

$$\begin{aligned} P(A_1) &= \frac{1}{100\,000} = 10^{-5} , \\ P(A_2) &= 1 - \frac{1}{100\,000} = 1 - 10^{-5} = 0.99999 , \\ P(B|A_1) &= 0.95 , \\ P(B|A_2) &= 0.005 . \end{aligned}$$

Se facciamo le approssimazioni $P(A_2) \approx 1$ e $P(B|A_1) \approx 1$, si ha:

$$P(A_1|B) \approx \frac{1}{1 + \frac{P(B|A_2)}{P(A_1)}} .$$

Poiché $P(B|A_2) \gg P(A_1)$, si ha che $P(A_1|B) \ll 1$. In effetti, sostituendo i valori numerici, si trova $P(A_1|B) \approx 2 \cdot 10^{-3}$; nonostante il test sembri abbastanza affidabile, la rarità della malattia lo rende praticamente inutile. Dalle considerazioni fatte, si comprende che per avere $P(A_1|B) \approx 1$ (un buon test) debba risultare $P(B|A_2) \ll P(A_1)$, cioè dovrei avere un test con probabilità di "falsa positività" molto più piccola della probabilità di occorrenza della malattia. Pertanto, si capisce perché effettuare uno *screening* di massa per individuare una malattia rara sia spesso considerato economicamente poco conveniente. ◀

2.3 Indipendenza tra eventi

Un concetto fondamentale nella teoria della probabilità è quello dell'indipendenza tra eventi, che può intuitivamente ricavarsi dal concetto di probabilità condizionale. Infatti, si considerino due eventi A e B : in base ad un ragionamento intuitivo, se gli eventi sono indipendenti, ci aspettiamo che sapere che B si sia verificato non altera in nessun modo la probabilità che si verifichi A . In formule, deve risultare:

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

Si noti che sostituendo tale espressione nella legge di probabilità composta (2.2) si ha:

$$P(AB) = P(A) P(B) \quad (2.3)$$

ed inoltre dalla definizione di probabilità condizionale (2.1) si ha pure

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

e quindi l'indipendenza è una proprietà simmetrica (se A è indipendente da B , anche B è indipendente da A). Peraltro, la (2.3) implica sia $P(A|B) = P(A)$ che $P(B|A) = P(B)$, per cui, sebbene meno intuitiva, si assume per simmetria proprio la (2.3) come definizione di indipendenza tra due eventi:

Definizione (indipendenza). Due eventi A e B si dicono indipendenti se

$$P(AB) = P(A) P(B).$$

Tale definizione afferma che la probabilità congiunta $P(AB)$ si fattorizza nel prodotto delle probabilità $P(A)$ e $P(B)$, che prendono il nome di *probabilità marginali*.

► *Esempio 2.6.* Consideriamo l'esempio, già visto, del lancio di due monete uguali, o di una moneta due volte. Lo spazio campione è $\Omega = \{TT, TC, CT, CC\}$, e abbiamo assunto gli eventi elementari equiprobabili e con probabilità pari ad $1/4$, per simmetria. Consideriamo ora gli eventi:

$$\begin{aligned} A &= \{\text{testa al primo lancio}\}, \\ B &= \{\text{testa al secondo lancio}\}, \end{aligned}$$

e verifichiamo che essi sono indipendenti. Si ha:

$$\begin{aligned} A &= \{TT, TC\}, \\ B &= \{CT, TT\}, \\ P(AB) &= P(\{TT\}) = 1/4, \\ P(A) &= 2/4 = 1/2, \\ P(B) &= 2/4 = 1/2, \end{aligned}$$

per cui $P(AB) = P(A) P(B)$, e quindi gli eventi sono indipendenti. Ragionando allo stesso modo, è facile provare che risultano statisticamente indipendenti tutti gli eventi del tipo $\{T/C \text{ al primo lancio}\}$ e $\{T/C \text{ al secondo lancio}\}$. ◀

Abbiamo già osservato che la definizione di indipendenza implica che $P(A|B) = P(A)$ e $P(B|A) = P(B)$. Inoltre, se A e B sono indipendenti, è facile provare che risultano indipendenti anche A e \bar{B} , \bar{A} e B , \bar{A} e \bar{B} .

Prova. Infatti, si ha:

$$\begin{aligned} P(A\bar{B}) &= P(\bar{B}|A) P(A) = [1 - P(B|A)] P(A) = [1 - P(B)] P(A) = P(\bar{B}) P(A); \\ P(\bar{A}B) &= P(\bar{A}|B) P(B) = [1 - P(A|B)] P(B) = [1 - P(A)] P(B) = P(\bar{A}) P(B); \\ P(\bar{A}\bar{B}) &= P(\bar{A}|\bar{B}) P(\bar{B}) = [1 - P(B|\bar{A})] P(\bar{A}) = [1 - P(B)] P(\bar{A}) = P(\bar{B}) P(\bar{A}), \end{aligned}$$

come volevasi dimostrare. \square

2.3.1 Indipendenza di tre o più eventi

Il concetto di indipendenza si può estendere a tre o più eventi, con qualche cautela:

Definizione (indipendenza di tre eventi). Gli eventi A , B e C si dicono indipendenti se:

1. sono indipendenti a coppie, cioè $P(AB) = P(A)P(B)$, $P(AC) = P(A)P(C)$, $P(BC) = P(B)P(C)$;
2. $P(ABC) = P(A)P(B)P(C)$.

Osserviamo esplicitamente che non è possibile assumere solo la seconda fattorizzazione come definizione di indipendenza, in quanto è possibile costruire esempi per i quali risulta $P(ABC) = P(A)P(B)P(C)$ mentre $P(AB) \neq P(A)P(B)$ etc.

Generalizzando il discorso al caso di n eventi, conviene definire separatamente il concetto di *indipendenza a coppie* e quello di *indipendenza*:

Definizione (indipendenza a coppie). Gli eventi $\{A_i\}_{i=1}^n$ si dicono indipendenti a coppie se

$$P(A_i A_j) = P(A_i)P(A_j), \quad \forall i \neq j.$$

Definizione (indipendenza di n eventi). Gli eventi $\{A_i\}_{i=1}^n$ si dicono indipendenti se

$$P\left(\bigcap_{i \in I} A_i\right) = \prod_{i \in I} P(A_i),$$

per ogni insieme I di indici diversi.

Le precedenti definizioni si possono estendere al caso di una infinità numerabile di eventi. In pratica, la definizione di indipendenza afferma che qualunque probabilità congiunta di un numero arbitrario di eventi A_i *distinti* si fattorizza nel prodotto delle corrispondenti probabilità marginali. Va osservato esplicitamente che l'indipendenza a coppie non implica necessariamente l'indipendenza, mentre l'indipendenza implica non solo l'indipendenza a coppie, ma anche a terne, a quaterne, etc.

► **Esempio 2.7.** Un esempio¹ di eventi indipendenti a coppie ma non indipendenti è il seguente: dato lo spazio campione $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}$, con gli eventi elementari equiprobabili, si considerino gli eventi:

$$A = \{\omega_1, \omega_2\}, \quad B = \{\omega_1, \omega_3\}, \quad C = \{\omega_1, \omega_4\}.$$

¹Tale esempio è attribuito al matematico S. N. Bernstein ed è menzionato nella originale monografia di Kolmogorov (cfr. nota pag. 7).

Si ha, con facili calcoli:

$$\begin{aligned} P(A) &= P(B) = P(C) = 1/2; \\ P(AB) &= P(BC) = P(AC) = 1/4 = (1/2)(1/2); \\ P(ABC) &= 1/4 \neq (1/2)(1/2)(1/2); \end{aligned}$$

per cui gli eventi sono indipendenti a coppie, ma non indipendenti. Per un altro esempio, si veda l'esercizio 2.17. ◀

► **Esempio 2.8 (eventi indipendenti ed eventi mutuamente esclusivi).** Due concetti talvolta confusi sono quelli di eventi indipendenti e di eventi mutuamente esclusivi. Mentre infatti l'indipendenza equivale alla fattorizzazione $P(AB) = P(A)P(B)$, due eventi si dicono mutuamente esclusivi se $AB = \emptyset$, per cui risulta necessariamente $P(AB) = 0$. Inoltre il concetto di eventi mutuamente esclusivi ha una chiara interpretazione sui diagrammi di Venn (gli insiemi A e B non si sovrappongono), mentre il concetto di eventi indipendenti no (se utilizziamo l'analogia in termini di aree normalizzate, la condizione di indipendenza si può esprimere dicendo che "l'area dell'intersezione AB è pari al prodotto delle aree di A e di B ", che non ha una chiara interpretazione sul diagramma di Venn, in quanto non si riconduce a relazioni di inclusione/esclusione). In definitiva, i due concetti non hanno alcuna relazione reciproca, salvo nel caso banale in cui $P(A) = 0$ oppure $P(B) = 0$. ◀

2.3.2 Indipendenza condizionale tra eventi

È possibile anche definire il concetto di *indipendenza condizionale* tra due eventi dato un terzo evento:

Definizione (indipendenza condizionale tra eventi). Due eventi A e B si dicono condizionalmente indipendenti, dato un terzo evento C , se

$$P(AB|C) = P(A|C)P(B|C).$$

Si noti che l'indipendenza condizionale *non* implica l'indipendenza di A e B , se non nel caso in cui $C = \Omega$. Allo stesso modo, per quanto meno intuitivamente comprensibile, l'indipendenza tra A e B non implica l'indipendenza condizionale rispetto ad un terzo evento C (si veda l'esercizio 2.18).

2.4 Esperimenti combinati

In molti casi interessa affrontare il seguente problema: dati più esperimenti aleatori, ognuno dei quali descritto in termini probabilistici, descrivere l'esperimento combinato, risultante dalla combinazione dei singoli esperimenti. Per far questo, è necessario costruire un *nuovo* spazio di probabilità, denominato *spazio di probabilità prodotto*, sull'esperimento combinato. Tale concetto è sviluppato nell'esempio seguente.

► **Esempio 2.9.** Supponiamo di avere due esperimenti aleatori, cui siano associati due spazi di probabilità $(\Omega_1, \mathcal{S}_1, P_1)$ e $(\Omega_2, \mathcal{S}_2, P_2)$. Per fissare le idee, si consideri come primo esperimento il lancio di una moneta, con spazio campione $\Omega_1 = \{T, C\}$, e come secondo esperimento il lancio di un dado, con spazio campione $\Omega_2 = \{1, 2, 3, 4, 5, 6\}$. Gli spazi di probabilità associati a ciascuno di tali esperimenti si costruiscono nel modo consueto (assumendo l'equiprobabilità degli eventi elementari).

Consideriamo adesso l'esperimento combinato (lancio di una moneta e di un dado), che ha come spazio campione il *prodotto cartesiano* di Ω_1 ed Ω_2 :

$$\Omega = \Omega_1 \times \Omega_2 = \{(T, 1), (T, 2), \dots, (T, 6), (C, 1), (C, 2), \dots, (C, 6)\}$$

costituito da $2 \times 6 = 12$ coppie ordinate. Poichè Ω è un insieme finito, possiamo considerare il σ -campo $\mathcal{S} = \mathcal{P}(\Omega)$ contenente tutti i 2^{12} sottoinsiemi di Ω . Notiamo che tale σ -campo \mathcal{S} conterrà, tra gli altri, eventi del tipo $A \times B$, con $A \in \mathcal{S}_1$ e $B \in \mathcal{S}_2$. Ad esempio, se $A = \{T\}$ e $B = \{\text{pari}\}$, si avrà $A \times B = \{(T, 2), (T, 4), (T, 6)\}$. Possiamo interpretare l'evento $A \times B$ nel modo seguente: si verifica l'evento $A \times B$ nell'esperimento combinato se si verifica l'evento A nell'esperimento 1 e l'evento B nell'esperimento 2. Tuttavia non tutti gli eventi di \mathcal{S} sono del tipo $A \times B$: si pensi ad esempio all'evento $C = \{(T, 1), (C, 2)\}$, che non può essere interpretato come $A \times B$.

A questo punto per completare la descrizione probabilistica dell'esperimento combinato resta da fissare la legge di probabilità P su \mathcal{S} . Osserviamo che si ha:

$$P(A \times \Omega_2) = P_1(A) \quad (2.4)$$

$$P(\Omega_1 \times B) = P_2(B) \quad (2.5)$$

dove P_1 è la legge di probabilità su Ω_1 e P_2 è la legge di probabilità su Ω_2 . Infatti $A \times \Omega_2$ è l'evento dell'esperimento combinato corrispondente al fatto che nel primo esperimento si verifichi l'evento A e nel secondo si verifichi l'evento *certo* Ω_2 . Pertanto la probabilità dev'essere pari a quella relativa al solo esperimento 1, ovvero a $P_1(A)$. In maniera analoga si ragiona per la seconda relazione. Osserviamo allora che P_1 e P_2 possono servire solo a calcolare probabilità di eventi del tipo $A \times \Omega_2$ e $\Omega_1 \times B$, ma non consentono certo di determinare la probabilità P di un *qualunque* evento di Ω ; ciò significa che la legge P può essere assegnata con una certa libertà, a patto di rispettare le condizioni (2.4) e (2.5). ◀

Dall'esempio precedente, abbiamo osservato che non tutti gli eventi di $\Omega_1 \times \Omega_2$ sono del tipo $A \times B$, e quindi \mathcal{S} non è semplicemente dato da $\mathcal{S}_1 \times \mathcal{S}_2$: d'altra parte, se Ω_1 ed Ω_2 sono insiemi finiti di cardinalità n_1 ed n_2 , e se \mathcal{S}_1 e \mathcal{S}_2 sono le classi di tutti i sottoinsiemi di Ω_1 e Ω_2 , rispettivamente, si ha che $\text{card}(\mathcal{S}_1) = 2^{n_1}$ e $\text{card}(\mathcal{S}_2) = 2^{n_2}$, per cui $\text{card}(\mathcal{S}_1 \times \mathcal{S}_2) = 2^{n_1+n_2}$ mentre $\text{card}(\mathcal{S}) = 2^{n_1 n_2} > 2^{n_1+n_2}$. In generale è possibile costruire il σ -campo \mathcal{S} partendo da $\mathcal{S}_1 \times \mathcal{S}_2$ e aggiungendo complementi, unioni e intersezioni di un numero finito o infinito numerabile di insiemi. I precedenti concetti possono essere formalizzati nella seguente definizione di *spazio di probabilità prodotto*:

Definizione (spazio di probabilità prodotto). Si considerino due spazi di probabilità $(\Omega_1, \mathcal{S}_1, P_1)$ e $(\Omega_2, \mathcal{S}_2, P_2)$. Si definisce *spazio di probabilità prodotto* lo spazio di probabilità (Ω, \mathcal{S}, P) dove:

- lo spazio campione Ω è il prodotto cartesiano $\Omega_1 \times \Omega_2$, ovvero i risultati ω dell'esperimento combinato sono le coppie ordinate $\omega = (\omega_1, \omega_2)$, con $\omega_1 \in \Omega_1$ e $\omega_2 \in \Omega_2$;
- il σ -campo degli eventi \mathcal{S} è il più piccolo σ -campo contenente gli eventi del tipo $A \times B$, con $A \in \mathcal{S}_1$ e $B \in \mathcal{S}_2$;
- la legge di probabilità P definita su \mathcal{S} deve soddisfare le seguenti proprietà (di consistenza):

$$P(A \times \Omega_2) = P_1(A), \quad \forall A \in \mathcal{S}_1; \quad (2.6)$$

$$P(\Omega_1 \times B) = P_2(B), \quad \forall B \in \mathcal{S}_2. \quad (2.7)$$

La precedente definizione può essere facilmente estesa al caso di $n > 2$ esperimenti.

Abbiamo già notato che le (2.6) e (2.7) consentono di determinare, a partire dagli spazi di probabilità sui singoli esperimenti, solo le probabilità di eventi del tipo $A \times \Omega_2$ e $\Omega_1 \times B$, ma non quelle di un *qualsiasi* evento di \mathcal{S} . D'altra parte, in generale, è intuitivamente accettabile che assegnare solo le leggi di probabilità P_1 e P_2 sui due esperimenti componenti non consente di determinare la legge di probabilità dell'esperimento combinato: abbiamo bisogno di qualche informazione sulla *relazione di dipendenza* che c'è tra i due esperimenti.

2.4.1 Esperimenti indipendenti

Un caso particolarmente semplice è quello in cui si suppone che gli esperimenti siano *indipendenti*:

Definizione (esperimenti indipendenti). Siano $(\Omega_1, \mathcal{S}_1, P_1)$ e $(\Omega_2, \mathcal{S}_2, P_2)$ due spazi di probabilità, e sia (Ω, \mathcal{S}, P) lo spazio di probabilità prodotto. Gli esperimenti si diranno indipendenti se gli eventi $(A \times \Omega_2)$ e $(\Omega_1 \times B)$ dello spazio prodotto sono indipendenti per ogni $A \in \mathcal{S}_1$ e $B \in \mathcal{S}_2$.

In sostanza, dalla precedente definizione di indipendenza,² si ha che per tutti gli eventi di \mathcal{S} che possono essere espressi come $A \times B$, con $A \in \mathcal{S}_1$ e $B \in \mathcal{S}_2$, poiché risulta:

$$A \times B = (A \times \Omega_2) \cap (\Omega_1 \times B),$$

si ha:

$$P(A \times B) = P[(A \times \Omega_2) \cap (\Omega_1 \times B)] = P(A \times \Omega_2)P(\Omega_1 \times B) = P_1(A)P_2(B).$$

In particolare, osserviamo che, per gli eventi elementari di Ω , si ha $(\omega_1, \omega_2) = \{\omega_1\} \times \{\omega_2\}$, per cui

$$P(\omega_1, \omega_2) = P_1(\omega_1)P_2(\omega_2). \quad (2.8)$$

È facile dimostrare a questo punto, almeno per spazi di probabilità discreti, che l'ipotesi di indipendenza consente di calcolare completamente le probabilità dello spazio prodotto in termini delle probabilità degli spazi componenti. Infatti, un qualunque evento appartenente al σ -campo costruito sullo spazio di probabilità prodotto potrà essere espresso come unione al più numerabile di eventi elementari dello spazio prodotto, e quindi la sua probabilità si potrà calcolare, a partire dalle probabilità degli eventi elementari (2.8), adoperando l'assioma di numerabile additività. Concetti più sofisticati di teoria della misura mostrano che è possibile procedere in maniera simile anche per spazi di probabilità continui. *In definitiva, allora, nel caso di esperimenti indipendenti è possibile specificare la legge di probabilità P sullo spazio prodotto semplicemente a partire dalle leggi di probabilità P_1 e P_2 definite sugli spazi componenti.*

► *Esempio 2.10.* Torniamo al caso del lancio di una moneta e di un dado. Se supponiamo che gli esperimenti siano indipendenti, e la moneta ed il dado non siano truccati, avrò ad esempio:

$$P(T, 1) = P_1(T)P_2(1) = \frac{1}{2} \cdot \frac{1}{6} = \frac{1}{12}.$$

Poiché lo spazio campione dell'esperimento combinato è composto da 12 elementi, è facile riconoscere che i risultati dell'esperimento combinato sono equiprobabili e con probabilità $1/12$. ◀

► *Esempio 2.11.* Consideriamo adesso il lancio di una moneta 2 volte (i lanci sono assunti indipendenti). È chiaro che quest'esperimento si può vedere come il prodotto di due esperimenti, ciascuno dei quali si riferisce ad un singolo lancio. Possiamo introdurre, allora, i seguenti spazi campione:

$$\begin{aligned} \Omega_1 &= \{T, C\}, \\ \Omega_2 &= \Omega_1 = \{T, C\}, \\ \Omega &= \Omega_1 \times \Omega_2 = \Omega_1^2 = \{TT, TC, CT, CC\}. \end{aligned}$$

²Notiamo che tale definizione può apparire più elaborata della precedente definizione di indipendenza di eventi appartenenti ad un medesimo spazio di probabilità, vale a dire $P(AB) = P(A)P(B)$, ma non è sostanzialmente differente: bisogna infatti osservare che per parlare di indipendenza tra due eventi bisogna che i due eventi A e B appartengano ad uno *stesso* spazio di probabilità; pertanto, bisogna *prima* costruire lo spazio di probabilità prodotto.

Notiamo poi che poiché i lanci sono effettuati con la stessa moneta (supposta bilanciata), risulta $P_1 = P_2$, è poichè i due lanci sono assunti indipendenti, allora si ha:

$$P(TT) = P_1(T) P_2(T) = P_1(T)P_2(T) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$$

e similmente per le altre probabilità. Si trova in questo caso che i quattro possibili risultati dell'esperimento combinato sono *equiprobabili*. ◀

► *Esempio 2.12.* L'esempio 2.11 può facilmente essere generalizzato al caso di n lanci indipendenti di una moneta bilanciata. In questo caso lo spazio campione dell'esperimento prodotto è $\Omega = \Omega_1^n$ i cui elementi sono le 2^n stringhe di lunghezza n composte dai simboli T e C : ad esempio, per $n = 4$ si ha:

$$\Omega = \Omega_1^4 = \{T, C\}^4 = \{TTTT, TTTC, TTCT, \dots, CCCC\}$$

Poiché gli n lanci sono effettuati con la stessa moneta, si ha $P_1 = P_2 = \dots = P_n$; poiché poi sono assunti indipendenti, allora la probabilità di una qualunque successione (stringa) di lanci si calcola facilmente, in quanto, considerando, ad esempio, la stringa composta da n teste, si ha

$$P(TTT \dots T) = \underbrace{P_1(T) P_1(T) \dots P_1(T)}_{n \text{ termini}} = \frac{1}{2} \cdot \frac{1}{2} \dots \frac{1}{2} = \left(\frac{1}{2}\right)^n.$$

Anche in questo caso, le 2^n stringhe sono tutte equiprobabili. Questo può sembrare controintuitivo, poiché potremmo pensare (data l'equiprobabilità tra testa e croce) che le stringhe con circa $n/2$ teste ed $n/2$ croci debbano essere *più probabili* di una stringa con n teste o n croci. In realtà il risultato ottenuto è corretto, e l'apparente contraddizione va sanata portando in conto il differente *numero* di tali stringhe. Tale problema sarà analizzato più in dettaglio quando si studierà il problema delle prove ripetute e la distribuzione binomiale di probabilità) (cfr. 3.5.2). ◀

Notiamo che nella pratica l'indipendenza tra due o più esperimenti si giustifica con considerazioni di natura fisica o intuitiva. Ciò significa che, in molti casi, l'indipendenza è una assunzione o una ipotesi di lavoro (come l'equiprobabilità), spesso motivata in accordo al principio di ragione insufficiente.

► *Esempio 2.13.* Consideriamo come primo esperimento il seguente: si sceglie a caso una persona in un gruppo di persone, il risultato dell'esperimento è la sua altezza h . Come secondo esperimento, possiamo considerare il seguente: si sceglie a caso una persona in un gruppo di persone, il risultato dell'esperimento è il suo peso p . L'esperimento combinato può essere descritto nel modo seguente: si sceglie a caso una persona in un gruppo di persone, il risultato dell'esperimento è una coppia (h, p) rappresentante l'altezza ed il peso della persona. È chiaro, da considerazioni intuitive, che i due esperimenti non sono indipendenti, perchè esiste una chiara relazione di dipendenza fisica tra altezza e peso di una persona.

Consideriamo, invece, un terzo esperimento definito nel modo seguente: si sceglie a caso una persona in un gruppo, il risultato dell'esperimento è il colore dei suoi occhi. È chiaro adesso che le stesse motivazioni intuitive ci indurranno a ritenere indipendenti il primo ed il terzo esperimento, così come il secondo ed il terzo, in quanto non esiste nessuna relazione evidente e dimostrata tra l'altezza ed il colore degli occhi di una persona, oppure tra il peso ed il colore degli occhi di una persona. Se anche sospettassimo l'esistenza di una relazione del genere, non sapremmo quantificarla, e quindi non ci resterebbe che assumere ugualmente gli esperimenti indipendenti. ◀

► *Esempio 2.14.* Si hanno due scatole:

- la scatola S_1 contiene 10 palle bianche e 5 rosse;
- la scatola S_2 contiene 20 palle bianche e 20 rosse.

Si estrae una palla da ogni scatola. Calcolare la probabilità che la palla estratta dalla scatola S_1 sia bianca e quella estratta dalla scatola S_2 sia rossa.

Gli spazi campione e le leggi di probabilità associate ai singoli esperimenti sono i seguenti (si assumono gli eventi elementari equiprobabili):

$$\begin{aligned}\Omega_1 &= \{10 \text{ bianche, } 5 \text{ rosse}\} \Rightarrow P_1(\omega_i) = 1/15; \\ \Omega_2 &= \{20 \text{ bianche, } 20 \text{ rosse}\} \Rightarrow P_2(\omega_i) = 1/40.\end{aligned}$$

Lo spazio campione dell'esperimento combinato $\Omega = \Omega_1 \times \Omega_2$ ha $15 \times 40 = 600$ elementi, ognuno con probabilità $1/600$. Sia:

$$\begin{aligned}A &= \{\text{si estrae una palla bianca da } S_1\} \Rightarrow P_1(A) = \frac{10}{15} = \frac{2}{3}; \\ B &= \{\text{si estrae una palla rossa da } S_2\} \Rightarrow P_2(B) = \frac{20}{40} = \frac{1}{2}.\end{aligned}$$

Si ha allora:

$$P(\text{si estrae una palla bianca da } S_1 \text{ ed una rossa da } S_2) = P(A \times B) = P_1(A) P_2(B) = \frac{2}{3} \cdot \frac{1}{2} = \frac{1}{3},$$

per l'indipendenza tra le estrazioni delle palle dalle 2 scatole. ◀

► *Esempio 2.15.* Riprendiamo il problema dell'incontro (esempio 1.16), per mostrare come esso si possa interpretare come esperimento combinato. Possiamo infatti costruire due spazi di probabilità $(\Omega_1, \mathcal{S}_1, P_1)$ e $(\Omega_2, \mathcal{S}_2, P_2)$ che descrivono, rispettivamente, l'istante x di arrivo di Tizio e l'istante y di arrivo di Caio. Risulta $\Omega_1 = \Omega_2 = [0, T]$, come σ -campo $\mathcal{S}_1 = \mathcal{S}_2$ sceglieremo la classe degli insiemi di Borel e come legge di probabilità su Ω_1 ed Ω_2 potremo assumere quella uniforme, cioè se $A = [a, b] \subseteq \Omega_1$, si avrà:

$$P_1(A) = \frac{\text{misura}(A)}{\text{misura}(\Omega_1)} = \frac{b-a}{T}$$

e similmente se $B = [c, d] \subseteq \Omega_2$ si avrà:

$$P_2(B) = \frac{\text{misura}(B)}{\text{misura}(\Omega_2)} = \frac{d-c}{T}.$$

Lo spazio campione prodotto sarà $\Omega = \Omega_1 \times \Omega_2 = [0, T] \times [0, T]$, cioè il quadrato di lato T , con \mathcal{S} costruito come il più piccolo σ -campo contenente i rettangoli aperti di $[0, T] \times [0, T]$. Se assumiamo che l'arrivo di Tizio sia indipendente da quello di Caio, gli esperimenti saranno indipendenti, per cui possiamo porre:

$$P((x, y) \in [a, b] \times [c, d]) = P(A \times B) = P_1(A) P_2(B) = \frac{(b-a)(d-c)}{T^2},$$

che è lo stesso risultato che abbiamo già ricavato nell'esempio 1.16. Notiamo che questa rappresenta solo la probabilità di eventi di tipo rettangolare, ovvero esprimibili come $A \times B$, con $A \in \mathcal{S}_1$ e $B \in \mathcal{S}_2$. Utilizzando un teorema fondamentale di teoria della misura (teorema di Carathéodory) si può provare che tale probabilità si può estendere a tutto il σ -campo \mathcal{S} , come $P(A) = \frac{\text{misura}(A)}{\text{misura}(\Omega)}$. ◀

2.5 Elementi di un sistema di comunicazione

I concetti studiati nei precedenti paragrafi, e particolarmente quelli di esperimenti combinati e di eventi indipendenti, possono essere applicati ad un caso di particolare interesse nell'ingegneria dell'informazione, e cioè quello della trasmissione dell'informazione da una sorgente ad una destinazione, mediante un canale di comunicazione. Tale canale di comunicazione, in pratica, è un cavo metallico, una fibra ottica, o lo spazio libero (nel caso di trasmissione via radio).

Per discutere il problema in un caso semplice, consideriamo lo schema (astratto e semplificato) di un sistema di comunicazione riportato in Fig. 2.4, nel quale una *sorgente* emette simboli X , per

semplicità supposti appartenenti ad un insieme discreto $\Omega_1 = \{x_1, x_2, \dots, x_K\}$ (alfabeto sorgente), che vengono immessi su un *canale* di comunicazione, il quale restituisce infine alla *destinazione* simboli Y appartenenti ad un insieme discreto $\Omega_2 = \{y_1, y_2, \dots, y_M\}$ (alfabeto destinazione), che può essere differente (in generale) dall'alfabeto Ω_1 di ingresso.³ Il problema fondamentale della comunicazione è che, per inevitabili limitazioni fisiche (ad esempio, per la presenza di *rumore termico* dovuto al moto degli elettroni nei conduttori e per l'attenuazione di potenza che subisce un qualunque segnale che viaggia su di un canale fisico), qualsiasi canale introduce errori casuali, per cui la trasmissione dell'informazione dalla sorgente alla destinazione non è completamente affidabile.



Fig. 2.4. Schema semplificato di un sistema di comunicazione. La sorgente emette simboli X , che sono trasformati dal canale in simboli Y , che giungono alla destinazione.

2.5.1 Sorgente di informazione

Per iniziare il nostro studio, dobbiamo fornire un modello probabilistico per la sorgente di informazione. Faremo per il momento l'ipotesi (implicita nello schema di Fig. 2.4) che la sorgente emetta un unico simbolo X in un determinato istante di tempo e poi rimanga per sempre in quiete. In questo caso ideale, il modello è estremamente semplice: il simbolo emesso dalla sorgente appartiene infatti ad uno spazio campione Ω_1 discreto, con K risultati possibili, per cui la descrizione probabilistica richiede solo l'assegnazione dei K valori di probabilità p_1, p_2, \dots, p_K associati ai simboli x_1, x_2, \dots, x_K , garantendo che la condizione di normalizzazione $\sum_{k=1}^K p_k = 1$ sia soddisfatta. Nel caso particolare di un sorgente con simboli appartenenti ad un alfabeto binario, ovvero $\Omega_1 = \{0, 1\}$, il modello sarebbe concettualmente simile a quello relativo al lancio di una moneta; una tale sorgente prende il nome di *sorgente binaria*, e se i simboli sono equiprobabili la sorgente si dirà anche *simmetrica*. Per quanto osservato, un singolo lancio di una moneta bilanciata rappresenta un esempio concreto di realizzazione di una sorgente binaria simmetrica (*binary symmetric source*, BSS).

2.5.2 Canale di comunicazione e canale binario simmetrico (BSC)

Concentriamo ora l'attenzione sul canale di comunicazione, e per evitare inutili complicazioni consideriamo dall'inizio il caso di alfabeto sorgente e destinazione entrambi *binari*, ovvero $\Omega_1 = \Omega_2 = \{0, 1\}$; in questo caso il canale, accettando in ingresso e restituendo in uscita simboli binari, si dirà *canale binario*.

Poichè tale canale introduce errori (scambi di 0 con 1 e viceversa) in maniera non prevedibile a priori, allora va modellato anch'esso in termini probabilistici. Abbiamo già visto come sia relativamente semplice costruire uno spazio di probabilità $(\Omega_1, \mathcal{S}_1, P_1)$ che descriva la sorgente; le proprietà del canale entrano in gioco quando si vuole costruire uno spazio di probabilità sullo

³In questo paragrafo, l'introduzione del simbolo X (un discorso analogo vale per Y) consente di esprimere sinteticamente l'evento: "il simbolo all'ingresso del canale è 0" mediante la notazione $\{X = 0\}$. Vedremo poi nel capitolo 3 che X rappresenta un semplice esempio di *variabile aleatoria*.

spazio prodotto sorgente-destinazione $\Omega = \Omega_1 \times \Omega_2 = \{00, 01, 10, 11\}$. Non è ragionevole in questo caso assumere ingresso ed uscita del canale indipendenti, perchè anzi ci aspettiamo che, se il canale è sufficientemente *affidabile*, ci sia una forte dipendenza del simbolo di uscita Y dal simbolo di ingresso X . Addirittura, se il canale è ideale (senza errori), dovrà risultare $Y = X$, cioè si avrà una dipendenza *deterministica*; più in generale, si avrà solo una dipendenza *probabilistica* o statistica, cioè si avrà $Y = X$ con elevata probabilità.

La strada più conveniente per descrivere matematicamente il canale è quella di assegnare le *probabilità condizionali* dei simboli in uscita Y , dati i simboli in ingresso X . Ad esempio, possiamo assegnare la probabilità che si abbia in uscita $Y = 0$, sapendo che in ingresso si ha $X = 0$:

$$P(0|0) \triangleq P(Y = 0|X = 0)$$

e analogamente è possibile assegnare le probabilità $P(0|1)$, $P(1|0)$, e $P(1|1)$. Tale descrizione in termini di probabilità condizionate è particolarmente conveniente perchè risulta svincolata dalle caratteristiche della sorgente (i simboli di ingresso sono fissati, e quindi le loro probabilità non compaiono esplicitamente).

Poichè, per un fissato evento condizionante, la probabilità condizionale è una legge di probabilità, devono valere le consuete condizioni di normalizzazione, vale a dire:

$$P(0|0) + P(1|0) = 1,$$

$$P(0|1) + P(1|1) = 1,$$

per cui, delle quattro probabilità condizionali menzionate, solo due possono essere assegnate ad arbitrio, restando univocamente determinate le altre due. Una volta assegnate tali probabilità, se conosciamo le probabilità dei simboli X emessi dalla sorgente, siano esse:

$$P_1(X = 0) = q,$$

$$P_1(X = 1) = p,$$

è chiaro che possiamo descrivere lo spazio di probabilità prodotto applicando la legge della probabilità composta, avendosi, ad esempio,

$$P(X = 0, Y = 0) = P(Y = 0|X = 0) P_1(X = 0) = P(0|0) q$$

e similmente per tutte le altre. Evidentemente, restano anche univocamente determinate le probabilità dei simboli di uscita. Si ha, infatti,

$$P_2(Y = 0) = P(X = 0, Y = 0) + P(X = 1, Y = 0),$$

$$P_2(Y = 1) = P(X = 0, Y = 1) + P(X = 1, Y = 1).$$

Osserviamo che le due probabilità $P(0|1)$ e $P(1|0)$ rappresentano le probabilità di scambiare un simbolo di ingresso con il suo complemento in uscita, e per questo motivo si chiamano *probabilità di scambio* del canale; se tali probabilità di scambio sono uguali tra loro, cioè si ha:

$$P(0|1) = P(1|0) = \varepsilon,$$

il canale binario si dirà *simmetrico* (*binary symmetric channel*, BSC), e sarà descritto dal solo parametro ε . Per la condizione di normalizzazione, risulta anche:

$$P(0|0) = P(1|1) = 1 - \varepsilon,$$

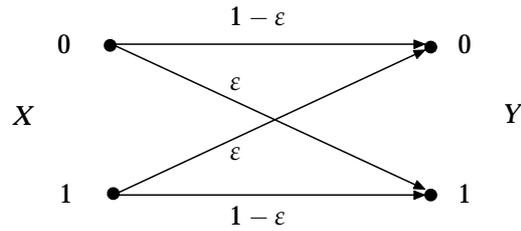


Fig. 2.5. Grafo di un canale binario simmetrico (BSC) con ingresso X , uscita Y e parametro di scambio ε ; i valori indicati sugli archi del grafo rappresentano le probabilità condizionali dei simboli di uscita dati i simboli di ingresso.

dove evidentemente $P(0|0)$ e $P(1|1)$ rappresentano *probabilità di non scambio*. Un canale binario simmetrico è convenientemente rappresentato da un grafo orientato (Fig. 2.5).

Ricaviamo esplicitamente, per un canale binario simmetrico, le probabilità di tutti gli elementi dello spazio prodotto $\Omega = \Omega_1 \times \Omega_2$. Con notazione sintetica, si ha:

$$P(00) = P(0|0) P_1(0) = (1 - \varepsilon) q,$$

$$P(01) = P(1|0) P_1(0) = \varepsilon q,$$

$$P(10) = P(0|1) P_1(1) = \varepsilon p,$$

$$P(11) = P(1|1) P_1(1) = (1 - \varepsilon) p,$$

e a partire da queste possiamo ricavare le probabilità dei simboli di uscita Y :

$$P_2(0) = P(00) + P(10) = (1 - \varepsilon) q + \varepsilon p,$$

$$P_2(1) = P(01) + P(11) = \varepsilon q + (1 - \varepsilon) p.$$

È conveniente raggruppare tali probabilità in una tabella (Tab. 2.1). Possiamo osservare che sommando lungo le righe otteniamo le probabilità P_1 dei simboli X , mentre sommando lungo le colonne otteniamo le probabilità P_2 dei simboli Y .

		Y	
		0	1
X	0	$(1 - \varepsilon) q$	εq
	1	εp	$(1 - \varepsilon) p$

Tab. 2.1. Probabilità caratteristiche di un canale binario simmetrico (BSC).

Se i simboli di ingresso sono equiprobabili ($P_1(0) = P_1(1) = 1/2$, ovvero $p = q = 1/2$), si ricava facilmente che anche i simboli di uscita sono equiprobabili ($P_2(0) = P_2(1) = 1/2$); si noti che questa proprietà consegue dalla simmetria del canale, oltre che dalla equiprobabilità dei simboli di ingresso.

Poichè l'affidabilità di un canale di comunicazione dipende da quanto frequentemente il canale introduca errori, calcoliamo la *probabilità di errore* $P(e)$:

$$\begin{aligned} P(e) &= P(Y \neq X) = P(01 \cup 10) = P(01) + P(10) = P(1|0) P_1(0) + P(0|1) P_1(1) \\ &= \varepsilon q + \varepsilon p = \varepsilon(p + q) = \varepsilon. \end{aligned}$$

Notiamo allora che $P(e) = \varepsilon$, ovvero la probabilità di errore coincide con la probabilità di scambio, indipendentemente dalla distribuzione di probabilità della sorgente. È chiaro allora che ε determina l'affidabilità del canale di comunicazione; quanto più ε è piccolo (valori tipici sono nell'ambito da 10^{-3} a 10^{-9}), tanto più il canale è affidabile.⁴

Possiamo utilizzare il canale binario simmetrico per chiarire meglio il significato di probabilità a priori e a posteriori. Se non osserviamo l'uscita del canale, potremo dire che il simbolo emesso dalla sorgente è 0 con probabilità $P_1(0)$ oppure 1 con probabilità $P_1(1)$ (probabilità a priori). Se però osserviamo l'uscita del canale, sia ad esempio $Y = 1$, tali probabilità a priori si trasformano nelle probabilità a posteriori:

$$\begin{aligned} P(X = 0|Y = 1) &= \frac{P(01)}{P_2(1)} = \frac{\varepsilon q}{\varepsilon q + (1 - \varepsilon)p}, \\ P(X = 1|Y = 1) &= \frac{P(11)}{P_2(1)} = \frac{(1 - \varepsilon)p}{\varepsilon q + (1 - \varepsilon)p}. \end{aligned}$$

Se il canale è affidabile, ovvero se $\varepsilon \ll 1/2$, allora si ha:

$$\begin{aligned} P(X = 0|Y = 1) &\approx \frac{\varepsilon q}{\varepsilon q + p}, \\ P(X = 1|Y = 1) &\approx \frac{p}{\varepsilon q + p}, \end{aligned}$$

per cui si verifica che $P(X = 1|Y = 1) \gg P(X = 0|Y = 1)$ (al limite, per $\varepsilon \rightarrow 0$, si ha $P(X = 1|Y = 1) \rightarrow 1$ e $P(X = 0|Y = 1) \rightarrow 0$); ciò significa che osservare l'uscita $Y = 1$ fa aumentare significativamente la probabilità che la sorgente abbia emesso il simbolo 1. In questo caso il canale trasmette l'informazione dalla sorgente alla destinazione in maniera affidabile.

È interessante individuare la condizione di massima incertezza, nella quale $P(X = 0|Y = 1) = P(X = 1|Y = 1)$. Si trova:

$$\varepsilon q = (1 - \varepsilon)p \Rightarrow \varepsilon = p.$$

Impostando un problema analogo per il caso in cui si osserva l'uscita $Y = 0$, si ricava simmetricamente $\varepsilon = q$. Le due condizioni sono entrambe soddisfatte se $p = q = \varepsilon$, il che ovviamente implica $\varepsilon = 0.5$. Per cui il canale meno affidabile in assoluto è quello caratterizzato da $\varepsilon = 0.5$; si noti che tale canale ha una probabilità di errore anch'essa pari a 0.5, cioè sbaglia il 50% delle volte.

Si può facilmente verificare che per un BSC con $\varepsilon = 0.5$ tutte le probabilità congiunte si fattorizzano nel prodotto delle probabilità marginali. Questo equivale a dire che gli spazi di probabilità Ω_1 e Ω_2 sono indipendenti. È chiaro che avere un'uscita indipendente dall'ingresso è la condizione più sfavorevole che possa capitare se si trasmette informazione su un canale di comunicazione; per decidere quale simbolo è stato trasmesso, una volta osservato un valore dell'uscita, tanto vale lanciare una moneta e scegliere $X = 0$ se esce testa, $X = 1$ se esce croce (o viceversa).

⁴A dire il vero, osserviamo che un canale con ε prossimo ad 1, ad esempio $\varepsilon = 1 - 10^{-3}$, è altrettanto affidabile di un canale con $\varepsilon = 10^{-3}$; infatti il primo canale inverte sistematicamente i simboli, ma questo può facilmente essere compensato alla destinazione. Per evitare questa incongruenza, considereremo solo valori di ε in $[0, 1/2]$.

2.5.3 Sorgenti e canali senza memoria★

Il caso considerato nel § 2.5.1, di una sorgente che emette un simbolo “una tantum”, è ovviamente irrealistico: nella pratica, una sorgente emette non un solo simbolo, ma una *sequenza* X_1, X_2, \dots , di più simboli (al limite, con uno sforzo di idealizzazione, potremmo considerare anche una sequenza *infinita* di simboli). Per fissare le idee, supponiamo che la sorgente emetta sequenze di simboli di lunghezza n , che denomineremo *blocchi* di lunghezza n ; ad esempio, per una sorgente binaria con alfabeto $\Omega_1 = \{0, 1\}$, tali blocchi di simboli sono in effetti le stringhe di n bit, in numero pari a 2^n . L'emissione di un blocco di n simboli binari è esattamente equivalente a quella del lancio di una moneta n volte, che può essere modellata, come abbiamo già visto nel § 2.4 (cfr. esempio 2.12), in termini di esperimento combinato, avente come spazio campione $\Omega = \Omega_1^n$; in generale, abbiamo anche visto che l'assegnazione di una legge di probabilità per l'esperimento combinato non si ottiene semplicemente a partire dalla legge di probabilità assegnata su Ω_1 . Il caso più semplice è quello in cui si suppone che i simboli emessi in successione dalla sorgente siano *indipendenti*; in tal caso la probabilità di un qualunque blocco di simboli emessi dalla sorgente si ottiene semplicemente moltiplicando tra loro le probabilità dei simboli che compongono il blocco; ad esempio, per una sorgente binaria simmetrica, tutti i blocchi di n simboli avranno una probabilità pari a $(1/2) \times (1/2) \times \dots \times (1/2) = (1/2)^n$. Una sorgente discreta che emette simboli indipendenti prende anche il nome di *sorgente discreta senza memoria* (*discrete memoryless source*, DMS).

Consideriamo adesso la trasmissione di un blocco di n simboli $\mathbf{X} = (X_1, X_2, \dots, X_n)$ su un canale di trasmissione, e immaginiamo per semplicità che $\Omega_1 = \Omega_2 = \{0, 1\}$, ovvero ci riferiamo al caso di sorgente e canale entrambi binari, come nel § 2.5.2. In questo caso, è chiaro che alla destinazione sarà consegnato un blocco di n simboli $\mathbf{Y} = (Y_1, Y_2, \dots, Y_n)$ in generale differente da quello di ingresso. Per descrivere il canale completamente, in tal caso, dovremo assegnare le probabilità condizionate di avere uno tra i qualunque 2^n blocchi alla destinazione dato uno qualunque tra i 2^n blocchi di ingresso; ad esempio, per $n = 3$, tali probabilità sono del tipo:

$$P(\mathbf{Y} = 101 | \mathbf{X} = 001) = P(Y_1 = 1, Y_2 = 0, Y_3 = 1 | X_1 = 0, X_2 = 0, X_3 = 1).$$

Un caso semplice è quello di *canale discreto senza memoria* (*discrete memoryless channel*, DMC), corrispondente concettualmente al caso in cui la trasmissione di un blocco avvenga trasmettendo *indipendentemente* i simboli che lo compongono. In tal caso, la probabilità precedente si fattorizza come:

$$P(\mathbf{Y} = 101 | \mathbf{X} = 001) = P(Y_1 = 1 | X_1 = 0) P(Y_2 = 0 | X_2 = 0) P(Y_3 = 1 | X_3 = 1)$$

e se il canale è simmetrico (BSC) si avrà:

$$P(\mathbf{Y} = 101 | \mathbf{X} = 001) = \varepsilon(1 - \varepsilon)\varepsilon = \varepsilon^2(1 - \varepsilon).$$

Possiamo esprimere il tutto in forma abbastanza compatta, se osserviamo che in pratica la probabilità condizionata precedente si può vedere come il prodotto di ε , elevato al numero di *disaccordi* tra le stringhe 101 e 001, moltiplicato per $(1 - \varepsilon)$, elevato al numero di accordi. Il numero di disaccordi tra due stringhe binarie \mathbf{x} e \mathbf{y} prende il nome di *distanza di Hamming* tra le due stringhe $d_H(\mathbf{x}, \mathbf{y})$, ed in pratica si può calcolare effettuando la somma modulo due⁵ o OR esclusivo (XOR)

⁵La somma modulo due è un'operazione binaria, si denota con il simbolo \oplus ed è caratterizzata dalla seguente tabella

delle due stringhe e contando il numero di 1 del risultato, ovvero il cosiddetto *peso di Hamming* $p_H(\mathbf{x})$ della stringa \mathbf{x} . Matematicamente, si ha:

$$d_H(\mathbf{x}, \mathbf{y}) = p_H(\mathbf{x} \oplus \mathbf{y})$$

Ad esempio, la distanza di Hamming tra le stringhe $\mathbf{x} = 101$ e $\mathbf{y} = 100$ si può calcolare come segue:

$$d_H(101, 100) = p_H(101 \oplus 100) = p_H(001) = 1.$$

Utilizzando questo formalismo, le probabilità condizionate che descrivono il canale assumono la forma generale:

$$P(\mathbf{Y} = \mathbf{y} | \mathbf{X} = \mathbf{x}) = \varepsilon^{d_H(\mathbf{x}, \mathbf{y})} (1 - \varepsilon)^{n - d_H(\mathbf{x}, \mathbf{y})}$$

per cui si vede che la caratterizzazione del canale è particolarmente semplice, in quanto si può esprimere in funzione dell'unico parametro ε e della distanza di Hamming tra le stringhe all'ingresso ed all'uscita.

di verità:

$$0 \oplus 1 = 1 \oplus 0 = 1$$

$$0 \oplus 0 = 1 \oplus 1 = 0$$

2.6 Esercizi proposti

Esercizio 2.1. Da un mazzo di carte francesi (senza jolly) si sottrae una carta senza guardarla. Poi si gira un'altra carta: con quale probabilità questa è di fiori? [*Risposta:* $\frac{1}{4}$]

Esercizio 2.2. Risolvere l'esercizio 1.16 utilizzando le leggi della probabilità condizionale.

Esercizio 2.3. Risolvere l'esercizio 1.18 utilizzando le leggi della probabilità condizionale.

Esercizio 2.4. (Paradosso dei due figli). Considerate le seguenti due formulazioni del medesimo problema:

- una coppia ha due figli; sapendo che uno dei due è maschio, qual è la probabilità che anche l'altro sia maschio?
- una coppia ha due figli; sapendo che *il maggiore* dei due è maschio, qual è la probabilità che anche l'altro sia maschio?

Calcolate le due probabilità e discutete il risultato. [*Risposta:* $\frac{1}{3}, \frac{1}{2}$]

★ *Esercizio 2.5.* A e B giocano a dadi, a turno tirano due dadi (comincia A) e vince chi per primo ottiene un punteggio maggiore o uguale a 7. Si determinino le rispettive probabilità di vittoria. [*Risposta:* probabilità che vinca A = $\frac{12}{17}$; probabilità che vinca B = $\frac{5}{17}$]

Esercizio 2.6. Una scatola contiene tre dadi, di cui uno è truccato in modo tale che $P(6) = 2/3$, mentre gli altri due sono bilanciati. Si estrae a caso un dado e lo si lancia ottenendo un 6. Qual è la probabilità che sia il dado truccato? Ripetere il calcolo sapendo che lanciando lo stesso dado una seconda volta si riottiene un 6. [*Risposta:* $\frac{2}{3}, \frac{8}{9}$]

Esercizio 2.7. Una compagnia di assicurazione ha tre tipologie di clienti: ad alto rischio, a medio rischio, e a basso rischio. In particolare, il 20% dei clienti è ad alto rischio, il 30% è a medio rischio, ed il 50% è a basso rischio. Inoltre, la probabilità che un cliente abbia almeno un incidente durante l'anno è pari a 0.25 per clienti ad alto rischio, a 0.16 per clienti a medio rischio, ed a 0.10 per clienti a basso rischio.

- Determinare la probabilità che un cliente scelto a caso abbia almeno un incidente durante l'anno.
- Determinare la probabilità che un cliente sia ad alto rischio, sapendo che ha avuto almeno un incidente durante l'anno.

[*Risposta:* 0.148, 0.338]

Esercizio 2.8. Si hanno due monete, una bilanciata e l'altra con due teste. Si sceglie una moneta a caso e si lancia due volte, ottenendo due teste. Qual è la probabilità che si sia scelta la moneta bilanciata? [*Risposta:* $\frac{1}{5}$]

Esercizio 2.9. Un calcolatore elettronico smette di funzionare se entrambi i componenti A e B si guastano. Il componente A si guasta con probabilità 0.01, ed il componente B con probabilità 0.005. Tuttavia, la probabilità che B si guasti aumenta di un fattore 4 se A si è guastato. Calcolare:

- la probabilità che il calcolatore vada fuori servizio;
- la probabilità che A sia guasto se B si è guastato.

[*Risposta:* 0.0002, 0.04]

★ *Esercizio 2.10.* (Urna di Polya). Un'urna contiene b palle blu e c palle ciano. Si estrae una palla a caso, si verifica il colore, e si reintroduce nell'urna insieme con d palle dello stesso colore⁶. La procedura viene ripetuta all'infinito. Qual è la probabilità che:

- la prima palla estratta sia ciano;

⁶Questo schema fu introdotto dal matematico G. Polya per descrivere gli effetti di una malattia contagiosa. Infatti l'estrazione di una palla di un colore aumenta la probabilità di estrarre successivamente una palla dello stesso colore, il che rappresenta un modello semplificato per il contagio di una malattia, nelle quali il verificarsi di un caso aumenta la probabilità che ci siano ulteriori casi.

- b) la seconda palla estratta sia ciano?
 c) la prima palla estratta sia ciano, sapendo che la seconda palla estratta è ciano?

[Risposta: $\frac{c}{b+c}$, $\frac{c}{b+c}$, $\frac{c+d}{b+c+d}$]

Esercizio 2.11. Se N studenti nati nel 1983 stanno seguendo il corso di Teoria dei Fenomeni Aleatori, qual è la probabilità che almeno due di essi festeggino il compleanno nello stesso giorno? Che cosa cambierebbe se gli studenti fossero nati nel 1984? [Risposta: $1 - \frac{365!}{365^N(365-N)!}$]

Esercizio 2.12. Se $P(A) = 1/3$ e $P(\bar{B}) = 1/4$, A e B possono essere indipendenti? Possono essere mutualmente esclusivi? Motivare le risposte.

Esercizio 2.13. (Paradosso di de Meré). Dimostrare che è più probabile ottenere almeno un 6 lanciando un dado 4 volte che un doppio 6 lanciando due dadi 24 volte.⁷

Esercizio 2.14. Si considerino N punti p_1, p_2, \dots, p_N presi indipendentemente in un cerchio C di raggio R , con $P(p_i \in A) = \text{misura}(A)/\text{misura}(C)$, $\forall A \subseteq C$, dove $\text{misura}(A)$ rappresenta l'area di A . Determinare la probabilità che il punto più vicino al centro abbia da esso distanza maggiore di $r \leq R$.

★ **Esercizio 2.15.** Dovete affrontare in un torneo di scacchi i maestri Alekhine, Botvinnik e Capablanca, una volta ciascuno. Le vostre probabilità di vittoria contro i tre sono, rispettivamente, $p_A > p_B > p_C$; vi aggiudicate il torneo se vincete due partite consecutive, altrimenti perdete. Avete però la possibilità di scegliere in che ordine affrontare i tre avversari. Mostrate che per massimizzare la vostra probabilità di vittoria dovete affrontare Alekhine per secondo.

Esercizio 2.16. Siano A, B e C tre eventi indipendenti. Mostrare che risultano indipendenti:

- a) l'evento A e l'evento BC ;
 b) l'evento A e l'evento $B \cup C$.

Esercizio 2.17. Nel lancio di due dadi, si considerino i seguenti eventi:

$$\begin{aligned} A &= \{\text{esce dispari al primo lancio}\}, \\ B &= \{\text{esce dispari al secondo lancio}\}, \\ C &= \{\text{la somma dei due lanci è un numero pari}\}. \end{aligned}$$

Verificare che A, B e C sono indipendenti a coppie, ma non sono indipendenti.

Esercizio 2.18. Nel lancio di due dadi, si considerino i seguenti eventi:

$$\begin{aligned} A &= \{\text{esce 6 al primo lancio}\}, \\ B &= \{\text{esce 6 al secondo lancio}\}, \\ C &= \{\text{la somma dei due lanci è maggiore o uguale a 10}\}. \end{aligned}$$

Verificare che A e B sono indipendenti, ma non sono condizionalmente indipendenti dato C .

★ **Esercizio 2.19.** (Paradosso di Monty Hall). In un gioco televisivo a premi un concorrente è invitato a scegliere una tra tre porte chiuse: dietro due di tali porte ci sono due capre, mentre dietro la rimanente c'è una lussuosa automobile. Si supponga che il concorrente scelga la porta numero 1: a questo punto il conduttore del gioco apre la porta numero 2 dietro la quale vi è una capra, e chiede al concorrente se questi voglia cambiare la propria scelta della porta oppure no. Qual è la scelta più conveniente per il concorrente?⁸

⁷Questo è il calcolo originariamente effettuato nel 1654 dal matematico e filosofo francese B. Pascal (1623-1662) su richiesta di un famoso scommettitore e matematico dilettante, il cavaliere de Meré, che riteneva erroneamente che i due eventi avessero la stessa probabilità.

⁸Questo problema fu effettivamente proposto agli ospiti di un celebre gioco a premi televisivo americano "Let's make a deal", il cui conduttore era appunto Monty Hall, e suscitò una accesa controversia sulla rivista "Parade" nel 1990 su quale fosse la scelta più conveniente (si veda P. Hoffman, "L'uomo che amava solo i numeri", ed. Mondadori (1999)).

★ *Esercizio 2.20.* (Paradosso dei prigionieri). Tre prigionieri A, B, e C sono condannati a morte. Il governatore decide di concedere la grazia ad uno di essi scelto a caso, ed informa il secondino della sua scelta, chiedendogli di non rivelare tale nome ai prigionieri. Il giorno successivo, A cerca inutilmente di sapere dal secondino chi sia stato graziato. Allora A chiede al secondino di rivelargli almeno chi tra B e C sarà giustiziato, ed il secondino, dopo averci pensato un attimo, gli rivela che B sarà giustiziato. A è soddisfatto della risposta del secondino, perchè ritiene che la probabilità di essere stato graziato sia cresciuta da $1/3$ ad $1/2$. Ha ragione?⁹

Esercizio 2.21. Nel codice telegrafico Morse si utilizzano punti (dot) e linee (dash), nella proporzione di 3 : 4, per codificare le lettere dell'alfabeto. Si supponga che errori nella trasmissione possano far interpretare erroneamente in ricezione un punto come una linea con probabilità $1/4$, ed una linea come un punto con probabilità $1/3$.

- Mostrare che il problema può essere descritto da un modello di *canale binario non simmetrico*.
- Sapendo che è stata ricevuta una linea, calcolare la probabilità che sia stata trasmessa una linea.
- Supponendo che le successive trasmissioni siano indipendenti, nell'ipotesi che sia stato ricevuto il messaggio punto-punto, calcolare la distribuzione di probabilità dei quattro possibili messaggi trasmessi.

Esercizio 2.22. Caratterizzare il canale binario equivalente ottenuto collegando in cascata tre BSC indipendenti con probabilità di scambio ε_i , $i = 1, 2, 3$. Discutere in particolare il caso $\varepsilon_i = \varepsilon = 10^{-3}$. L'affidabilità della trasmissione aumenta o diminuisce?

Esercizio 2.23. Caratterizzare il canale binario equivalente ottenuto trasmettendo tre volte lo stesso simbolo su un BSC di parametro ε e decidendo a maggioranza in ricezione (si supponga che le differenti trasmissioni siano indipendenti). Discutere in particolare il caso $\varepsilon = 10^{-3}$. L'affidabilità della trasmissione aumenta o diminuisce?

Esercizio 2.24. Un simbolo binario è trasmesso in parallelo su tre BSC indipendenti con probabilità di scambio ε_i , $i = 1, 2, 3$. In ricezione si decide per il simbolo presente in almeno due delle uscite dei tre canali. Determinare la probabilità di scambio del canale binario equivalente, discutendo in particolare il caso $\varepsilon_i = \varepsilon = 10^{-3}$. L'affidabilità della trasmissione aumenta o diminuisce?

Esercizio 2.25. Tre sorgenti binarie indipendenti emettono il simbolo 1 con probabilità p_i , $i = 1, 2, 3$, e sono connesse mediante un interruttore ad un BSC. L'interruttore è connesso per il 50% del tempo alla prima sorgente, e per il 25% del tempo a ciascuna delle altre due (indipendentemente dai simboli emessi dalle sorgenti). Determinare:

- la probabilità dei simboli in uscita al BSC;
- la probabilità che il canale sia connesso alla prima sorgente avendo osservato uno zero in uscita al BSC.

★ *Esercizio 2.26.* Per aumentare l'affidabilità nella trasmissione di una coppia di bit, ad essi viene concatenato un terzo bit (bit di parità) in modo che il numero di bit alti in ciascuna terna risulti pari (oppure zero). Le terne di bit così ottenute vengono trasmesse in serie su un canale BSC con probabilità di scambio ε (si suppongano le successive trasmissioni indipendenti). In ricezione, se la terna di bit non soddisfa la regola di parità, si richiede una ritrasmissione della terna, altrimenti si elimina il bit di parità riottenendo l'originaria coppia di bit.

- Calcolare la probabilità p_R di ritrasmissione, la probabilità p_C di ricevere correttamente la coppia di bit trasmessi, la probabilità p_E di commettere un errore che il bit di parità non è in grado di individuare (ovviamente deve risultare $p_R + p_C + p_E = 1$);
- calcolare la probabilità p_T di errore complessiva tenendo conto anche delle ritrasmissioni;
- confrontare p_T con la probabilità di errore che si otterrebbe trasmettendo direttamente la coppia di bit senza alcun controllo di parità (si assuma $\varepsilon = 10^{-3}$).

⁹La formulazione di questo problema è simile a quella del precedente paradosso di Monty Hall; la soluzione classica non è difficile da ottenere, ma considerazioni più approfondite evidenziano le ambiguità che possono sorgere nella costruzione di spazi di probabilità prodotto.

Variabili aleatorie

L concetto di variabile aleatoria, introdotto in questo capitolo, è anch'esso fondamentale nello studio della probabilità, in quanto consente di associare ad ogni risultato di un esperimento un numero reale, e quindi di trasformare lo spazio campione in un insieme di numeri reali. Il vantaggio è quello di poter applicare alla risoluzione dei problemi di probabilità i potenti strumenti dell'analisi matematica, al prezzo forse di una maggiore astrazione nella descrizione del problema. Nel capitolo si forniscono i principali strumenti per la descrizione di una singola variabile aleatoria (i casi di più variabili aleatorie sono trattati nei capitoli che seguono); in particolare, si introducono le funzioni necessarie per la caratterizzazione statistica di una variabile aleatoria (funzione di distribuzione cumulativa, funzione di densità di probabilità, funzione distribuzione di probabilità). Infine, dopo aver classificato le variabili aleatorie in continue, discrete e miste, si presentano alcune tra le variabili aleatorie discrete e continue maggiormente utilizzate nelle applicazioni, tra cui la variabile aleatoria binomiale, associata al problema delle prove ripetute, e la variabile aleatoria gaussiana o normale.

3.1 Introduzione

Nei precedenti capitoli abbiamo mostrato come costruire spazi di probabilità a partire da esperimenti i cui risultati non sono necessariamente numerici, quali ad esempio il lancio di una moneta o di un dado, o l'estrazione di una carta da un mazzo di carte francesi. Nelle scienze fisiche e nell'ingegneria, tuttavia, nasce la necessità di descrivere i risultati di un esperimento e gli eventi associati ad esso in maniera *numerica*. Un modo semplice di ottenere ciò consiste nell'associare ad ogni risultato dell'esperimento un numero reale; ciò conduce al fondamentale concetto di *variabile aleatoria*.

► *Esempio 3.1.* Consideriamo il lancio di una moneta, il cui spazio campione è $\Omega = \{T, C\}$. Un esempio di variabile aleatoria si ottiene associando al risultato T il numero 1, ed al risultato C il numero 0. ◀

► *Esempio 3.2.* Consideriamo il lancio di un dado, il cui spazio campione è $\Omega = \{\omega_1, \omega_2, \dots, \omega_6\}$, dove con ω_i abbiamo indicato il risultato che nel lancio si presenti la i -esima faccia del dado. Possiamo costruire una variabile aleatoria semplicemente associando a ω_i il valore i . Si noti che abbiamo già (implicitamente) effettuato questa corrispondenza, quando abbiamo assunto come spazio campione per il lancio di un dado l'insieme $\Omega = \{1, 2, \dots, 6\}$, invece dell'insieme $\Omega = \{\omega_1, \omega_2, \dots, \omega_6\}$. ◀

Negli esempi precedenti, costruire una variabile aleatoria X su Ω equivaleva semplicemente a *cambiare nome* (associando dei valori numerici) ai risultati dell'esperimento, in quanto le funzioni costruite erano biunivoche; tuttavia, in molti altri casi, si può utilizzare opportunamente il concetto di variabile aleatoria per ottenere una significativa riduzione di complessità nella descrizione dell'esperimento, come mostrato dall'esempio seguente.

► *Esempio 3.3.* Consideriamo un sondaggio di opinione, nel quale si intervistano 1000 persone, per sapere se sono d'accordo oppure no su un certo argomento. Lo spazio campione Ω associato a tale esperimento è composto da $2^{1000} \approx 10^{300}$ risultati, ognuno dei quali è una stringa di 1000 simboli, scelti tra S (corrispondente alla risposta "sì") ed N (corrispondente alla risposta "no"), che è un numero eccezionalmente grande (provate a calcolarlo sulla vostra calcolatrice tascabile!). Tuttavia, nella pratica quello che interessa sapere è quante persone rispondono sì e quante no. Allora ad ogni punto (stringa) dello spazio campione possiamo associare il numero dei simboli S presenti nella stringa stessa, ottenendo un numero intero tra 0 e 1000. In questo modo abbiamo descritto il problema in maniera più semplice e soprattutto più aderente alle nostre finalità. ◀

Generalizzando, per un dato esperimento, una variabile aleatoria X (Fig. 3.1) è una *funzione* costruita su Ω e che assume valori nell'insieme $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, \infty\}$:

$$X: \omega \in \Omega \rightarrow X(\omega) \in \mathcal{X} \subseteq \overline{\mathbb{R}}$$

dove abbiamo denotato con \mathcal{X} il *codominio* della funzione X , ovvero l'insieme dei possibili valori assunti da X . Tale funzione deve soddisfare certe condizioni di regolarità, come vedremo meglio nel seguito.

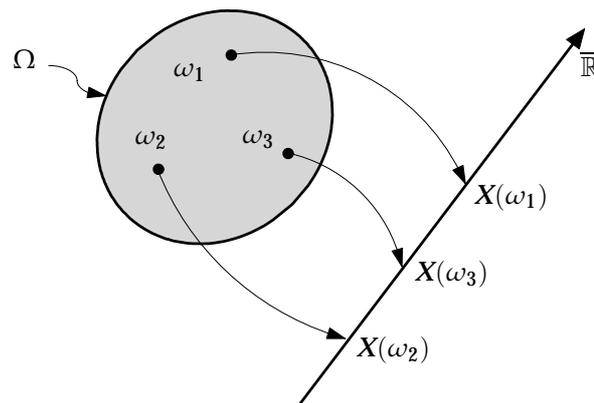


Fig. 3.1. Una variabile aleatoria X è una funzione definita nello spazio campione Ω e a valori in $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, \infty\}$.

► *Esempio 3.4.* Consideriamo ancora il lancio di un dado, per il quale lo spazio campione è $\Omega = \{\omega_1, \omega_2, \dots, \omega_6\}$. Tre diverse variabili aleatorie definite su Ω sono:

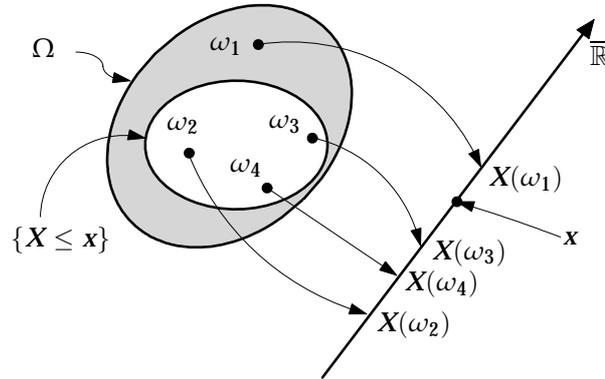


Fig. 3.2. L'evento $\{X \leq x\}$ è il sottoinsieme $A = \{\omega_2, \omega_3, \omega_4\}$ di Ω (in bianco), ottenuto considerando gli elementi $\omega \in \Omega$ la cui immagine attraverso X risulta minore o uguale ad x .

1. $X(\omega_i) = i$;
2. $X(\omega_i) = 10i$;
3. $X(\omega_i) = \begin{cases} 1, & \text{se } i \text{ è pari;} \\ 0, & \text{se } i \text{ è dispari.} \end{cases}$

Notiamo che qui e nel seguito, in analogia alla notazione comunemente utilizzata in matematica, indicheremo con X la legge di corrispondenza (funzione o variabile aleatoria), e con $X(\omega)$ il *valore* della funzione in corrispondenza del risultato ω . ◀

Il successivo passo per una corretta comprensione del concetto di variabile aleatoria è capire in che modo, se Ω è un insieme dotato di struttura di spazio di probabilità, una variabile aleatoria X costruita su Ω “conservi” informazioni sulle probabilità degli eventi di Ω . A tale scopo, di fondamentale importanza è chiarire il significato della notazione

$$\{X \leq x\} \quad (3.1)$$

per un dato $x \in \overline{\mathbb{R}}$. Dal punto di vista dell'analisi reale, l'insieme dei valori *reali* minori o uguali di un dato valore x è una *semiretta sinistra* (chiusa a destra), che si denota anche con $] -\infty, x]$. Tuttavia, il senso che noi daremo alla notazione (3.1) è completamente differente: con essa intenderemo riferirci al sottoinsieme A di Ω così definito:

$$A = \{\omega \in \Omega \text{ tali che } X(\omega) \leq x\},$$

ovvero all'insieme dei valori $\omega \in \Omega$ la cui immagine (Fig. 3.2) attraverso la funzione X è minore o uguale¹ ad x . Pertanto, $\{X \leq x\}$ non va considerato come un sottoinsieme di $\overline{\mathbb{R}}$, ma come un sottoinsieme di Ω .

Se allora $A = \{X \leq x\}$ è un evento $\forall x \in \overline{\mathbb{R}}$, è possibile calcolarne la probabilità $P(A)$; se tale assunzione è verificata, sarà più in generale possibile calcolare la probabilità dell'insieme $B = \{X \in T\} = \{\omega \in \Omega \text{ tali che } X(\omega) \in T\}$, se tale insieme si può ottenere come complemento, unione o intersezione numerabile di eventi del tipo $\{X \leq x\}$; intuitivamente, ciò equivale a dire che l'insieme *numerico* $T \subseteq \overline{\mathbb{R}}$ si può ottenere come complemento, unione o intersezione numerabile di semirette sinistre.

¹Ovviamente, l'ordinamento su $\overline{\mathbb{R}}$ è tale che $-\infty \leq x \leq \infty, \forall x \in \overline{\mathbb{R}}$.

► *Esempio 3.5.* Con riferimento all'esempio precedente (lancio di un dado) e alla variabile aleatoria definita al punto 2, vale a dire $X(\omega_i) = 10i$, si ha:

$$\begin{aligned} \{X \leq 35\} &= \{\omega_1, \omega_2, \omega_3\} \Rightarrow P(X \leq 35) = 1/2 \\ \{X \leq 5\} &= \{\emptyset\} \Rightarrow P(X \leq 5) = 0 \\ \{20 \leq X \leq 35\} &= \{\omega_2, \omega_3\} \Rightarrow P(20 \leq X \leq 35) = 2/6 = 1/3 \end{aligned}$$

◀

3.1.1 Definizione formale di variabile aleatoria

Sulla base dei concetti introduttivi e degli esempi del precedente paragrafo, siamo ora in grado di dare la seguente definizione *formale* di variabile aleatoria:²

Definizione (variabile aleatoria). Dato uno spazio di probabilità (Ω, \mathcal{S}, P) , una variabile aleatoria (v.a.) X è una funzione definita in Ω ed a valori in $\mathcal{X} \subseteq \overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, +\infty\}$, tale che

1. $\{X \leq x\}$ è un evento, $\forall x \in \overline{\mathbb{R}}$;
2. $P(\{X = +\infty\}) = P(\{X = -\infty\}) = 0$.

Il significato della proprietà 1 è stato discusso precedentemente; con la proprietà 2, per motivi matematici che qui non è il caso di approfondire, si consente alla funzione X di assumere il valore $+\infty$ oppure $-\infty$, ma gli eventi $\{X = +\infty\}$ e $\{X = -\infty\}$ devono avere probabilità nulla. Infine, una osservazione sulla notazione: benchè sia più corretta la notazione $P(\{X \leq x\})$, che evidenzia la natura di evento di $\{X \leq x\}$, nel seguito useremo quasi sempre la notazione semplificata, ma più imprecisa, $P(X \leq x)$.

In conclusione, osserviamo che definire una variabile aleatoria su uno spazio di probabilità (Ω, \mathcal{S}, P) equivale in pratica a costruire un nuovo spazio di probabilità, nel quale lo spazio campione diventa $\mathcal{X} \subseteq \overline{\mathbb{R}}$, gli eventi sono sottoinsiemi di \mathcal{X} che si ottengono per complementazioni, unioni ed intersezioni di semirette sinistre, e la legge di probabilità è, per così dire, "indotta" dalla legge di probabilità P .

3.2 Funzione di distribuzione cumulativa (CDF)

La funzione che esprime la probabilità dell'evento $\{X \leq x\}$ al variare di x in $\overline{\mathbb{R}}$ prende il nome di *funzione di distribuzione cumulativa* (CDF) della variabile aleatoria X :

Definizione (funzione di distribuzione cumulativa). Data una variabile aleatoria X , la sua *funzione di distribuzione cumulativa* (CDF) è:

$$F(x) \triangleq P(X \leq x), \quad \forall x \in \overline{\mathbb{R}}.$$

Ha senso calcolare questa probabilità perchè nella definizione di variabile aleatoria abbiamo richiesto (proprietà 1) che $\{X \leq x\}$ sia un evento, $\forall x \in \overline{\mathbb{R}}$. Notiamo anche che, sebbene il codominio di X sia \mathcal{X} , la CDF è definita in tutto $\overline{\mathbb{R}}$.

²Osserviamo che il termine *variabile* aleatoria è fuorviante, trattandosi piuttosto di una *funzione* aleatoria; tuttavia esso è quello più comunemente utilizzato.

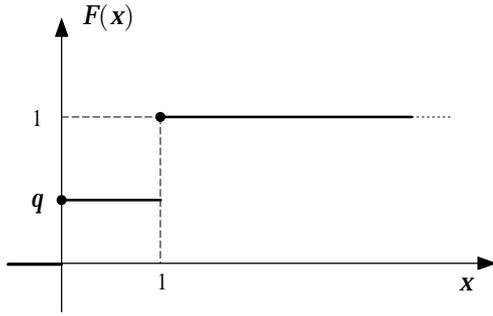


Fig. 3.3. La CDF $F(x)$ della variabile aleatoria dell'esempio 3.6.

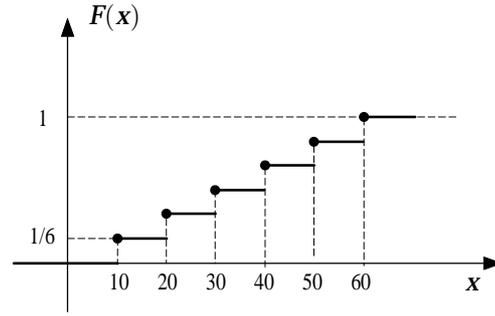


Fig. 3.4. La CDF $F(x)$ della variabile aleatoria dell'esempio 3.7.

In alternativa alla notazione $F(x)$, useremo la notazione $F_X(x)$ quando vorremo specificare esplicitamente che si tratta della CDF della variabile aleatoria X (quindi, ad esempio, quando avremo a che fare con più variabili aleatorie). Osserviamo esplicitamente che il *pedice* X (maiuscolo) rappresenta la variabile aleatoria (ovvero la legge di corrispondenza), mentre la variabile indipendente della funzione x (minuscolo) è un numero reale. Notazioni come $F_X(y)$ oppure $F_X(w)$ sono ovviamente lecite.

► **Esempio 3.6.** Si consideri la variabile aleatoria definita su $\Omega = \{T, C\}$ nel seguente modo:

$$\begin{aligned} X(T) &= 1, \\ X(C) &= 0. \end{aligned}$$

Se $P(T) = p$ e $P(C) = q$, con $p + q = 1$, la CDF di X è la seguente:

$$F(x) = \begin{cases} 0, & x < 0; \\ q, & 0 \leq x < 1; \\ 1, & x \geq 1. \end{cases}$$

Infatti:

- per $x < 0$, si ha $F(x) = P(X \leq x) = P(X \leq x < 0) = P(\emptyset) = 0$;
- per $0 \leq x < 1$, si ha $F(x) = P(X \leq x) = P(X = 0) = P(C) = q$;
- per $x \geq 1$, si ha $F(x) = P(X \leq x) = P(\{X = 0\} \cup \{X = 1\}) = P(C) + P(T) = q + p = 1$.

Osserviamo che tale CDF (Fig. 3.3) ha un andamento costante a tratti. Si parla in questo caso di variabile aleatoria *discreta* (cfr. § 3.2.2). ◀

► **Esempio 3.7.** Sia $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4, \omega_5, \omega_6\}$, con risultati equiprobabili, e si consideri la variabile aleatoria $X(\omega_i) = 10i$. La CDF si calcola facilmente:

$$F(x) = \begin{cases} 0, & x < 10; \\ 1/6, & 10 \leq x < 20; \\ 2/6, & 20 \leq x < 30; \\ 3/6, & 30 \leq x < 40; \\ 4/6, & 40 \leq x < 50; \\ 5/6, & 50 \leq x < 60; \\ 1, & x \geq 60; \end{cases}$$

ed, anche in questo caso, ha un andamento costante a tratti (Fig. 3.4) per cui X è una variabile aleatoria discreta (cfr. § 3.2.2). ◀

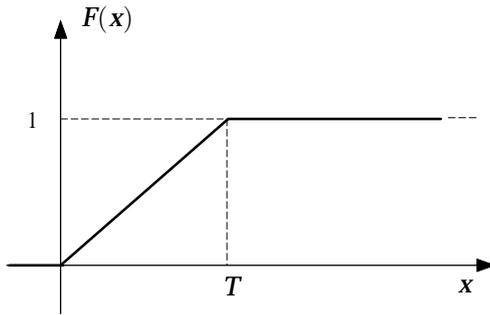


Fig. 3.5. La CDF $F(x)$ della variabile aleatoria dell'esempio 3.8.

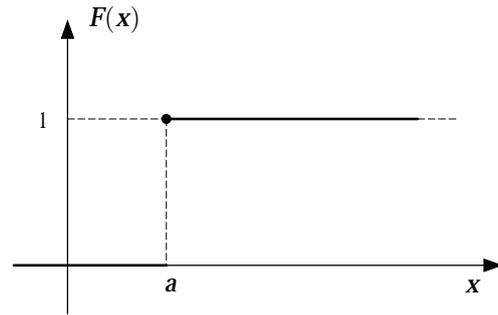


Fig. 3.6. La CDF $F(x)$ della variabile aleatoria dell'esempio 3.9.

► **Esempio 3.8.** Consideriamo l'esperimento consistente nell'arrivo a caso di una telefonata nell'intervallo $[0, T]$, e denotiamo con t l'istante di arrivo della telefonata. Lo spazio campione è $\Omega = [0, T]$, gli eventi sono complementi, unioni ed intersezioni numerabili di intervalli aperti $]a, b[\subseteq \Omega$. Come legge di probabilità, porremo (legge uniforme):

$$P(t \in (a, b)) = \frac{b - a}{T}, \quad \text{con } 0 \leq a \leq b \leq T.$$

Essendo il risultato dell'esperimento già numerico, possiamo definire una variabile aleatoria su Ω semplicemente come la trasformazione identica

$$X(t) = t.$$

È allora semplice calcolare la CDF:

- per $x < 0$, si ha $F(x) = P(X \leq x) = P(X \leq x < 0) = P(\emptyset) = 0$;
- per $0 \leq x < T$, si ha $F(x) = P(X \leq x) = P(0 \leq X \leq x) = x/T$;
- per $x \geq T$, si ha $F(x) = P(X \leq x) = P(0 \leq X \leq T) = T/T = 1$.

In questo caso la CDF (Fig. 3.5) non è una funzione costante a tratti, ma è una funzione *continua* su tutto l'insieme di definizione. Si parla allora di variabile aleatoria *continua* (cfr. § 3.2.2). ◀

► **Esempio 3.9 (variabile aleatoria costante o deterministica).** Sia X una variabile aleatoria definita su un qualunque spazio campione Ω nel seguente modo:

$$X(\omega) = a, \quad \forall \omega \in \Omega.$$

In tal caso, il calcolo della CDF si effettua come segue:

- per $x < a$, si ha $F(x) = P(X \leq x) = P(X \leq x < a) = P(\emptyset) = 0$;
- per $x \geq a$, si ha $F(x) = P(X \leq x) = P(\Omega) = 1$.

Tale CDF (Fig. 3.6) è una funzione costante a tratti, per cui la variabile aleatoria X è di tipo discreto (cfr. § 3.2.2). ◀

3.2.1 Proprietà della CDF

I precedenti esempi hanno mostrato che la CDF di una variabile aleatoria è una funzione a valori in $[0, 1]$ e non decrescente. Oltre a queste due evidenti proprietà, la CDF possiede altre proprietà, riassunte e dimostrate di seguito:³

³Nelle formule che seguono, con $F(x^+)$ e $F(x^-)$ intendiamo il limite *da destra* e *da sinistra* della funzione $F(x)$ nel punto x , ovvero $F(x^+) \triangleq \lim_{\varepsilon \rightarrow 0} F(x + \varepsilon)$ e $F(x^-) \triangleq \lim_{\varepsilon \rightarrow 0} F(x - \varepsilon)$, con $\varepsilon > 0$.

$$1. \quad F(+\infty) = 1, F(-\infty) = 0.$$

Prova. Si ha, banalmente, $F(+\infty) = P(X \leq +\infty) = P(\Omega) = 1$ e $F(-\infty) = P(X \leq -\infty) = P(X = -\infty) = 0$ (per la seconda, si sfrutta la proprietà 2 della definizione di variabile aleatoria). \square

$$2. \quad F(x) \text{ è una funzione monotona crescente, ovvero } x_1 < x_2 \Rightarrow F(x_1) \leq F(x_2).$$

Prova. Se $x_1 < x_2$, si ha che $\{X \leq x_1\} \subseteq \{X \leq x_2\}$ e quindi, per le proprietà elementari della probabilità, $P(X \leq x_1) \leq P(X \leq x_2)$, da cui l'asserto. \square

Notiamo che, se $F(x_0) = 0$, in base a tale proprietà risulta $F(x) = 0, \forall x \leq x_0$. Conseguentemente, se $X(\omega) > 0, \forall \omega \in \Omega$, risulta $F(0) = P(X \leq 0) = 0$ e quindi $F(x) = 0$ per $x \leq 0$. Una tale variabile aleatoria si dice *positiva*, ed ha pertanto CDF nulla per $x \leq 0$.

$$3. \quad P(X > x) = 1 - F(x).$$

Prova. Basta osservare che $\{X \leq x\} \cup \{X > x\} = \Omega$, e che i due eventi sono mutuamente esclusivi, per cui $P(X \leq x) + P(X > x) = P(\Omega) = 1$, da cui l'asserto. \square

La funzione $\bar{F}(x) \triangleq 1 - F(x)$ prende il nome di *CDF complementare* o anche di *funzione di affidabilità*.⁴

$$4. \quad F(x) \text{ è continua da destra, ovvero } F(x^+) = F(x).$$

Prova. Dobbiamo provare che $\lim_{\varepsilon \rightarrow 0} F(x + \varepsilon) = F(x)$, per $\varepsilon > 0$. Notiamo che poiché per la proprietà 2 la $F(x)$ è monotona crescente (e limitata), sicuramente in ogni punto esiste finito il limite da destra e da sinistra (teorema sull'esistenza del limite per le funzioni monotone). Allora, per calcolare il limite da destra, non è restrittivo considerare $\varepsilon = 1/n$ e far tendere $n \rightarrow \infty$ (cioè far tendere ε a zero su una particolare successione di valori). Osserviamo allora che $F(x + 1/n) = P(X \leq x + 1/n) = P(A_n)$, dove abbiamo posto $A_n = \{X \leq x + 1/n\}$; si noti che A_n è una successione decrescente e tale che $\bigcap_{n=1}^{\infty} A_n = A = \{X \leq x\}$, per cui possiamo applicare la proprietà di continuità della probabilità (cfr. § 1.4.5) e scrivere:

$$F(x^+) = \lim_{n \rightarrow \infty} F(x + 1/n) = \lim_n P(A_n) = P(A) = P(X \leq x) = F(x)$$

cioè l'asserto. \square

$$5. \quad P(x_1 < X \leq x_2) = F(x_2) - F(x_1).$$

Prova. Per l'evento $\{x_1 < X \leq x_2\}$, vale la seguente identità:

$$\{X \leq x_1\} \cup \{x_1 < X \leq x_2\} = \{X \leq x_2\}$$

nella quale i due eventi a primo membro sono mutuamente esclusivi, per cui:

$$\underbrace{P(X \leq x_1)}_{=F(x_1)} + P(x_1 < X \leq x_2) = \underbrace{P(X \leq x_2)}_{=F(x_2)}$$

da cui si ha l'asserto. \square

⁴La denominazione di "funzione di affidabilità" deriva dal fatto che, se si interpreta la variabile aleatoria X come il "tempo di vita" di un dispositivo, la funzione $\bar{F}(x)$ descrive la probabilità che il dispositivo "viva" per un tempo maggiore o uguale a x , e quindi misura l'affidabilità del dispositivo stesso.

$$6. \quad P(X = x) = F(x) - F(x^-).$$

Prova. Poniamo $A_n = \{x - 1/n < X \leq x\}$: tale successione di eventi è chiaramente decrescente e tale che $\bigcap_{n=1}^{\infty} A_n = A = \{X = x\}$. Dalla proprietà 5, per $x_1 = x - 1/n$ e $x_2 = x$, si ha:

$$P(A_n) = P(x - 1/n < X \leq x) = F(x) - F(x - 1/n),$$

Passando al limite per $n \rightarrow \infty$, sfruttando la proprietà di continuità della probabilità (cfr. § 1.4.5) si ha $\lim_n P(A_n) = P(A) = P(X = x)$ al primo membro; d'altra parte, come già osservato, la $F(x)$ essendo monotona e limitata ammette sicuramente limite finito da sinistra nel punto x , e quindi si ha:

$$P(X = x) = F(x) - \lim_n F(x - 1/n) = F(x) - F(x^-)$$

cioè l'asserto. □

$$7. \quad P(x_1 \leq X \leq x_2) = F(x_2) - F(x_1^-).$$

Prova. Si ha, banalmente,

$$\{x_1 \leq X \leq x_2\} = \{x_1 < X \leq x_2\} \cup \{X = x_1\}$$

e gli eventi a secondo membro sono mutuamente esclusivi. Si ha allora, per le proprietà 5 e 6,

$$\begin{aligned} P(x_1 \leq X \leq x_2) &= P(x_1 < X \leq x_2) + P(X = x_1) = F(x_2) - F(x_1) + F(x_1) - F(x_1^-) = \\ &= F(x_2) - F(x_1^-) \end{aligned}$$

cioè l'asserto. □

Notiamo che in particolare le proprietà 5 e 7 consentono di calcolare la probabilità che la variabile aleatoria assuma valori nell'intervallo $]x_1, x_2]$ ed $[x_1, x_2]$, rispettivamente. D'altra parte, utilizzando anche la proprietà 6, si trova (la verifica, banale, è lasciata al lettore per esercizio):

$$8. \quad P(x_1 \leq X < x_2) = F(x_2^-) - F(x_1^-).$$

$$9. \quad P(x_1 < X < x_2) = F(x_2^-) - F(x_1).$$

Si noti che se $F(x)$ è continua (cioè se la variabile aleatoria è continua, cfr. § 3.2.2), i limiti da sinistra e da destra coincidono tutti con il valore assunto dalla funzione nel punto, e quindi le probabilità calcolate sulle base delle proprietà 5, 7, 8, 9 sono tutte uguali tra loro, e pari a $F(x_2) - F(x_1)$ (indipendentemente dal fatto che gli estremi appartengano oppure no all'intervallo).

3.2.2 Variabili aleatorie discrete, continue, miste

La variabile aleatoria X si dirà *discreta* se la sua CDF $F(x)$ è una funzione costante a tratti (Figg. 3.3, 3.4, 3.6). In tal caso, detti x_k i punti di discontinuità di $F(x)$, si ha, per la proprietà 6 della CDF,

$$P(X = x) = F(x) - F(x^-) = \begin{cases} 0, & \text{se } x \neq x_k \text{ è un punto di continuità;} \\ p_k, & \text{se } x = x_k \text{ è un punto di discontinuità.} \end{cases}$$

Quindi in pratica una variabile aleatoria discreta X assume i valori x_k con probabilità p_k date dai valori dei salti di discontinuità della sua CDF, e pertanto l'insieme \mathcal{X} dei valori assunti da X è un insieme discreto, cioè $\mathcal{X} = \{x_1, x_2, \dots\}$ (finito o infinito numerabile).

Un caso particolare di variabili aleatorie discrete sono quelle di tipo *reticolare*, caratterizzate dal fatto che i valori assunti x_k sono equispaziati (appartengono, cioè, ad un reticolo monodimensionale), e si può porre quindi $x_k = a + bk$, con $a, b \in \mathbb{R}$ e $k \in K \subseteq \mathbb{Z}$.

La variabile aleatoria X si dirà *continua* se la sua CDF $F(x)$ è una funzione continua (Fig. 3.5). La continuità di $F(x)$ implica che $F(x) = F(x^+) = F(x^-)$ e quindi $P(X = x) = 0, \forall x \in \overline{\mathbb{R}}$. In altri termini, una variabile aleatoria continua assumerà ogni valore del suo codominio con probabilità nulla. L'insieme \mathcal{X} dei valori assunti da una variabile aleatoria continua è un insieme continuo, quale ad esempio un intervallo (a, b) , o anche tutto $\overline{\mathbb{R}}$.

Infine, la variabile aleatoria X si dirà *mista* se la sua CDF $F(x)$ è discontinua, ma non costante a tratti. L'insieme \mathcal{X} dei valori assunti da X sarà l'unione di un insieme continuo, ad esempio un intervallo, e di un insieme discreto (eventualmente vuoto). Esempi di variabili aleatorie miste saranno presentati nel cap. 4.

► **Esempio 3.10 (variabile aleatoria indicatrice di un evento).** Sia Ω uno spazio campione qualunque, e sia $A \subseteq \Omega$ un evento di Ω . Definiamo una variabile aleatoria X_A su Ω come segue:

$$X_A(\omega) = \begin{cases} 1, & \text{se } \omega \in A; \\ 0, & \text{se } \omega \notin A. \end{cases}$$

Tale variabile aleatoria X_A è di tipo *discreto*, in quanto assume solo i valori 0 ed 1 con le seguenti probabilità:

$$\begin{aligned} P(X_A = 1) &= P(\omega \in A) = P(A); \\ P(X_A = 0) &= P(\omega \notin A) = 1 - P(A); \end{aligned}$$

e quindi la sua CDF è a costante a tratti e, se poniamo $p = P(A)$, è la stessa di quella dell'esempio 3.6 (Fig. 3.3). Tale variabile aleatoria prende il nome di *variabile aleatoria indicatrice dell'evento A*. ◀

L'ultimo esempio mette in evidenza che è possibile costruire variabili aleatorie discrete su un qualunque spazio di probabilità (discreto o continuo). Osserviamo infatti che, se Ω è uno spazio di probabilità discreto, tutte le variabili aleatorie costruite su Ω saranno necessariamente discrete. Se invece Ω è uno spazio di probabilità continuo, su di esso è possibile costruire sia variabili aleatorie continue che discrete (e ovviamente anche miste).

3.2.3 Percentile e mediana ★

A partire dalla definizione di CDF, è possibile ricavare direttamente i concetti di *percentile* e *mediana*:

Definizione (percentile). Dato $u \in [0, 1]$, il *percentile u-esimo* di una variabile aleatoria è il più piccolo numero x_u tale che

$$P(X \leq x_u) = F(x_u) \geq u. \quad (3.2)$$

Osserviamo che u deve necessariamente assumere valori in $[0, 1]$, perchè rappresenta un valore di probabilità. L'interpretazione del percentile è la seguente (Fig. 3.7): il percentile x_u rappresenta quel valore della variabile aleatoria che *non* è superato con probabilità maggiore o uguale ad u . Ad esempio, posto $u = 0.75$, il percentile $x_{0.75}$ rappresenta quel valore che non è superato nel 75% o più dei casi, e viene chiamato *quartile superiore*. Similmente, il percentile $x_{0.25}$ rappresenta il valore che non è superato con probabilità maggiore o uguale a 0.25, e viene chiamato *quartile inferiore*.

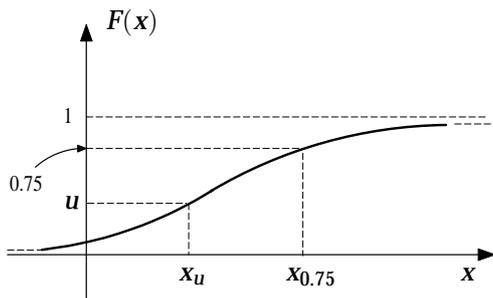


Fig. 3.7. Il percentile u -esimo della variabile aleatoria X con CDF $F(x)$ è x_u ; $x_{0.75}$ rappresenta il valore che non è superato con probabilità maggiore o uguale a 0.75 (quartile superiore).

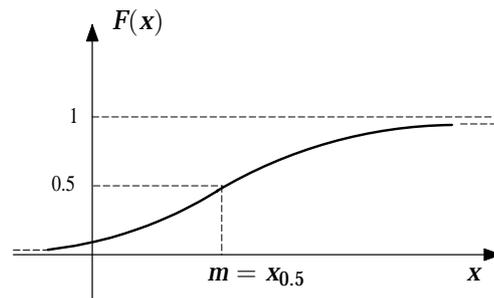


Fig. 3.8. La mediana m della variabile aleatoria X con CDF $F(x)$ è il valore che non è superato con probabilità maggiore o uguale a 0.5 (coincide con il percentile $x_{0.5}$).

Notiamo inoltre che se $F(x)$ assume tutti i valori in $[0, 1]$ (non ha salti di discontinuità, ovvero la variabile aleatoria è continua) allora la definizione (3.2), per la monotonia di $F(x)$, si può scrivere come:

$$P(X \leq x_u) = F(x_u) = u.$$

per cui, se $F(x)$ è anche una funzione *invertibile*, si ha

$$x_u = F^{-1}(u),$$

e quindi la curva che fornisce i percentili si ottiene semplicemente considerando l'inversa della CDF, ovvero scambiando gli assi del diagramma cartesiano di $F(x)$. In pratica le considerazioni precedenti valgono anche se la CDF è localmente invertibile in corrispondenza dell'ordinata u .

In tutti gli altri casi (CDF discontinua, oppure CDF non invertibile, il che accade tipicamente se $F(x)$ presenta uno o più tratti costanti) si può determinare il percentile direttamente applicando la definizione (3.2), ovvero come

$$x_u = \inf\{x \in \overline{\mathbb{R}} \text{ tali che } F(x) \geq u\}. \quad (3.3)$$

Notiamo che la funzione $x_u = F^{-1}(u)$ definita implicitamente dalla (3.3) prende il nome di *inversa sinistra* della CDF $F(x)$, e si riduce all'inversa convenzionale quando la CDF è invertibile (tale inversa sinistra ricorre anche nel problema della generazione di variabili aleatorie discrete, si veda anche il § 4.3.1 per ulteriori dettagli). In pratica l'inversa sinistra si ottiene graficamente scambiando gli assi della CDF, anche quando la CDF non è invertibile. Notiamo peraltro che il concetto di percentile è maggiormente utilizzato quando la variabile aleatoria ha una CDF continua ed invertibile.

Definizione (mediana). La mediana è il percentile per $u = 0.5$, ovvero è il più piccolo numero m che soddisfa la relazione:

$$F(m) \geq 0.5.$$

Per la determinazione della mediana (Fig. 3.8) valgono considerazioni analoghe a quelle effettuate per il percentile, essendo di fatto $m = x_{0.5}$. Osserviamo che la mediana è un primo esempio di grandezza *media* relativa ad una variabile aleatoria: nel seguito incontreremo altre grandezze simili, quali la *moda* e la *media* propriamente detta.

► **Esempio 3.11.** Consideriamo la CDF $F(x)$ dell'esempio 3.8, diagrammata in Fig. 3.5. Poiché l'andamento di $F(x)$, per $x \in [0, T]$, è lineare, è immediato invertirla per ottenere il percentile. Si ha:

$$u = F(x_u) = \frac{x_u}{T} \Rightarrow x_u = T u$$

per cui il quartile inferiore è $x_{0.25} = 0.25 T$, il quartile superiore è $x_{0.75} = 0.75 T$, mentre la mediana è $m = 0.5 T$. ◀

3.3 Funzione densità di probabilità (pdf)

Accanto alla CDF, la funzione densità di probabilità (pdf) gioca un ruolo fondamentale nella descrizione di una variabile aleatoria X :

Definizione (densità di probabilità). La funzione densità di probabilità (pdf) di una variabile aleatoria X è la derivata della CDF $F(x)$:

$$f(x) \triangleq \frac{d}{dx} F(x). \quad (3.4)$$

Per quanto riguarda la notazione, useremo anche qui la notazione $f_X(x)$ quando vorremo specificare esplicitamente che si tratta della pdf di X .

Nella (3.4), la derivata va intesa *in senso generalizzato*, ovvero possono comparire degli *impulsi di Dirac*⁵ in corrispondenza delle discontinuità di $F(x)$. A tale proposito, osserviamo che se la variabile aleatoria X è continua, $F(x)$ è una funzione continua, e quindi la pdf $f(x)$ non può contenere impulsi. Viceversa, se X è discreta, $F(x)$ è costante a tratti, con salti di discontinuità in corrispondenza dei valori x_i : l'ampiezza dei salti di discontinuità rappresenta (per la proprietà 6 della CDF) la probabilità p_i che la variabile aleatoria assuma il valore x_i . Pertanto, derivando tale CDF, si ottiene una pdf costituita da *soliti* impulsi di Dirac, centrati nei valori *discreti* $x_i \in \mathcal{X}$:

$$f(x) = \sum_{x_i \in \mathcal{X}} p_i \delta(x - x_i),$$

dove $p_i \triangleq P(X = x_i)$. Infine, se X è mista, la pdf conterrà una parte continua (la derivata convenzionale) e impulsi di Dirac in corrispondenza delle discontinuità di $F(x)$.

► **Esempio 3.12.** Consideriamo la CDF $F(x)$ dell'esempio 3.6, diagrammata in Fig. 3.3. Poiché si tratta di una variabile aleatoria discreta (CDF costante a tratti), la pdf sarà una somma di impulsi di Dirac. Applicando le proprietà di derivazione dell'impulso di Dirac, si trova

$$f(x) = q \delta(x) + p \delta(x - 1),$$

che è raffigurata in Fig. 3.9. Un risultato simile si ottiene calcolando le pdf associate alle CDF degli esempi 3.7 e 3.9. ◀

⁵Si suppone che il lettore conosca la definizione e le principali proprietà dell'impulso di Dirac; tali proprietà sono comunque brevemente richiamate nell'Appendice D.

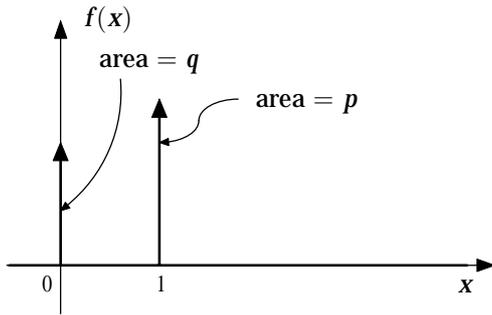


Fig. 3.9. La pdf $f(x)$ della variabile aleatoria dell'esempio 3.12.

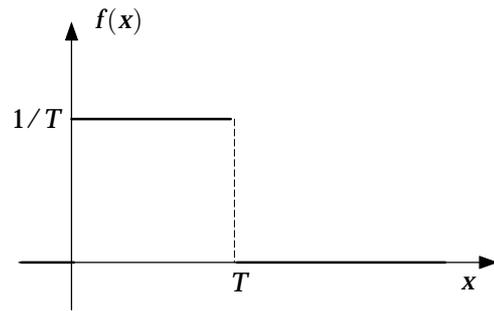


Fig. 3.10. La pdf $f(x)$ della variabile aleatoria dell'esempio 3.13.

► **Esempio 3.13.** Consideriamo la CDF $F(x)$ dell'esempio 3.8, diagrammata in Fig. 3.5. Poichè si tratta di una variabile aleatoria continua (con CDF quindi continua), la pdf non conterrà impulsi di Dirac, ma la derivata si calcolerà in senso ordinario. Si ha:

$$f(x) = \begin{cases} \frac{1}{T}, & \text{se } x \in]0, T[; \\ 0, & \text{se } x \in]-\infty, 0[\cup]T, \infty[; \end{cases}$$

che è raffigurata in Fig. 3.10. Notiamo che la derivata (e quindi la pdf) non è definita nei punti $x = 0$ e $x = T$ (punti angolosi della curva della CDF). Ciò, tuttavia, non rappresenta un problema perchè, come vedremo, la pdf viene utilizzata sempre all'interno di un integrale, e quindi i valori assunti in punti isolati non giocano alcun ruolo (a patto, ovviamente, che in tali punti non siano presenti impulsi di Dirac). ◀

3.3.1 Proprietà della pdf

Di seguito sono elencate le principali proprietà della pdf di una variabile aleatoria X :

1. $f(x) \geq 0.$

Prova. La proprietà discende dal fatto che $F(x)$ è una funzione monotona crescente, e quindi la sua derivata è non negativa. In corrispondenza delle discontinuità, la proprietà va interpretata nel senso che gli impulsi di Dirac ivi presenti hanno *area* positiva. □

2. $F(x) = \int_{-\infty}^x f(y) dy.$

Prova. Poichè $f(x) \triangleq \frac{d}{dx}F(x)$, integrando ambo i membri si ha:

$$\int_{-\infty}^x f(y) dy = \int_{-\infty}^x \frac{d}{dy}F(y) dy = F(x) - F(-\infty).$$

Ma $F(-\infty) = 0$ per cui ho l'asserto. □

3. $\int_{-\infty}^{\infty} f(x) dx = 1.$

Prova. Dalla proprietà 2, per $x = +\infty$, si ha l'asserto, tenendo conto che $F(+\infty) = 1$. □

Tale proprietà prende il nome di *proprietà di normalizzazione della pdf*.

$$4. \quad P(x_1 < X \leq x_2) = F(x_2) - F(x_1) = \int_{x_1}^{x_2} f(x) dx.$$

Prova. Per la proprietà 5 della CDF, e per la proprietà 2 precedente, si ha:

$$P(x_1 < X \leq x_2) = F(x_2) - F(x_1) = \int_{-\infty}^{x_2} f(y) dy - \int_{-\infty}^{x_1} f(y) dy = \int_{x_1}^{x_2} f(y) dy,$$

come volevasi dimostrare. \square

La proprietà va impiegata con qualche cautela nel caso in cui la pdf $f(x)$ contenga impulsi di Dirac (e quindi se la variabile aleatoria X è discreta oppure mista); in particolare, in accordo con il fatto che si sta calcolando la probabilità dell'evento $\{x_1 < X \leq x_2\}$, l'integrale tra x_1 ed x_2 della pdf va inteso come $\lim_{\varepsilon \rightarrow 0} \int_{x_1+\varepsilon}^{x_2} f(x) dx$, con $\varepsilon > 0$, in maniera da *non* portare in conto l'eventuale presenza di un impulso in x_1 , mentre un (eventuale) impulso in x_2 va portato in conto. Se viceversa la variabile aleatoria X è continua, la pdf $f(x)$ non contiene impulsi di Dirac e quindi $P(X = x_1) = 0$, per cui $P(x_1 \leq X \leq x_2) = P(X = x_1) + P(x_1 < X \leq x_2) = \int_{x_1}^{x_2} f(x) dx$. Più in generale, per variabili aleatorie continue, la probabilità di eventi del tipo $\{X \in (x_1, x_2)\}$ è la stessa, indipendentemente dal fatto che gli estremi x_1 ed x_2 si considerino appartenenti all'intervallo oppure no, e si calcola integrando (in senso ordinario) la pdf tra x_1 ed x_2 .

$$5. \quad X \text{ continua, con pdf } f(x) \text{ continua} \Rightarrow P(x \leq X \leq x + \Delta x) \approx f(x) \Delta x, \text{ per } \Delta x \ll 1.$$

Prova. Dalla proprietà 4, ponendo $x_1 = x$ e $x_2 = x + \Delta x$, ed osservando che per una variabile aleatoria continua la probabilità non cambia se includiamo il limite sinistro x_1 oppure no, si ha:

$$P(x \leq X \leq x + \Delta x) = P(x < X \leq x + \Delta x) = \int_x^{x+\Delta x} f(y) dy$$

Per l'ipotesi di continuità della pdf $f(x)$, possiamo applicare il teorema della media per il calcolo integrale:

$$P(x \leq X \leq x + \Delta x) = \int_x^{x+\Delta x} f(y) dy = f(x + \theta \Delta x) \Delta x \approx f(x) \Delta x.$$

con $\theta \in [0, 1]$, dove l'ultima approssimazione vale per $\Delta x \ll 1$. \square

Notiamo che quest'ultima proprietà giustifica il nome di *densità* di probabilità: infatti da essa discende che, se $f(x)$ è continua, si ha:

$$f(x) = \lim_{\Delta x \rightarrow 0} \frac{P(x \leq X \leq x + \Delta x)}{\Delta x} \approx \frac{P(x \leq X \leq x + \Delta x)}{\Delta x}$$

e quindi il valore $f(x)$ nel punto x rappresenta la probabilità che X assuma valori in un intervallo $(x, x + \Delta x)$ prossimo a x , divisa per l'ampiezza dell'intervallo Δx , cioè proprio una *densità* di probabilità. Per questo motivo, poiché $f(x)$ è una densità di probabilità e non una probabilità, può assumere valori maggiori di 1.

Osserviamo inoltre che, per la stessa proprietà, la probabilità che $X \in [x, x + \Delta x]$ è proporzionale (se $\Delta x \ll 1$) a $f(x)$ ed è (localmente) massima se $[x, x + \Delta x]$ contiene il valore x_m dove $f(x)$ è (localmente) massima. Ognuno di tali punti x_m si dice *valore modale* o *moda*,

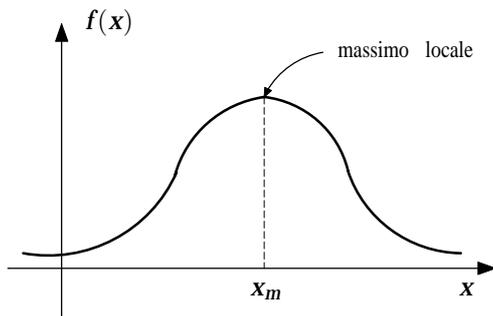


Fig. 3.11. La moda x_m della variabile aleatoria X corrisponde ad un massimo locale. La pdf $f(x)$ in figura ha una sola moda, quindi è unimodale.

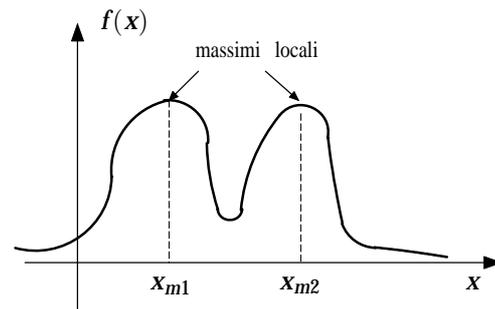


Fig. 3.12. La pdf $f(x)$ in figura ha due mode x_{m1} ed x_{m2} , corrispondenti a due massimi locali, quindi è multimodale.

e rappresenta un valore (localmente) più probabile di X (Fig. 3.11). Una variabile aleatoria si dice *unimodale* se ha un solo valore modale (Fig. 3.11), altrimenti si dice *multimodale* (Fig. 3.12).

Notiamo in conclusione che, come già osservato, definire una variabile aleatoria significa, in sostanza, sostituire allo spazio di probabilità (Ω, \mathcal{S}, P) un nuovo spazio di probabilità, in cui lo spazio campione è $\mathcal{X} \subseteq \overline{\mathbb{R}}$. Se, in particolare, la variabile aleatoria è continua, allora $\mathcal{X} \subseteq \overline{\mathbb{R}}$ è un insieme continuo, per cui la definizione di una legge di probabilità su tale insieme avviene, come descritto nel § 1.6.2, definendo una funzione densità di probabilità $f(x)$ che, di fatto, possiamo adesso interpretare come la pdf di una variabile aleatoria X (si noti in particolare che vale la condizione di normalizzazione (1.13) tipica delle pdf). In definitiva, allora, definire una legge di probabilità su uno spazio continuo è equivalente ad assegnare la pdf di una variabile aleatoria X . Il lettore è invitato a rileggere il § 1.6.2 alla luce di questa interpretazione.

3.4 Funzione distribuzione di probabilità (DF)

Abbiamo visto che, se X è una variabile aleatoria discreta, essa assume solo i valori $x_i \in \mathcal{X}$ con probabilità p_i , e pertanto la sua pdf è di tipo puramente impulsivo (cfr. esempio 3.6). In tal caso, appare più semplice e immediato, in luogo della CDF o pdf, definire una funzione che restituisca direttamente le probabilità con cui la variabile aleatoria assume i suoi valori. Tale funzione prende il nome di *funzione distribuzione di probabilità* (DF):

Definizione (distribuzione di probabilità). La funzione distribuzione di probabilità (DF) di una variabile aleatoria *discreta* X a valori in \mathcal{X} è:

$$p(x) = P(X = x) \quad (3.5)$$

con $x \in \mathcal{X}$.

Anche per la DF, come per la CDF e la pdf, utilizzeremo la notazione $p_X(x)$ quando vorremo esplicitamente denotare che essa si riferisce alla variabile aleatoria X .

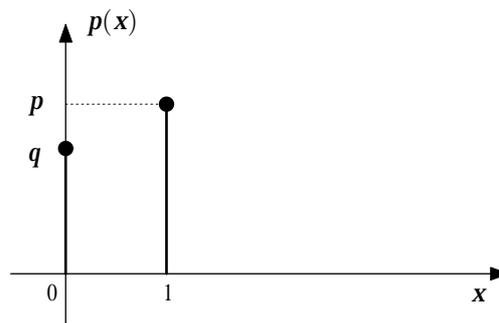


Fig. 3.13. La DF $p(x)$ della variabile aleatoria dell'esempio 3.14.

► *Esempio 3.14.* Si consideri la variabile aleatoria dell'esempio 3.6, che assume i due valori $X = 1$ con probabilità p e $X = 0$ con probabilità q . La DF di X è:

$$p(x) = \begin{cases} q, & x = 0; \\ p, & x = 1; \end{cases}$$

ed è raffigurata in Fig. 3.13. Si noti che la pdf (Fig. 3.9) della stessa variabile aleatoria è:

$$f(x) = q\delta(x) + p\delta(x-1)$$

Il vantaggio nell'uso della DF è quello di disporre di una funzione ordinaria che non contiene impulsi di Dirac. ◀

Notiamo esplicitamente che per una variabile aleatoria continua non ha senso introdurre la DF, in quanto essa risulterebbe identicamente nulla, $\forall x \in \mathcal{X}$, perché una variabile aleatoria continua assume tutti i valori di \mathcal{X} con probabilità zero. Per lo stesso motivo, la DF fornisce una descrizione incompleta di una variabile aleatoria mista, e non è utilizzata neanche in questo caso.

3.4.1 Proprietà della DF

La DF gode delle seguenti proprietà, che presentano una stretta analogia con quelle della pdf:

1. $p(x) \geq 0.$

Prova. La prova è banale perchè $p(x)$ è una probabilità. ◻

2. $F(x) = \sum_{u \in \mathcal{X}, u \leq x} p(u).$

Prova. Si ha, sfruttando le proprietà elementari della probabilità,

$$F(x) = P(X \leq x) = P\left(\bigcup_{\substack{u \in \mathcal{X} \\ u \leq x}} \{X = u\}\right) = \sum_{u \in \mathcal{X}, u \leq x} P(X = u) = \sum_{u \in \mathcal{X}, u \leq x} p(u).$$

◻

$$3. \quad \sum_{u \in \mathcal{X}} p(u) = 1.$$

Prova. Si ricava dalla precedente; infatti:

$$F(+\infty) = 1 = \sum_{u \in \mathcal{X}, u \leq +\infty} p(u) = \sum_{u \in \mathcal{X}} p(u).$$

□

$$4. \quad P(x_1 < X \leq x_2) = \sum_{u \in]x_1, x_2] \cap \mathcal{X}} p(u).$$

Prova. Si ha:

$$P(x_1 < X \leq x_2) = P\left(\bigcup_{u \in]x_1, x_2] \cap \mathcal{X}} \{X = u\}\right) = \sum_{u \in]x_1, x_2] \cap \mathcal{X}} P(X = u) = \sum_{u \in]x_1, x_2] \cap \mathcal{X}} p(u).$$

□

Concludiamo osservando che la CDF, pdf e DF di una variabile aleatoria sono collettivamente denominate *funzioni di distribuzione* della variabile aleatoria: per *caratterizzazione statistica* di una variabile aleatoria, allora, si intende la conoscenza di almeno una tra le sue funzioni di distribuzione.

3.5 Variabili aleatorie notevoli

Nel corso di questo capitolo, abbiamo introdotto le variabili aleatorie come funzioni definite su uno spazio campione Ω , dotato di struttura di spazio di probabilità. Tale definizione richiede l'individuazione esplicita di un esperimento aleatorio e la descrizione dello spazio di probabilità costruito su di esso. D'altra parte, nella pratica spesso si introducono variabili aleatorie semplicemente assegnando le loro funzioni di distribuzione: tale semplificazione è possibile in virtù del seguente *teorema di esistenza*, che enunciamo senza dimostrazione (gli interessati vedano [3, cap. 4.3]):

Teorema 3.1 (teorema di esistenza). Data una funzione $F(x)$ che soddisfa le proprietà di CDF (o alternativamente data una funzione $f(x)$ tale che $F(x) = \int_{-\infty}^x f(y) dy$ soddisfi le proprietà di CDF, o una funzione $p(x)$ tale che $F(x) = \sum_{u \in \mathcal{X}, u \leq x} p(u)$ soddisfi le proprietà di CDF), è possibile costruire uno spazio di probabilità (Ω, \mathcal{S}, P) ed una variabile aleatoria X con CDF $F(x)$ (o pdf $f(x)$, o DF $p(x)$).

Sulla base di questo teorema, potremo o costruire la variabile aleatoria su un determinato spazio di probabilità, oppure in alternativa introdurre direttamente le variabili aleatorie attraverso le loro funzioni di distribuzione (CDF, pdf o DF), senza specificare esplicitamente l'esperimento su cui sono definite.

Nel seguito del paragrafo introdurremo alcune delle variabili aleatorie più comunemente utilizzate. Per le variabili discrete, riporteremo la descrizione in termini di funzione di distribuzione di probabilità (DF), lasciando per esercizio al lettore di ricavare le corrispondenti pdf e CDF, peraltro scarsamente utilizzate nel caso discreto. Notiamo preliminarmente che tutte le

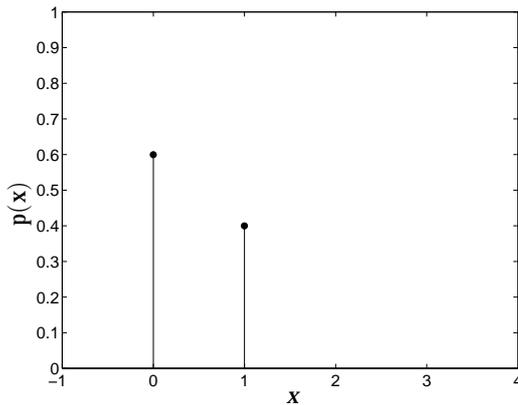


Fig. 3.14. La DF $p(x)$ di una variabile aleatoria bernoulliana ($p = 0.4$).

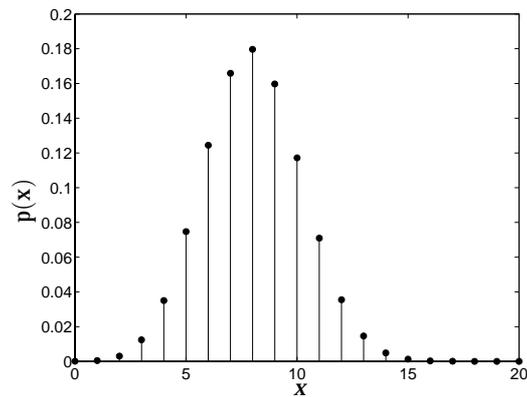


Fig. 3.15. La DF $p(x)$ di una variabile aleatoria binomiale ($n = 20$, $p = 0.4$).

variabili aleatorie discrete che introdurremo saranno di tipo *reticolare*. A differenza di quelle discrete, le variabili aleatorie continue saranno descritte attraverso la pdf e la CDF (risultando la DF identicamente nulla).

3.5.1 Variabile aleatoria di Bernoulli

La variabile aleatoria X si dice di Bernoulli o bernoulliana, e si denota $X \sim \text{Bern}(p)$, se essa assume il valore 1 con probabilità p ed il valore 0 con probabilità $q = 1 - p$ ($\mathcal{X} = \{0, 1\}$), per cui la sua DF è (Fig. 3.14):

$$p(k) = \begin{cases} q, & \text{se } k = 0; \\ p, & \text{se } k = 1. \end{cases}$$

Una variabile aleatoria di Bernoulli si può anche interpretare come variabile aleatoria indicatrice di un evento A che si verifica con probabilità p (vedi esempio 3.10). Notiamo infine che una variabile aleatoria di Bernoulli è un caso particolare (per $n = 1$) della variabile aleatoria binomiale, discussa nel paragrafo seguente.

3.5.2 Variabile aleatoria binomiale e problema delle prove ripetute

Anziché fornire direttamente la sua DF, in questo caso è istruttivo mostrare come la variabile aleatoria binomiale si possa costruire su uno spazio di probabilità sufficientemente generale ed applicabile alla descrizione di numerosi problemi pratici. Tale spazio di probabilità fa riferimento al cosiddetto problema delle *prove ripetute*, per il cui studio si applicano i concetti di esperimento combinato (cfr. § 2.4) nonché di indipendenza statistica (cfr. § 2.3).

Si consideri un esperimento, descritto dallo spazio di probabilità $(\Omega_1, \mathcal{S}_1, P_1)$, e si supponga di ripeterlo n volte, nelle medesime condizioni, assumendo che le successive ripetizioni dell'esperimento siano *indipendenti*. Lo spazio campione dell'esperimento combinato sarà evidentemente

$$\Omega = \Omega_1^n = \underbrace{\Omega_1 \times \Omega_1 \cdots \times \Omega_1}_{n \text{ volte}},$$

il σ -campo \mathcal{S} sarà il più piccolo σ -campo contenente gli eventi del tipo $A = A_1 \times A_2 \cdots \times A_n$, con $A_i \in \Omega_1$, ed infine la legge di probabilità P , nell'ipotesi di prove indipendenti, è indotta dalla

legge P_1 . Infatti, per qualunque evento $A \in \mathcal{S}$ che si possa esprimere come prodotto cartesiano del tipo precedentemente visto, risulta

$$P(A) = P_1(A_1) P_1(A_2) \cdots P_1(A_n).$$

La probabilità di un qualunque altro evento di \mathcal{S} si può ottenere facilmente a partire dalle probabilità del tipo precedente, ovvero utilizzando la proprietà di fattorizzazione. In particolare, se Ω_1 (e quindi Ω) è un insieme discreto, è possibile calcolare la probabilità degli eventi elementari $\omega = (\omega_1, \omega_2, \dots, \omega_N)$ come $P(\omega) = P_1(\omega_1) P_1(\omega_2) \cdots P_1(\omega_N)$ e quindi, a partire da esse, la probabilità di un qualunque evento di Ω .

Consideriamo ora il caso particolarmente interessante delle prove cosiddette *di Bernoulli* o *bernoulliane*,⁶ in cui l'attenzione si concentra su un evento A di Ω_1 (convenzionalmente denominato *successo*), che si verifica con probabilità $p = P(A)$; ovviamente, l'evento complementare \bar{A} (denominato *insuccesso*) si verificherà con probabilità $q = 1 - P(A) = 1 - p$. Data la natura essenzialmente *binaria* (successo/insuccesso) del problema, possiamo darne una descrizione estremamente semplificata, ottenuta utilizzando come spazio campione $\Omega_1 = \{0, 1\}$, in cui convenzionalmente associamo al successo il valore 1 e all'insuccesso il valore 0. In questo caso, lo spazio campione $\Omega = \Omega_1^n$ dell'esperimento combinato è rappresentato da tutte le stringhe binarie di lunghezza n , in numero pari evidentemente a 2^n . Costruiamo una variabile aleatoria X su Ω nel seguente modo: a ciascuna stringa binaria $\omega \in \Omega$ associamo il numero di 1 contenuti nella stringa, denominato anche *peso di Hamming* $p_H(\omega)$ della stringa: ad esempio, se $n = 8$ si ha:

$$\begin{aligned} \omega = 00000000 &\rightarrow X(\omega) = p_H(\omega) = 0 \\ \omega = 01100110 &\rightarrow X(\omega) = p_H(\omega) = 4 \\ \omega = 11100100 &\rightarrow X(\omega) = p_H(\omega) = 4 \\ \omega = 11111111 &\rightarrow X(\omega) = p_H(\omega) = 8 \end{aligned} \tag{3.6}$$

Per costruzione, la variabile aleatoria X assume lo stesso valore in corrispondenza di tutte le stringhe aventi lo stesso numero di 1, ovvero lo stesso peso di Hamming; pertanto determinare la DF $p(x)$ della variabile aleatoria X richiede senz'altro la determinazione del *numero* di tali stringhe. Notiamo che la variabile aleatoria X assume valori nell'insieme $\mathcal{X} = \{0, 1, \dots, n\}$ e che, ritornando all'interpretazione in termini di successi ed insuccessi, $p(k) \triangleq P(X = k)$ rappresenta la probabilità che, nelle n prove ripetute, *si abbiano esattamente k successi, in un qualunque ordine*.

Per capire come determinare il numero di configurazioni (stringhe) con $k \in \{0, 1, \dots, n\}$ successi (valori 1), consideriamo un esempio specifico. Se $n = 4$ e $k = 2$, l'evento A si verifica in 2 delle 4 prove, ed evidentemente nelle altre $n - k = 2$ prove si verificherà \bar{A} . Ovviamente l'evento A si potrà verificare nella prima e nella seconda prova, nella prima e nella terza, nella prima e nella quarta, nella seconda e nella terza, etc. Tutte le possibili configurazioni (sei, in questo caso)

⁶La denominazione deriva dal matematico svizzero J. Bernoulli (1654–1705), autore del fondamentale trattato di probabilità "Ars Conjectandi".

sono riportate di seguito (insieme con la loro rappresentazione “binaria”)

$$\begin{aligned}
 A \times A \times \bar{A} \times \bar{A} &\rightarrow 1100 \\
 A \times \bar{A} \times A \times \bar{A} &\rightarrow 1010 \\
 A \times \bar{A} \times \bar{A} \times A &\rightarrow 1001 \\
 \bar{A} \times A \times A \times \bar{A} &\rightarrow 0110 \\
 \bar{A} \times A \times \bar{A} \times A &\rightarrow 0101 \\
 \bar{A} \times \bar{A} \times A \times A &\rightarrow 0011
 \end{aligned}$$

In generale, per determinare il numero delle possibili configurazioni, posso ragionare come segue: ho n oggetti (le prove), e devo specificarne k (le prove in cui si hanno i successi), senza sostituzioni e senza ordinamento; pertanto il numero di possibili configurazioni è pari al numero delle disposizioni di n oggetti su k posti senza sostituzioni e senza ordinamento (cfr. Tab. B.1), espressa da $\binom{n}{k}$, che nel caso in esame vale appunto $\binom{4}{2} = 6$.⁷ Data l'indipendenza delle prove, la probabilità di una qualsiasi configurazione di k successi ed $n - k$ insuccessi vale sempre $p^k q^{n-k}$. Ad esempio,

$$P(A \times A \times \bar{A} \times \bar{A}) = P(1100) = P_1(A) P_1(A) P_1(\bar{A}) P_1(\bar{A}) = p^2 q^2.$$

Poichè le $\binom{n}{k}$ configurazioni con k successi sono tutte differenti, esse corrispondono ad eventi mutuamente esclusivi, ed allora la probabilità cercata si riduce alla *somma* delle probabilità di tutte le configurazioni. Le configurazioni sono tutte equiprobabili, con probabilità $p^k q^{n-k}$, ed in numero pari a $\binom{n}{k}$, per cui la $P(X = k)$ ovvero la DF della variabile aleatoria X è data da

$$p(k) = \binom{n}{k} p^k q^{n-k}, \quad k \in \mathcal{X} = \{0, 1, \dots, n\},$$

con $q = 1 - p$. Una variabile aleatoria avente tale DF si dice binomiale di parametri $n > 0$ e $p \in [0, 1]$, e si denota $X \sim B(n, p)$. I valori della DF hanno somma unitaria (proprietà 3 della DF), come si può facilmente provare utilizzando il teorema binomiale (cfr. Appendice A). Osserviamo che una variabile aleatoria bernoulliana si può riguardare come un caso particolare (per $n = 1$) della variabile aleatoria binomiale, ovvero le notazioni $X \sim \text{Bern}(p)$ e $X \sim B(1, p)$ sono equivalenti.

L'andamento della DF binomiale al variare di k , per $n = 20$ e $p = 0.4$, è illustrato in Fig. 3.15. Dalla Fig. 3.15 si può notare che, al variare di k , la $p(k)$ prima cresce, poi decresce, presentando un massimo per $k = np = 8$. Un'analisi più rigorosa mostra che il massimo si trova, in generale, in $k = \lfloor (n+1)p \rfloor$, dove con il simbolo $\lfloor x \rfloor$ denotiamo il più grande intero non superiore ad x . Se, tuttavia, $(n+1)p$ è intero, allora $p(k)$ è massima per due consecutivi valori di k , dati da $k_1 = (n+1)p - 1$ e $k_2 = (n+1)p$.

Il modello delle prove ripetute e la variabile aleatoria binomiale possono essere applicati a numerosi problemi pratici, come illustrato dai seguenti esempi.

► **Esempio 3.15.** Un'azienda produce componenti elettronici in lotti da $n = 1000$ componenti. La probabilità che un componente sia difettoso è pari a $p = 10^{-1}$, indipendentemente dagli altri. Qual è la probabilità che:

- il numero di componenti difettosi di un lotto sia pari a zero;
- il numero di componenti difettosi di un lotto sia minore o uguale a 80;

⁷La definizione di coefficiente binomiale $\binom{n}{k}$ e di fattoriale $n!$, insieme con le principali proprietà, sono riportate in Appendice A.

- il numero di componenti difettosi di un lotto sia compreso tra 80 e 120.

Se interpretiamo come “successo” è l’evento che il componente sia difettoso, abbiamo proprio un problema di prove ripetute, con $n = 1000$. Pertanto, il numero di componenti difettosi si può modellare come una variabile aleatoria binomiale $X \sim B(1000, 10^{-1})$. La probabilità che nessun componente sia difettoso è allora data da:

$$P(X = 0) = p(0) = \binom{1000}{0} p^0 q^{1000} = q^{1000} = 0.9^{1000} \approx 1.75 \cdot 10^{-46}$$

cioè del tutto trascurabile. La probabilità che il numero di componenti difettosi sia minore o uguale a 80 si calcola facilmente, in quanto si ha $\{X \leq 80\} = \cup_{k=0}^{80} \{X = k\}$. Poichè gli eventi elementari sono mutuamente esclusivi, la probabilità dell’unione è pari alla somma delle probabilità, e si ha:⁸

$$P(X \leq 80) = P\left(\cup_{k=0}^{80} \{X = k\}\right) = \sum_{k=0}^{80} P(X = k) = \sum_{k=0}^{80} p(k) = \sum_{k=0}^{80} \binom{1000}{k} p^k q^{1000-k} \approx 0.0176.$$

Infine, l’evento che X sia compreso tra 80 e 120 può essere anch’esso espresso come unione di eventi elementari mutuamente esclusivi, ovvero come $\{80 \leq X \leq 120\} = \cup_{k=80}^{120} \{X = k\}$, per cui

$$P(80 \leq X \leq 120) = P\left(\cup_{k=80}^{120} \{X = k\}\right) = \sum_{k=80}^{120} P(X = k) = \sum_{k=80}^{120} p(k) = \sum_{k=80}^{120} \binom{1000}{k} p^k q^{1000-k} \approx 0.9695.$$

◀

► **Esempio 3.16.** Un test a risposte multiple prevede $n = 20$ domande, con tre possibili risposte per ciascuna domanda. Uno studente poco preparato risponde a caso a tutte le domande; qual è la probabilità che totalizzi un punteggio maggiore o uguale a 12, che è la soglia minima per l’ammissione?

Anche qui possiamo ricondurre il problema allo schema delle prove ripetute. Rispondendo a caso a ciascuna domanda, lo studente individuerà la risposta esatta (successo) con probabilità $p = 1/3$, e sbaglierà (insuccesso) con probabilità $q = 1 - p = 2/3$. Pertanto, il numero di risposte esatte è una variabile aleatoria binomiale $X \sim B(20, 1/3)$, e quindi la probabilità cercata, con considerazioni analoghe a quelle dell’esempio precedente è data da:

$$P(X \geq 12) = \sum_{k=12}^{20} p(k) = \sum_{k=12}^{20} \binom{20}{k} p^k q^{20-k} \approx 0.0130,$$

che è una probabilità inferiore al 2%, per cui è estremamente difficile che lo studente superi il test, rispondendo a casaccio alle domande. ◀

3.5.3 Variabile aleatoria binomiale negativa

La variabile aleatoria X si dice binomiale negativa di parametri $r > 0$ e $p \in [0, 1]$, e si denota $X \sim NB(r, p)$, se la sua DF (Fig. 3.16) è la seguente:

$$p(k) = \binom{r+k-1}{k} p^r q^k, \quad k \in \mathcal{X} = \{0, 1, \dots\} = \mathbb{N}_0,$$

con $q = 1 - p$. Il nome *binomiale negativa* discende dal fatto che, per provare che la somma dei valori della DF è pari ad uno (proprietà 3 della DF), è necessario sfruttare l’espansione binomiale negativa (equazione (A.5) in Appendice A). Infatti, si ha:

$$\sum_{k=0}^{\infty} p(k) = p^r \sum_{k=0}^{\infty} \binom{r+k-1}{k} q^k = p^r (1 - q)^{-r} = 1.$$

⁸Per effettuare il calcolo numerico che segue, come anche per gli altri che ricorrono in questo esempio ed in quello seguente, è indispensabile ricorrere ad un calcolatore, ad esempio scrivendo un semplice script Matlab. In alternativa, si veda il § 3.5.12 per un’interessante approssimazione numerica.

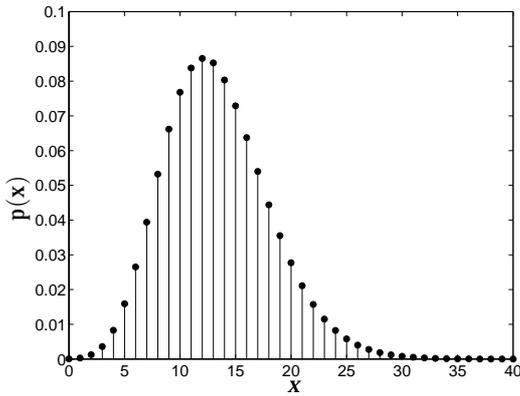


Fig. 3.16. La DF $p(x)$ di una variabile aleatoria binomiale negativa ($n = 20$, $p = 0.6$).

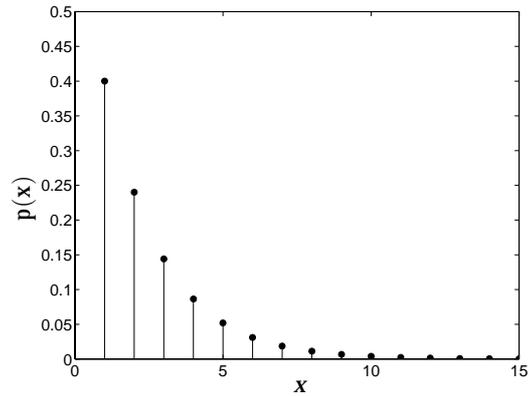


Fig. 3.17. La DF $p(x)$ di una variabile aleatoria geometrica ($p = 0.4$).

► **Esempio 3.17.** Come la variabile aleatoria binomiale, anche la variabile aleatoria binomiale negativa è associata al problema delle prove ripetute. Supponiamo infatti di voler calcolare la distribuzione di probabilità della variabile aleatoria Y che rappresenta la prova in cui si verifica l' r -esimo successo. Evidentemente, Y potrà assumere i valori $r, r + 1, r + 2, \dots$, in quanto, per avere r successi, è necessario effettuare almeno r prove. D'altra parte, l' r -esimo successo si verificherà nella prova $h \geq r$ se e solo se si verificano i seguenti due eventi:

- $A = \{\text{nelle } h - 1 \text{ prove precedenti, si hanno } r - 1 \text{ successi}\}$; tale evento ha una probabilità, descritta dalla legge binomiale, pari a $P(A) = \binom{h-1}{r-1} p^{r-1} q^{h-r}$;
- $B = \{\text{nella } h\text{-esima prova, si ha un successo}\}$; tale evento ha probabilità pari a $P(B) = p$.

Poiché gli eventi A e B sono indipendenti, si ha:

$$P(Y = h) = P(A) P(B) = \binom{h-1}{r-1} p^{r-1} q^{h-r} p = \binom{h-1}{r-1} p^r q^{h-r}$$

per $h = r, r + 1, \dots$, che può essere semplicemente espressa in termini di una variabile aleatoria binomiale negativa. Infatti, poiché $h \geq r$, basta porre $h = r + k$, con $k \geq 0$, e far riferimento ad una variabile $X = Y - r$ che può assumere i valori $k = 0, 1, \dots$. Si ha in tal caso:

$$P(X = k) = P(Y = r + k) = \binom{r+k-1}{k} p^r q^k$$

dove si sono sfruttate le proprietà del coefficiente binomiale (cfr. Appendice A). Notiamo allora che risulta $X \sim \text{NB}(r, p)$, che possiamo interpretare allora come la distribuzione del numero di prove che bisogna effettuare, successivamente alla r -esima, per ottenere l' r -esimo successo. ◀

3.5.4 Variabile aleatoria geometrica

La variabile aleatoria X si dice geometrica di parametro $p \in [0, 1]$, e si denota $X \sim \text{Geom}(p)$, se la sua DF (Fig. 3.17) è la seguente:

$$p(k) = p q^{k-1} \quad k \in \mathcal{X} = \{1, 2, \dots\} = \mathbb{N},$$

con $q = 1 - p$. Per provare che i valori della DF hanno somma unitaria, basta sfruttare la formula per la somma di una serie geometrica:

$$\sum_{k=1}^{\infty} p(k) = p \sum_{k=1}^{\infty} q^{k-1} = p \sum_{k=0}^{\infty} q^k = \frac{p}{1 - q} = 1.$$

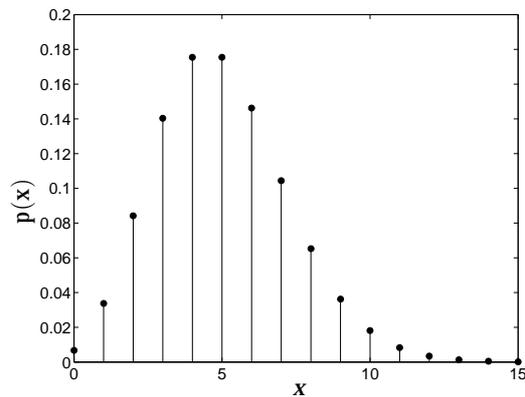


Fig. 3.18. La DF $p(x)$ di una variabile aleatoria di Poisson ($\lambda = 5$).

► *Esempio 3.18.* Come la variabile aleatoria binomiale negativa, anche la variabile aleatoria geometrica è associata al problema delle prove ripetute. Infatti, se denotiamo con X il numero di prove che intercorrono tra due successi consecutivi, tale variabile aleatoria assumerà valori in $1, 2, \dots$. Evidentemente, ci saranno k prove tra due successi consecutivi se e solo se si presenterà una sequenza di $k - 1$ insuccessi seguiti da un successo, il che avviene con probabilità $q^{k-1} p$, data l'indipendenza tra i successi in prove distinte. Pertanto, X è proprio una variabile aleatoria geometrica $X \sim \text{Geom}(p)$. ◀

3.5.5 Variabile aleatoria di Poisson

La variabile aleatoria X si dice di Poisson⁹ o poissoniana di parametro $\lambda > 0$, e si denota $X \sim \text{Poiss}(\lambda)$, se la sua DF (Fig. 3.18) è la seguente:

$$p(k) = \frac{\lambda^k}{k!} e^{-\lambda}, \quad k \in \mathcal{X} = \{0, 1, \dots\} = \mathbb{N}_0.$$

Sfruttando lo sviluppo in serie di Mc-Laurin della funzione esponenziale, si dimostra facilmente che i valori della DF hanno somma unitaria:

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

È possibile mostrare che anche la variabile aleatoria di Poisson è legata al problema delle prove ripetute; in particolare, essa rappresenta la forma limite della variabile aleatoria binomiale per p piccolo al divergere di n , ottenuta mantenendo il prodotto $\lambda = np$ costante (vedi [1, pagg. 153–154]).

3.5.6 Variabile aleatoria uniforme

La variabile aleatoria X si dice uniforme nell'intervallo (a, b) , con $a \leq b$, e si denota $X \sim U(a, b)$, se la sua pdf è (Fig. 3.19):

$$f(x) = \begin{cases} \frac{1}{b-a}, & x \in [a, b]; \\ 0, & \text{altrove.} \end{cases}$$

⁹Dallo studioso Siméon D. Poisson (1781–1840).

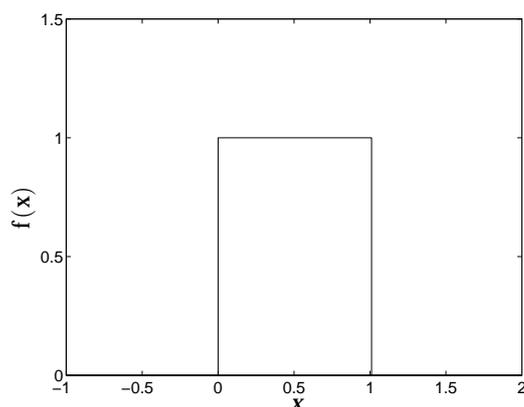


Fig. 3.19. La pdf $f(x)$ di una variabile aleatoria uniforme ($a = 0$, $b = 1$).

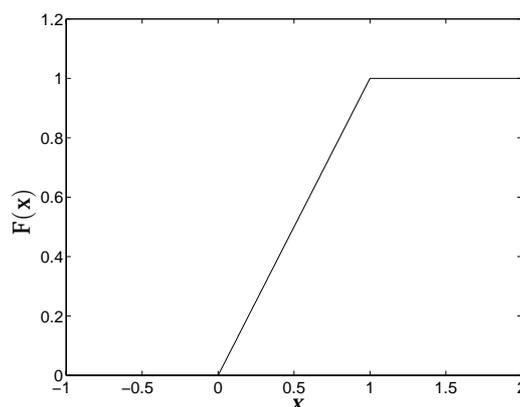


Fig. 3.20. La CDF $F(x)$ di una variabile aleatoria uniforme ($a = 0$, $b = 1$).

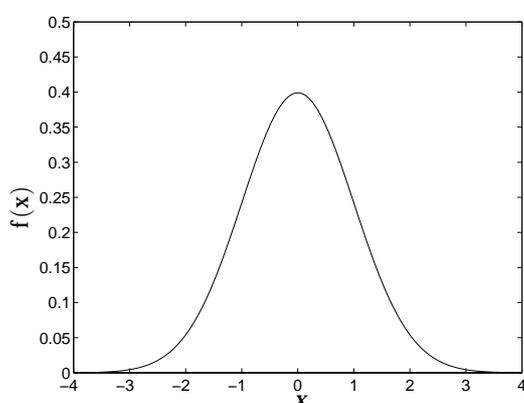


Fig. 3.21. La pdf $f(x)$ di una variabile aleatoria gaussiana ($\mu = 0$, $\sigma = 1$).

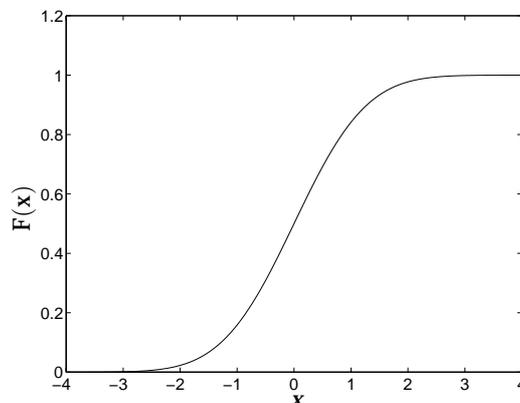


Fig. 3.22. La CDF $F(x)$ di una variabile aleatoria gaussiana ($\mu = 0$, $\sigma = 1$).

La CDF (Fig. 3.20) si calcola facilmente per integrazione, e vale:

$$F(x) = \begin{cases} 0, & x \in]-\infty, a[; \\ \frac{x-a}{b-a}, & x \in [a, b]; \\ 1, & x \in]b, \infty[. \end{cases}$$

3.5.7 Variabile aleatoria gaussiana o normale

La variabile aleatoria X si dice gaussiana o normale, e si denota $X \sim N(\mu, \sigma)$, se la sua pdf (Fig. 3.21) è:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad (3.7)$$

con $\mu, \sigma \in \mathbb{R}$ e $\sigma > 0$. La forma della pdf gaussiana (Fig. 3.21) è quella tipica di una “campana”, centrata in μ e la cui larghezza è governata dal parametro σ : a valori elevati di σ corrisponde una campana “larga”, mentre a valori piccoli di σ corrisponde una campana “stretta”.

Osserviamo che non è possibile calcolare la CDF associata alla (3.7) in forma chiusa, ma è solo

possibile scrivere:

$$F(x) = \int_{-\infty}^x \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(y-\mu)^2}{2\sigma^2}} dy = \mathbb{G}\left(\frac{x-\mu}{\sigma}\right) \quad (3.8)$$

dove, dopo un banale cambio di variabile, abbiamo espresso la $F(x)$ (Fig. 3.22) in termini della funzione $\mathbb{G}(x)$:

$$\mathbb{G}(x) \triangleq \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{y^2}{2}} dy.$$

In particolare, dalla (3.8), notiamo che $\mathbb{G}(x)$ rappresenta la CDF di una variabile aleatoria gaussiana “standard” con $\mu = 0$ e $\sigma = 1$.

Dobbiamo osservare, tuttavia, che la funzione $\mathbb{G}(x)$ non è comunque una funzione elementare, per cui, per determinarne i valori, è necessario ricorrere a grafici, a tabelle o a programmi al calcolatore. Un grafico della funzione $\mathbb{G}(x)$, in scala naturale, è riportato in Fig. 3.23; notiamo tuttavia che tale grafico non consente la determinazione accurata dei valori della funzione. Si veda l'Appendice C per un grafico più accurato (Fig. C.1), per una tabella dei valori (Tab. C.1) e per programmi Matlab utili per il calcolo; nella stessa Appendice sono riportate le principali proprietà della funzione $\mathbb{G}(x)$ e le relazioni con altre funzioni frequentemente utilizzate.

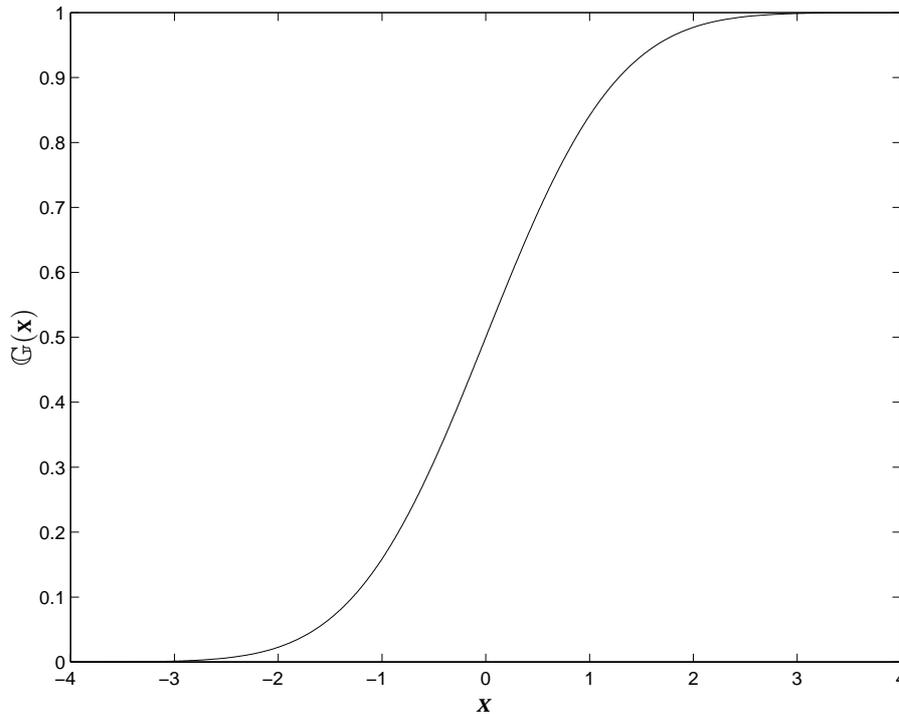


Fig. 3.23. Grafico in scala naturale della funzione $\mathbb{G}(x)$.

Una forma alternativa per la CDF di una variabile aleatoria gaussiana si può ottenere definendo la funzione $Q(x)$ (più nota, con terminologia inglese, come “Q-function”)

$$Q(x) \triangleq 1 - \mathbb{G}(x) = \frac{1}{\sqrt{2\pi}} \int_x^{\infty} e^{-\frac{y^2}{2}} dy \quad (3.9)$$

che rappresenta la CDF complementare di una variabile aleatoria gaussiana con $\mu = 0$ e $\sigma = 1$, e pertanto si ha:

$$F(x) = \mathbb{G}\left(\frac{x-\mu}{\sigma}\right) = 1 - Q\left(\frac{x-\mu}{\sigma}\right).$$

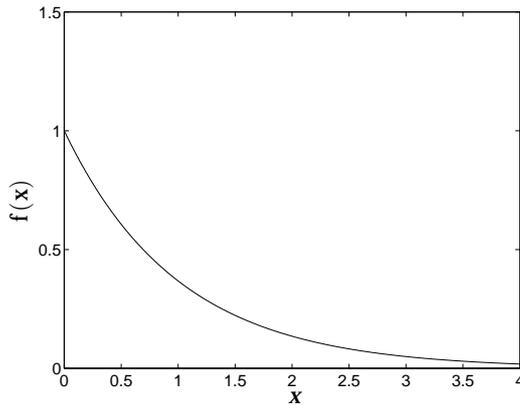


Fig. 3.24. La pdf $f(x)$ di una variabile aleatoria esponenziale ($\lambda = 1$).

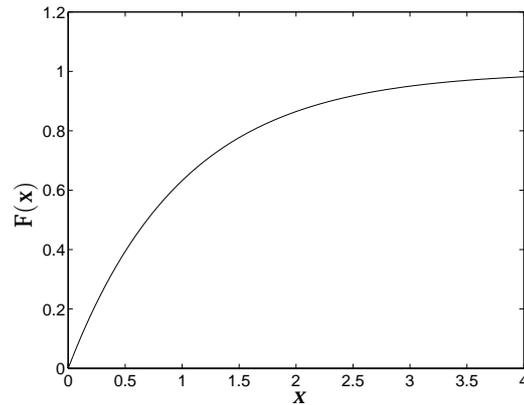


Fig. 3.25. La CDF $F(x)$ di una variabile aleatoria esponenziale ($\lambda = 1$).

Per calcolare la funzione $Q(x)$, è possibile utilizzare grafici, tabelle e programmi per il calcolo della $\mathbb{G}(x)$, tenendo conto della relazione (3.9). Inoltre, per ogni $x > 0$, vale la coppia di disuguaglianze

$$\frac{1}{x\sqrt{2\pi}} e^{-x^2/2} \left(1 - \frac{1}{x^2}\right) < Q(x) < \frac{1}{x\sqrt{2\pi}} e^{-x^2/2}.$$

Poichè il rapporto fra i due limiti vale $1 - 1/x^2$, al crescere di x essi diventano sempre più vicini e quindi entrambi approssimano la $Q(x)$ con notevole accuratezza.

La variabile aleatoria gaussiana gioca un ruolo preminente nella teoria della probabilità, principalmente in virtù del fatto che essa rappresenta una *distribuzione limite*: più precisamente, la pdf gaussiana rappresenta la distribuzione della *somma* di un numero elevato (al limite, infinito) di variabili aleatorie indipendenti e aventi pdf arbitrarie, a patto che il contributo di ciascuna variabile aleatoria alla somma sia trascurabile, una situazione che si verifica spesso in pratica (si pensi alla corrente elettrica che si può guardare come la somma dei contributi elementari di corrente dei singoli elettroni). Questa proprietà è formulata in maniera matematicamente corretta nel *teorema fondamentale del limite*, che vedremo nel capitolo 8 (cfr. § 8.6.2).

3.5.8 Variabile aleatoria esponenziale

La variabile aleatoria X si dice esponenziale (monolatera), e si denota $X \sim \text{Exp}(\lambda)$, se la sua pdf (Fig. 3.24) è:

$$f(x) = \lambda e^{-\lambda x} u(x)$$

con $\lambda > 0$. La CDF (Fig. 3.25) si calcola per integrazione:

$$F_X(x) = (1 - e^{-\lambda x}) u(x),$$

dove $u(x)$ rappresenta la funzione gradino unitario, definita come:

$$u(x) = \begin{cases} 1, & x \geq 0; \\ 0, & x < 0. \end{cases}$$

Osserviamo che la variabile aleatoria esponenziale monolatera è una variabile aleatoria *positiva*.

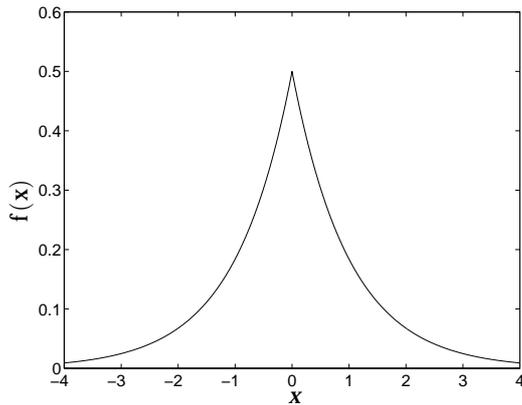


Fig. 3.26. La pdf $f(x)$ di una variabile aleatoria di Laplace ($\lambda = 1$).

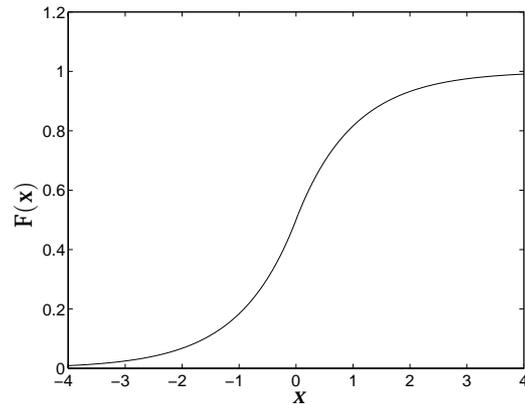


Fig. 3.27. La CDF $F(x)$ di una variabile aleatoria di Laplace ($\lambda = 1$).

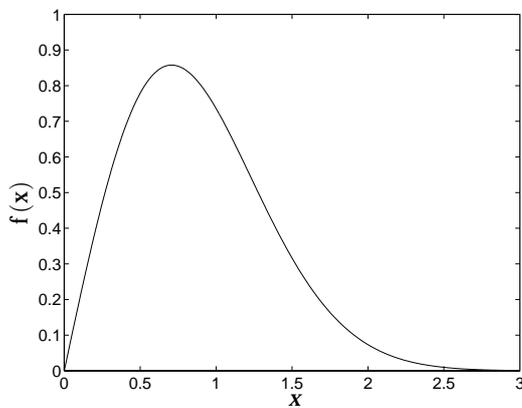


Fig. 3.28. La pdf $f(x)$ di una variabile aleatoria di Rayleigh ($b = 1$).

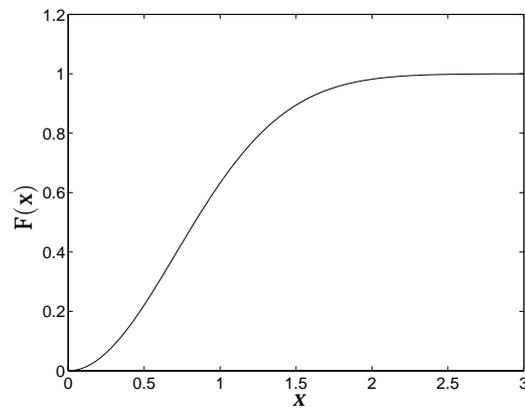


Fig. 3.29. La CDF $F(x)$ di una variabile aleatoria di Rayleigh ($b = 1$).

3.5.9 Variabile aleatoria di Laplace (esponenziale bilatera)

La variabile aleatoria X si dice di Laplace (o esponenziale bilatera), e si denota $X \sim \text{Lap}(\lambda)$, se la sua pdf (Fig. 3.26) è:

$$f(x) = \frac{\lambda}{2} e^{-\lambda|x|},$$

con $\lambda > 0$. La CDF (Fig. 3.27) si calcola per integrazione:

$$F(x) = \begin{cases} \frac{1}{2} e^{\lambda x}, & x < 0; \\ 1 - \frac{1}{2} e^{-\lambda x}, & x \geq 0. \end{cases}$$

3.5.10 Variabile aleatoria di Rayleigh

La variabile aleatoria X si dice di Rayleigh, e si denota $X \sim \text{Rayleigh}(b)$, se la sua pdf (Fig. 3.28) è:

$$f(x) = \frac{2x}{b} e^{-\frac{x^2}{b}} u(x),$$

con $b > 0$. La CDF (Fig. 3.28) si calcola per integrazione:

$$F(x) = (1 - e^{-\frac{x^2}{b}}) u(x).$$

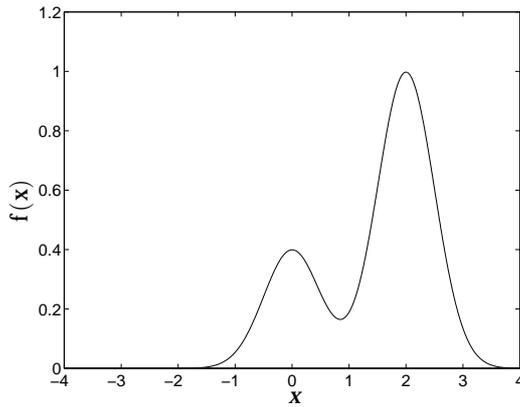


Fig. 3.30. La pdf $f(x)$ di una variabile aleatoria mixture di due pdf gaussiane, con $\mu_1 = 0$ $\mu_2 = 2$, $\sigma_1 = 0.5$, $\sigma_2 = 0.2$, $\gamma_1 = \gamma_2 = 0.5$.

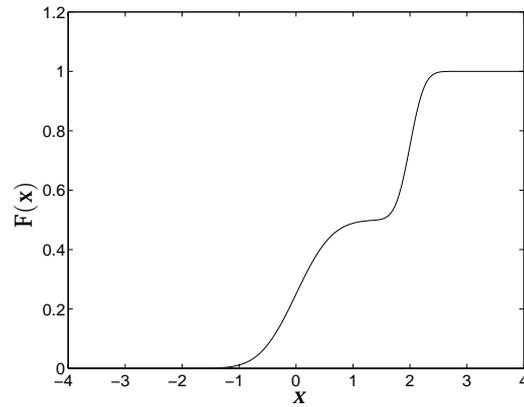


Fig. 3.31. La CDF $F(x)$ di una variabile aleatoria mixture di due CDF gaussiane, con $\mu_1 = 0$ $\mu_2 = 2$, $\sigma_1 = 0.5$, $\sigma_2 = 0.2$, $\gamma_1 = \gamma_2 = 0.5$.

Osserviamo che si tratta di una variabile aleatoria *positiva*.

3.5.11 Variabile aleatoria di tipo “mixture”

Consideriamo un esempio di variabile aleatoria non elementare, ottenuta cioè a partire da altre variabili aleatorie. Siano X_1 ed X_2 due variabili aleatorie arbitrarie, aventi rispettivamente pdf $f_1(x)$ ed $f_2(x)$. Definiamo una nuova pdf $f(x)$ come combinazione lineare delle due:

$$f(x) = \gamma f_1(x) + (1 - \gamma) f_2(x) \tag{3.10}$$

con $\gamma \in [0, 1]$. Osserviamo che effettivamente la (3.10) definisce una valida pdf, in quanto:

- $f(x) \geq 0$;
- $\int_{-\infty}^{\infty} f(x) dx = 1$.

La variabile aleatoria X avente tale pdf viene chiamata *mixture* (mistura) delle variabili aleatorie X_1 ed X_2 . Ovviamente, la CDF $F(x)$ sarà la combinazione lineare delle CDF $F_1(x)$ ed $F_2(x)$, con gli stessi coefficienti γ_1 e γ_2 . In Figg. 3.30 e 3.31 sono riportate la pdf e la CDF della variabile aleatoria ottenuta come mixture di due pdf (o CDF) gaussiane. Si noti in particolare dalla Fig. 3.30 la natura multimodale (in particolare, *bimodale*) della pdf risultante, una proprietà tipica delle variabile aleatoria mixture.

La definizione precedente può essere facilmente estesa al caso più generale di una pdf $f(x)$ ottenuta come *mixture* di $n > 2$ pdf:

$$f(x) = \sum_{i=1}^n \gamma_i f_i(x),$$

dove $\gamma_i > 0$ e $\sum_{i=1}^n \gamma_i = 1$.

► **Esempio 3.19.** Una variabile aleatoria $X \sim \text{Lap}(\lambda)$ di tipo Laplace si può vedere come mixture delle seguenti pdf (per $\gamma = 0.5$):

$$\begin{aligned} f_1(x) &= \lambda e^{-\lambda x} u(x) && \text{(esponenziale);} \\ f_2(x) &= \lambda e^{\lambda x} u(-x) && \text{(esponenziale negativa).} \end{aligned}$$

Infatti, si ha:

$$f(x) = 0.5 \lambda e^{-\lambda x} u(x) + 0.5 \lambda e^{\lambda x} u(-x) = \frac{\lambda}{2} e^{-\lambda|x|},$$

poichè per $x > 0$ risulta $u(x) = 1$ e $u(-x) = 0$, e viceversa per $x < 0$. ◀

3.5.12 Relazioni tra variabile aleatoria binomiale e gaussiana: i teoremi di de Moivre-Laplace★

Con riferimento al problema delle prove ripetute ed alla variabile aleatoria binomiale, gli esempi 3.15 e 3.16 hanno mostrato che un problema che si pone spesso in pratica è quello di valutare espressioni del tipo

$$\sum_{k=k_1}^{k_2} p(k), \quad (3.11)$$

dove $p(x)$ è la DF di una variabile aleatoria $X \sim B(n, p)$. Tale valutazione è computazionalmente difficoltosa quando il numero di termini della somma è elevato. Per valori elevati di n , tuttavia, è possibile trovare approssimazioni che semplificano il calcolo.

La prima approssimazione, nota come *teorema locale di de Moivre-Laplace*,¹⁰ afferma che se $npq \gg 1$, allora:

$$p(k) = \binom{n}{k} p^k q^{n-k} \approx \frac{1}{\sqrt{2\pi npq}} e^{-\frac{(k-np)^2}{2npq}}, \quad (3.12)$$

e l'approssimazione è accurata per k appartenente ad un intorno, di ampiezza \sqrt{npq} , di np . Dal confronto con la (3.7), notiamo che questa approssimazione consiste nell'approssimare i valori della DF della variabile aleatoria binomiale $X \sim B(n, p)$ (discreta) con quelli della pdf della variabile aleatoria gaussiana $X \sim N(np, \sqrt{npq})$ (continua).

Per mostrare la bontà dell'approssimazione fornita da tale teorema, in Fig. 3.32 riportiamo, al variare di k , la stessa $p(k)$ della Fig. 3.15 ($n = 20$, $p = 0.4$), insieme con la pdf gaussiana approssimante [secondo membro della (3.12)]. Nel caso in esame, si ha $np = 8$ e $\sqrt{npq} \approx 2.19$, per cui ci aspettiamo una approssimazione accurata all'incirca nell'intervallo $[6, 10]$; notiamo che invece si ha un buon accordo anche al di fuori di tale intervallo. Osserviamo che il parametro σ della pdf gaussiana approssimante è proprio pari a \sqrt{npq} ; per questo motivo, l'approssimazione del teorema di de Moivre-Laplace è buona nel centro della campana, e peggiora spostandosi verso le "code" della pdf gaussiana.

Una volta introdotta l'approssimazione del teorema locale di de Moivre-Laplace, possiamo trovare una approssimazione della (3.11). Si ha infatti, utilizzando la (3.12),

$$\sum_{k=k_1}^{k_2} p(k) \approx \sum_{k=k_1}^{k_2} \frac{1}{\sqrt{2\pi npq}} e^{-\frac{(k-np)^2}{2npq}}.$$

Poichè \sqrt{npq} rappresenta una misura della larghezza della pdf gaussiana, per $\sqrt{npq} \gg 1$ possiamo ritenere che tale pdf sia praticamente costante in ogni intervallo di ampiezza unitario. Allora la sommatoria tra k_1 e k_2 è una buona approssimazione dell'integrale, e si ha:

$$\sum_{k=k_1}^{k_2} p(k) \approx \frac{1}{\sqrt{2\pi npq}} \int_{k_1}^{k_2} e^{-\frac{(x-np)^2}{2npq}} dx.$$

¹⁰Il teorema fu dimostrato da A. de Moivre (1667-1754) nel caso particolare $p = 1/2$, e generalizzato da P. S. Laplace (1749-1827) al caso di p arbitrario. Per una dimostrazione moderna, si veda [1] oppure [2]: tale dimostrazione si basa sullo sviluppo asintotico (per valori elevati di n) del coefficiente binomiale utilizzando la formula di Stirling per il fattoriale (vedi Appendice A).

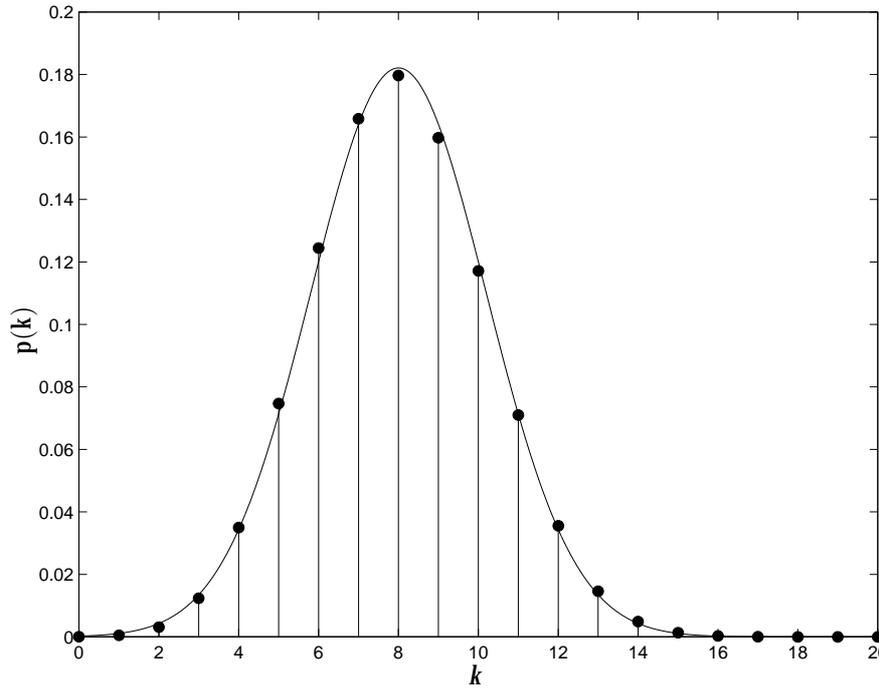


Fig. 3.32. Approssimazione del teorema locale di de Moivre-Laplace: pdf gaussiana (a tratto continuo) e DF binomiale $p(k)$, per $n = 20$ e $p = 0.4$.

Con il cambiamento di variabile $u = \frac{x-np}{\sqrt{npq}}$, l'integrale si riscrive:

$$\frac{1}{\sqrt{2\pi}} \int_{\frac{k_1-np}{\sqrt{npq}}}^{\frac{k_2-np}{\sqrt{npq}}} e^{-\frac{u^2}{2}} du = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{k_2-np}{\sqrt{npq}}} e^{-\frac{u^2}{2}} du - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{k_1-np}{\sqrt{npq}}} e^{-\frac{u^2}{2}} du$$

per cui, se introduciamo la funzione

$$\mathbb{G}(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{u^2}{2}} du,$$

possiamo porre infine:

$$\sum_{k=k_1}^{k_2} p(k) \approx \mathbb{G}\left(\frac{k_2-np}{\sqrt{npq}}\right) - \mathbb{G}\left(\frac{k_1-np}{\sqrt{npq}}\right).$$

Questa è l'espressione desiderata, che va sotto il nome di *teorema integrale di de Moivre-Laplace* e ci consente di calcolare la (3.11) come differenza della funzione $\mathbb{G}(x)$ in due punti.¹¹

► **Esempio 3.20.** Per mostrare la validità dell'approssimazione del teorema di de Moivre-Laplace, ricalcoliamo i risultati dell'esempio 3.15. Notiamo che si ha $np = 100$ e $\sqrt{npq} \approx 9.49$. Tralasciando il primo risultato (nessun componente difettoso) per il quale il calcolo diretto non presenta difficoltà, per il secondo (numero di componenti difettosi minore o uguale a 80) si ha (cfr. Tab. C.1 per i valori della $\mathbb{G}(x)$):

$$\begin{aligned} \sum_{k=0}^{80} p(k) &\approx \mathbb{G}\left(\frac{80-100}{9.49}\right) - \mathbb{G}\left(\frac{0-100}{9.49}\right) \\ &= \mathbb{G}(-2.11) - \mathbb{G}(-10.54) \approx \mathbb{G}(-2.11) = 1 - \mathbb{G}(2.11) = 1 - 0.9826 = 0.0174 \end{aligned}$$

¹¹Osserviamo che entrambi i teoremi di de Moivre-Laplace rappresentano un caso particolare di un teorema più generale, noto come *teorema limite fondamentale*, che vedremo in dettaglio nel § 8.6.

che va confrontato con il risultato esatto pari a 0.0176. Per il terzo risultato (numero di componenti difettosi compreso tra 80 e 120), si ha (cfr. Tab. C.1):

$$\begin{aligned} \sum_{k=80}^{120} p(k) &\approx \mathbb{G}\left(\frac{120-100}{9.49}\right) - \mathbb{G}\left(\frac{80-100}{9.49}\right) \\ &= \mathbb{G}(2.11) - \mathbb{G}(-2.11) = \mathbb{G}(2.11) - 1 + \mathbb{G}(2.11) = \\ &= 2 \mathbb{G}(2.11) - 1 = 2 \cdot 0.9826 - 1 = 0.9652 \end{aligned}$$

che va confrontato con il risultato esatto 0.9695. ◀

► **Esempio 3.21.** Un'altra applicazione interessante del teorema di de Moivre-Laplace ci consente di mettere in relazione tra loro il concetto di probabilità con quello di frequenza di successo. Sia $X \sim B(n, p)$ il numero di volte che si verifica un evento A , di probabilità p , in n prove; ci aspettiamo che, se il numero di prove è sufficientemente elevato, la frequenza di successo, definita come $\hat{p} \triangleq \frac{X}{n}$, debba essere prossima alla probabilità p . Questo enunciato vago può essere espresso in termini più precisi nel modo seguente: per n elevato, la probabilità che la variabile aleatoria \hat{p} si discosti da p di un ammontare pari ad ε deve essere piccola. Proviamo a calcolare tale probabilità, che si può esprimere come:

$$P(|\hat{p} - p| > \varepsilon).$$

Risulta più conveniente calcolare la probabilità complementare, ovvero:

$$\begin{aligned} P(|\hat{p} - p| \leq \varepsilon) &= P(p - \varepsilon \leq \hat{p} \leq p + \varepsilon) = P\left(p - \varepsilon \leq \frac{X}{n} \leq p + \varepsilon\right) \\ &= P[n(p - \varepsilon) \leq X \leq n(p + \varepsilon)] \end{aligned}$$

che ci aspettiamo grande per n sufficientemente elevato. Poichè siamo giunti ad una probabilità binomiale, possiamo scrivere, adoperando il teorema integrale di de Moivre-Laplace:

$$\begin{aligned} P(|\hat{p} - p| \leq \varepsilon) &= \sum_{k=n(p-\varepsilon)}^{n(p+\varepsilon)} p(k) \approx \mathbb{G}\left(\frac{n(p+\varepsilon) - np}{\sqrt{npq}}\right) - \mathbb{G}\left(\frac{n(p-\varepsilon) - np}{\sqrt{npq}}\right) \\ &= \mathbb{G}\left(\frac{n\varepsilon}{\sqrt{npq}}\right) - \mathbb{G}\left(\frac{-n\varepsilon}{\sqrt{npq}}\right) = 2 \mathbb{G}\left(\varepsilon \sqrt{\frac{n}{pq}}\right) - 1. \end{aligned}$$

Se, ad esempio, $p = 0.1$ e $\varepsilon = 0.1 p$, cioè lo scostamento è pari al 10% del valore di p , allora si trova che:

- per $n = 100$, la probabilità è pari a 0.2611;
- per $n = 1000$, la probabilità è pari a 0.7063;
- per $n = 10\,000$, la probabilità è pari a 0.9991.

In effetti, poiché $\mathbb{G}(\infty) = 1$, si ha che $\lim_{n \rightarrow \infty} P(|\hat{p} - p| \leq \varepsilon) = 1$. Si osserva allora che, se aumentiamo il numero di prove, effettivamente la frequenza di successo assume con probabilità tendente ad 1 valori prossimi a piacere alla probabilità p .¹² ◀

¹²Questa regolarità della frequenza di successo è nota come *legge dei grandi numeri*, e sarà approfondita più in dettaglio nel § 8.6.

3.6 Esercizi proposti

Esercizio 3.1. Si consideri lo spazio di probabilità (Ω, \mathcal{S}, P) associato al lancio di un dado ben bilanciato, e sia X la variabile aleatoria definita su Ω come segue:

$$X(\omega_1) = 2, \quad X(\omega_2) = 10, \quad X(\omega_3) = 2, \quad X(\omega_4) = 4, \quad X(\omega_5) = 0, \quad X(\omega_6) = -2.$$

Calcolare la CDF, la DF e la pdf della variabile aleatoria X e rappresentarle graficamente.

Esercizio 3.2. Si lanciano due dadi bilanciati, e si definisce la variabile aleatoria X come la somma dei punteggi ottenuti nei due lanci. Determinare la DF della variabile aleatoria X e rappresentarla graficamente.

★ *Esercizio 3.3.* Si lanciano tre dadi bilanciati, e si definisce la variabile aleatoria X come la somma dei punteggi ottenuti nei tre lanci. Determinare la DF della variabile aleatoria X e rappresentarla graficamente. [Suggerimento: risolvere per enumerazione ed utilizzando calcolo combinatorio elementare]

Esercizio 3.4. Si lancia un dado bilanciato finché non esca la stessa faccia due volte consecutive, e sia X la variabile aleatoria che rappresenta il numero di lanci. Calcolare la DF di X .

★ *Esercizio 3.5.* Stabilire per quale valore di c ciascuna delle seguenti funzioni $p(k)$, definite sui valori interi positivi $k = 1, 2, \dots$, è una valida DF:

a) $p(k) = c \frac{2^k}{k!}$

b) $p(k) = c p^k, p \in [0, 1];$

c) $p(k) = c \frac{p^k}{k}, p \in [0, 1];$

d) $p(k) = c \frac{1}{k(k+1)}$

[Risposta: $c = 1/(e^2 - 1)$; $c = (1 - p)/p$; $c = 1/\ln(1/(1 - p))$; $c = 1$]

Esercizio 3.6. Si consideri il seguente esperimento di probabilità: l'intensità di corrente che scorre attraverso un resistore R è una grandezza aleatoria $i \in \Omega = [-I_0, I_0]$. Assumendo per i una distribuzione uniforme di probabilità su Ω , si considerino le seguenti variabili aleatorie definite su (Ω, \mathcal{S}, P) :

a) la corrente $X(i) = i$;

b) la tensione $X(i) = Ri$ ai capi del resistore;

c) la potenza $X(i) = Ri^2$ dissipata dal resistore per effetto Joule.

Calcolare le CDF e le pdf delle variabili aleatorie X precedentemente definite e rappresentarle graficamente.

★ *Esercizio 3.7.* Un utente si reca ad uno sportello in un istante t qualunque dell'intervallo $\Omega = (0, T)$, senza sapere che lo sportello è occupato fino all'istante $T_0 < T$. Costruire una variabile aleatoria positiva X su Ω che descrive il tempo di attesa dell'utente e calcolarne CDF e pdf, rappresentandole graficamente; stabilire inoltre se X è una variabile aleatoria continua, discreta oppure mista.

Esercizio 3.8. Una coppia decide di continuare ad avere figli finché non nasce una bambina. Calcolare la DF della variabile aleatoria discreta X che rappresenta il numero di figli della coppia.

Esercizio 3.9. Il numero di persone in una fila è modellato come una variabile aleatoria $X \sim \text{Geom}(0.5)$.

a) Calcolare la probabilità che ci sia un numero *dispari* di persone in fila.

b) Calcolare la probabilità che ci sia un numero *pari* di persone in fila.

Esercizio 3.10. Sia $X = \frac{1}{2}N^2$, dove N è un numero intero aleatorio a valori equiprobabili in $-1 \leq N \leq 3$. Calcolare e diagrammare la CDF di X , ed utilizzarla per calcolare le probabilità dei seguenti eventi: $\{X \leq 0\}$, $\{2 < X \leq 3\}$, $\{X < 2\}$ e $\{X \geq 2\}$. [Risposta: $\frac{1}{5}, 0, \frac{3}{5}, \frac{2}{5}$]

★ *Esercizio 3.11.* In un cesto ci sono 12 mele sane e 4 mele marce, e voi estraete 3 mele a caso, simultaneamente.

- Descrivere l'esperimento in termini probabilistici, individuando lo spazio campione Ω e la legge di probabilità;
- determinare la DF della variabile aleatoria discreta X , definita su Ω , che rappresenta il numero di mele sane che estraete dal cesto. Qual è il valore di X più probabile?

[Suggerimento: utilizzare il calcolo combinatoriale]

[Risposta: $p_X(0) = \frac{1}{140}$, $p_X(1) = \frac{18}{140}$, $p_X(2) = \frac{66}{140}$, $p_X(3) = \frac{55}{140}$]

Esercizio 3.12. Determinare la pdf $f(x)$ associata alla CDF $F(x) = (1 - e^{-\alpha x}) u(x - c)$, con $\alpha > 0$ e $c \geq 0$. Stabilire se si tratta di una variabile aleatoria discreta, continua o mista.

Esercizio 3.13. Si consideri la funzione $f(x) = c x e^{-x} u(x)$.

- determinare c affinché $f(x)$ sia la valida pdf di una variabile aleatoria X ;
- utilizzando il valore di c determinato al passo precedente, calcolare la CDF $F(x)$, ed utilizzarla per valutare $P(X \leq 1)$, $P(1 < X \leq 2)$, e $P(X > 2)$.

[Risposta: $c = 1$; $1 - 2e^{-1}$, $2e^{-1} - 3e^{-2}$, $3e^{-2}$]

Esercizio 3.14. Una variabile aleatoria X ha la seguente CDF:

$$F(x) = \begin{cases} 0, & x \leq 0; \\ kx^2, & 0 < x \leq 10; \\ 100k, & x > 10. \end{cases}$$

Determinare k , valutare $P(X \leq 5)$ e $P(5 < X \leq 7)$, calcolare e diagrammare la pdf corrispondente. Si tratta di una variabile aleatoria discreta, continua o mista? [Risposta: $k = \frac{1}{100}$, $\frac{1}{4}$, $\frac{6}{25}$]

Esercizio 3.15. La pdf *triangolare* vale 0 ovunque, ad eccezione dell'intervallo limitato (a, b) , nel quale essa assume la forma di un triangolo isoscele.

- Determinare l'espressione di $f(x)$ e diagrammarla;
- determinare l'espressione di $F(x)$ e diagrammarla.

Esercizio 3.16. Una moneta viene lanciata 10 volte ed i lanci sono tutti indipendenti.

- Calcolare $P(10 \text{ teste})$.
- Calcolare $P(5 \text{ teste e } 5 \text{ croci in ordine qualsiasi})$.
- Dire se $P(\text{testa} \mid 10 \text{ teste})$ è minore, uguale o maggiore di 0.5.
- Stabilire se è più facile avere N teste e N croci su $2N$ lanci o $N + 1$ teste e $N + 1$ croci su $2N + 2$ lanci.

Esercizio 3.17. Una moneta viene lanciata 4 volte ed i lanci sono tutti indipendenti. Calcolare la probabilità di ottenere:

- almeno tre teste;
- esattamente tre teste;
- una sequenza di tre o più teste consecutive;
- una sequenza di esattamente tre teste consecutive.

[Risposta: $\frac{5}{16}$, $\frac{1}{4}$, $\frac{3}{16}$, $\frac{1}{8}$]

Esercizio 3.18. In un gioco a premi, un giocatore ha a disposizione 10 lanci per colpire un bersaglio, e vince se il bersaglio viene colpito almeno due volte. Supponendo che la probabilità di colpire il bersaglio in un singolo lancio sia $1/5$, e che i lanci siano indipendenti:

- calcolare la probabilità che il giocatore vinca il premio;

b) calcolare la probabilità che il giocatore vinca il premio, sapendo che ha colpito almeno una volta il bersaglio.

Esercizio 3.19. Si trasmettono messaggi di tre bit su un BSC con probabilità di scambio $\varepsilon = 1/5$, e sia X la variabile aleatoria discreta che descrive il numero di errori presenti in una terna di bit. Determinare la DF della variabile aleatoria X .

Esercizio 3.20. Calcolare la mediana ed il percentile u -esimo di una variabile aleatoria $X \sim \text{Exp}(\lambda)$.

Esercizio 3.21. Calcolare la mediana ed il percentile u -esimo di una variabile aleatoria $X \sim \text{Rayleigh}(b)$.

Esercizio 3.22. Calcolare la mediana ed il percentile u -esimo di una variabile aleatoria $X \sim N(\mu, \sigma)$. In particolare, determinare il valore dei quartile inferiore $x_{0.25}$, del quartile superiore $x_{0.75}$, e dei percentili $x_{0.90}$, $x_{0.95}$, $x_{0.99}$ in funzione dei parametri μ e σ .

[Suggerimento: utilizzare la tabella dei valori della funzione $\mathbb{G}(x)$]

Esercizio 3.23. Si misurano i valori di resistenza di componenti prodotti da una linea di produzione, e si accettano solo quei componenti la cui resistenza X è compresa tra 96 e 104 ohm. Determinare la percentuale dei componenti accettati, nei casi in cui:

- a) X è una variabile aleatoria uniforme tra 95 e 105 ohm;
- b) X è una variabile aleatoria gaussiana con $\mu = 100$ ohm e $\sigma = 2$ ohm.

[Risposta: 0.8, 0.9546]

Esercizio 3.24. In un processo per paternità contestata, un esperto testimonia che la lunghezza (espressa in giorni) di una gravidanza, dal concepimento alla nascita, è approssimativamente una variabile aleatoria $X \sim N(\mu, \sigma)$, con $\mu = 270$ e $\sigma = 10$. La difesa può provare che il suo cliente, imputato nel processo, si trovava all'estero nel periodo da 290 a 240 giorni prima della nascita del bambino. Qual è la probabilità che l'imputato si trovasse in Italia quando il bambino fu concepito? [Risposta: $2.41 \cdot 10^{-2}$]

Esercizio 3.25. L'esame finale del corso di Teoria dei Fenomeni Aleatori è congegnato in modo che il punteggio sia distribuito approssimativamente come una variabile aleatoria gaussiana $X \sim N(\mu, \sigma)$. Al punteggio X si associano cinque fasce di merito, da A (la migliore) fino a E (la peggiore), secondo la tabella seguente. Calcolare la frazione degli studenti che viene valutato A, B, C, D, E . [Risposta: 16%, 34%, 34%, 14%, 2%]

Intervallo di voti	Fascia
$X > \mu + \sigma$	A
$\mu < X \leq \mu + \sigma$	B
$\mu - \sigma < X \leq \mu$	C
$\mu - 2\sigma < X \leq \mu - \sigma$	D
$X \leq \mu - 2\sigma$	E

Trasformazioni di una variabile aleatoria

In questo capitolo si introduce e discute un argomento di interesse prevalentemente applicativo. Dopo aver fornito una definizione formale di trasformazione $Y = g(X)$, si introducono numerosi esempi che mostrano come caratterizzare statisticamente la variabile aleatoria Y , una volta nota la caratterizzazione statistica di X (problema diretto); in particolare, si espone il teorema fondamentale sulle trasformazioni di variabili aleatorie, che rappresenta uno strumento semplice e sufficientemente generale per la risoluzione del problema. Si affronta anche il cosiddetto problema inverso, consistente nel determinare la trasformazione g che consente di trasformare una variabile aleatoria X in una variabile aleatoria Y , entrambe con caratterizzazione statistica assegnata. Tale problema ricorre nella cosiddetta sintesi di variabili aleatorie, ovvero nella generazione automatica di variabili aleatorie mediante calcolatore: a tale proposito, si discute brevemente l'algoritmo lineare congruente, tra i più utilizzati nelle tecniche di simulazione.

4.1 Introduzione

Si presentano spesso casi in cui, a partire da una variabile aleatoria X , si ottiene una nuova variabile aleatoria Y mediante una opportuna *trasformazione* $Y = g(X)$. Ad esempio, supponiamo che la variabile aleatoria X rappresenti l'intensità di corrente che passa attraverso una resistenza (ideale) di 1 ohm; la potenza dissipata dalla resistenza si può scrivere come $Y = X^2$, e poichè X è una variabile aleatoria, anche Y sarà una variabile aleatoria. Un altro esempio è quello in cui la variabile aleatoria X rappresenta un angolo scelto a caso in $(0, 2\pi)$; il coseno $Y = \cos(X)$ di tale angolo è una variabile aleatoria ottenuta a partire da X . Formalizzando, possiamo dare la seguente definizione di trasformazione di una variabile aleatoria:

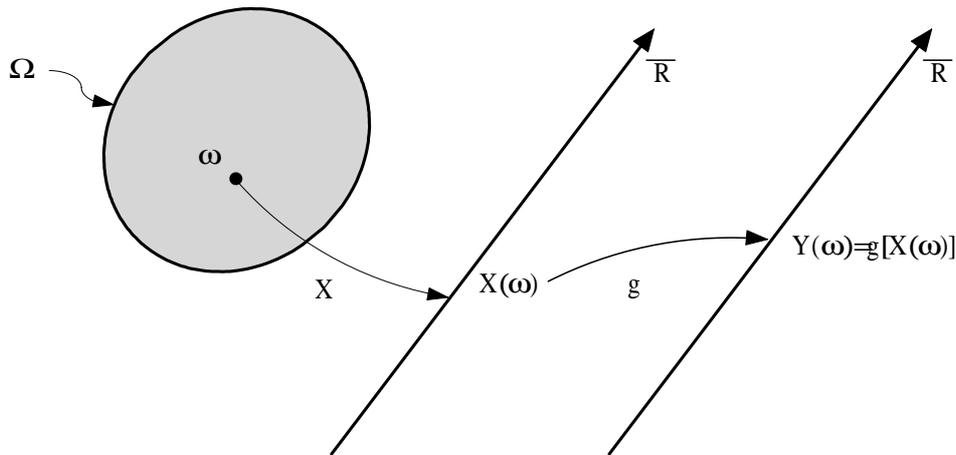


Fig. 4.1. La trasformazione $Y = g(X)$ definisce una nuova variabile aleatoria Y sullo spazio campione Ω .

Definizione (trasformazione di una variabile aleatoria). Sia X una variabile aleatoria definita sullo spazio di probabilità (Ω, \mathcal{S}, P) , e $g(x)$ una funzione definita in $\overline{\mathbb{R}}$ e a valori in $\overline{\mathbb{R}}$, tale che l'insieme di definizione di $g(x)$ contenga il codominio \mathcal{X} della funzione $X(\omega)$. La trasformazione $Y = g(X)$ definisce una nuova variabile aleatoria ottenuta associando a $\omega \in \Omega$ il valore $Y(\omega) = g[X(\omega)] \in \overline{\mathbb{R}}$.

In sostanza la nuova variabile aleatoria Y è definita su (Ω, \mathcal{S}, P) mediante una legge (Fig. 4.1) che è la *funzione composta* di X e g . La condizione richiesta sull'insieme di definizione di $g(x)$ ed il codominio \mathcal{X} di $X(\omega)$ serve semplicemente a garantire che tale funzione composta abbia un insieme di definizione non vuoto. Tuttavia, affinché $Y = g(X)$ sia effettivamente una variabile aleatoria, è necessario che la funzione g soddisfi qualche ulteriore condizione, come analizzato più in dettaglio nel seguente paragrafo.¹

4.1.1 Condizioni da imporre alla funzione $g(x)$ ★

Per capire se sia necessario richiedere qualche ulteriore condizione alla funzione g , dobbiamo ricordare che la definizione di variabile aleatoria (vedi § 3.1.1) richiede che per Y siano soddisfatte le seguenti due proprietà:

1. $\{Y \leq y\}$ deve essere un evento, $\forall y \in \overline{\mathbb{R}}$;
2. $P(\{Y = +\infty\}) = P(\{Y = -\infty\}) = 0$.

Per quanto riguarda la prima proprietà, osserviamo che, se $\{Y \leq y\}$ è un evento, la sua probabilità coincide proprio con la CDF di Y , e si ha:

$$F_Y(y) = P(Y \leq y) = P(g(X) \leq y) = P(g[X(\omega)] \leq y).$$

Per un dato $y \in \overline{\mathbb{R}}$, i valori di x tali che $g(x) \leq y$ (le soluzioni della disequazione) formano un sottoinsieme di $\overline{\mathbb{R}}$, sia esso R_y ; si ha allora:

$$F_Y(y) = P(X \in R_y). \quad (4.1)$$

¹Osserviamo peraltro che tali condizioni sono sempre verificate dalle trasformazioni che si utilizzano in pratica, per cui il lettore non interessato ad approfondimenti ulteriori può tranquillamente saltare il paragrafo che segue.

Pertanto, affinché $\{Y \leq y\}$ sia un evento, è necessario e sufficiente che $\{X \in R_y\}$ sia un evento per ogni $y \in \overline{\mathbb{R}}$, il che accade se l'insieme R_y si ottiene per complementazione, unione e/o intersezione (al più numerabile) di semirette chiuse a destra.

Per inciso, osserviamo che la (4.1) fornisce anche la strada per calcolare la CDF di Y in funzione della caratterizzazione statistica di X , argomento che sarà ulteriormente approfondito nei successivi paragrafi. Ricordando anche la seconda proprietà, dobbiamo richiedere che la trasformazione $Y = g(X)$ sia tale da soddisfare le seguenti condizioni:

1. per ogni $y \in \overline{\mathbb{R}}$, l'insieme $R_y = \{x \in \overline{\mathbb{R}} \text{ tali che } g(x) \leq y\}$ delle soluzioni della disequazione $g(x) \leq y$ dev'essere la complementazione, unione e/o intersezione (al più numerabile) di semirette chiuse a destra, cosicchè $\{Y \leq y\}$ sia un evento; una funzione g che possiede tale proprietà prende il nome di *funzione di Baire* [1];
2. gli eventi $\{g(X) = +\infty\}$ e $\{g(X) = -\infty\}$ devono avere probabilità zero.

Osserviamo che, mentre la prima proprietà coinvolge solo la funzione g , nella seconda entra in gioco anche la variabile aleatoria X . Peraltro, notiamo che praticamente tutte le funzioni elementari soddisfano la prima proprietà (sono cioè funzioni di Baire); per quanto riguarda la seconda proprietà, essa è spesso automaticamente soddisfatta, per il semplice motivo che la funzione g assume valori in \mathbb{R} e non in $\overline{\mathbb{R}}$. Nel seguito, per tutte le trasformazioni di variabili aleatorie che considereremo, riterremo sempre verificate le precedenti proprietà.

4.2 Caratterizzazione statistica di $Y = g(X)$

Data una trasformazione $Y = g(X)$, il problema che si pone in pratica è il seguente: nota la CDF (o la pdf, o la DF) di X , calcolare la CDF (o la pdf, o la DF) di Y . In breve, si parla di "caratterizzare statisticamente" la variabile aleatoria Y , nota la caratterizzazione statistica di X .

4.2.1 Calcolo della CDF di $Y = g(X)$

Consideriamo dapprima il calcolo della CDF di $Y = g(X)$. Possiamo scrivere

$$F_Y(y) = P(Y \leq y) = P(g(X) \leq y),$$

per cui per ogni $y \in \overline{\mathbb{R}}$ dobbiamo determinare i valori di X la cui immagine attraverso la funzione g è minore o uguale di y , e determinare la probabilità dell'evento corrispondente. Tale problema non ammette evidentemente una soluzione generale, ma si riconduce alla risoluzione di una o più disequazioni numeriche. I seguenti esempi, nei quali assumeremo che X sia una variabile aleatoria continua, chiariranno meglio la procedura da seguire in alcuni casi tipici.

► *Esempio 4.1.* Consideriamo la trasformazione lineare $Y = aX + b$, che è rappresentata graficamente in Fig. 4.2 nei casi $a > 0$ (a sinistra) e $a < 0$ (a destra). Nel caso $a > 0$, si ha:

$$F_Y(y) = P(Y \leq y) = P(aX + b \leq y) = P\left(X \leq \frac{y-b}{a}\right) = F_X\left(\frac{y-b}{a}\right).$$

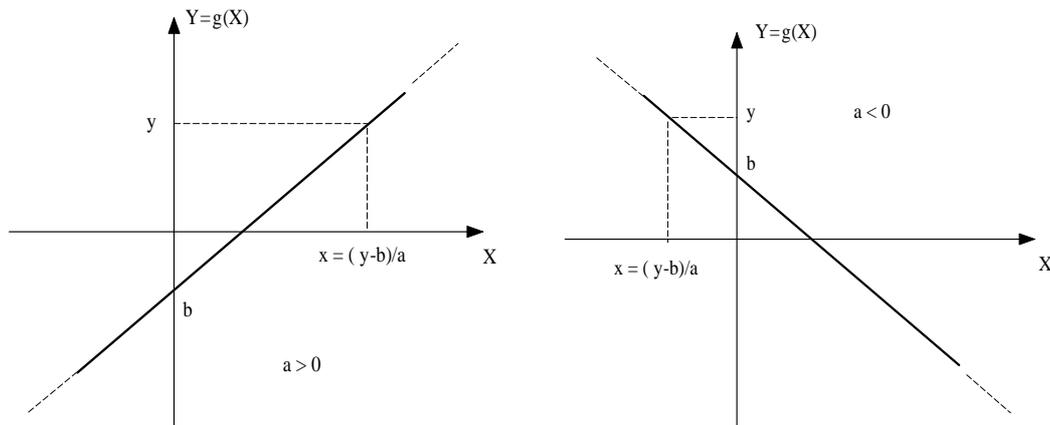


Fig. 4.2. La trasformazione $Y = aX + b$, nei casi $a > 0$ (a sinistra) e $a < 0$ (a destra).

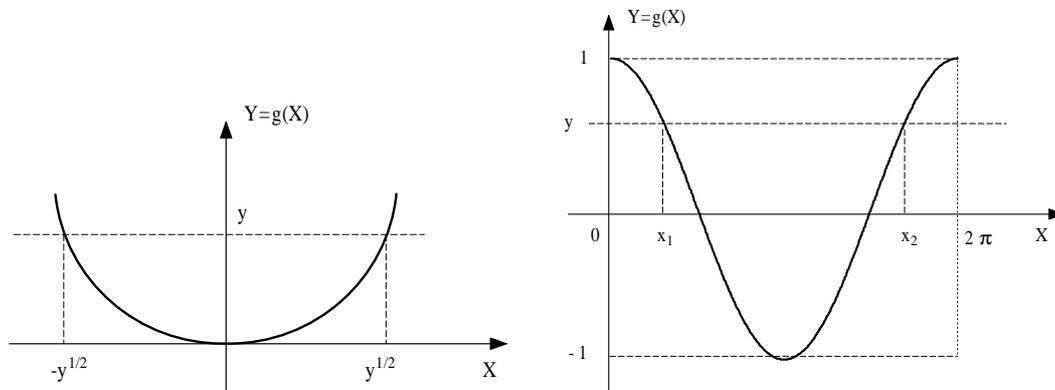


Fig. 4.3. La trasformazione $Y = X^2$.

Fig. 4.4. La trasformazione $Y = \cos(X)$.

Per $a < 0$, il verso della disuguaglianza si inverte, e si ha:²

$$\begin{aligned} F_Y(y) &= P(Y \leq y) = P(aX + b \leq y) = P\left(X \geq \frac{y-b}{a}\right) \\ &= 1 - P\left(X < \frac{y-b}{a}\right) = 1 - F_X\left(\frac{y-b}{a}\right). \end{aligned}$$

La pdf $f_X(x)$ si ottiene derivando la CDF, e per $a > 0$ si ha:

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

mentre per $a < 0$ si ha:

$$f_Y(y) = -\frac{1}{a} f_X\left(\frac{y-b}{a}\right).$$

Le due precedenti espressioni possono essere condensate nell'unica espressione, valida per $a \neq 0$:

$$f_Y(y) = \frac{1}{|a|} f_X\left(\frac{y-b}{a}\right).$$

²Si ricordi che, nell'ipotesi che X sia continua, si ha $P(X < x) = P(X \leq x)$.

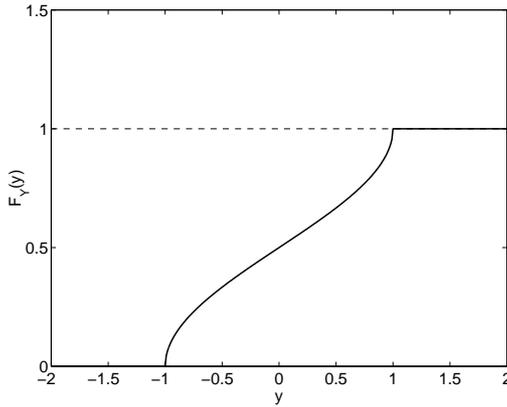


Fig. 4.5. La CDF $F(x)$ della variabile aleatoria $Y = \cos(X)$, con $X \sim U(0, 2\pi)$.

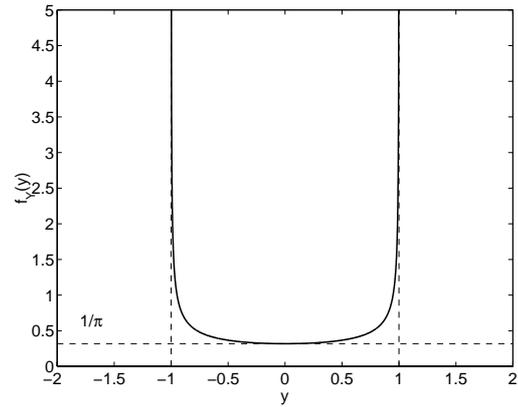


Fig. 4.6. La pdf $f(x)$ della variabile aleatoria $Y = \cos(X)$, con $X \sim U(0, 2\pi)$.

► **Esempio 4.2.** Consideriamo la trasformazione $Y = X^2$, che è rappresentata graficamente da una parabola (Fig. 4.3). Se $y < 0$, evidentemente $P(Y \leq y) = P(X^2 \leq y < 0) = P(\emptyset) = 0$. Viceversa, se $y \geq 0$, si ha che $P(Y \leq y) = P(X^2 \leq y) = P(X \in [-\sqrt{y}, \sqrt{y}]) = F_X(\sqrt{y}) - F_X(-\sqrt{y})$. In definitiva, si ha:

$$F_Y(y) = [F_X(\sqrt{y}) - F_X(-\sqrt{y})] u(y)$$

e derivando si ottiene la pdf

$$f_Y(y) = \frac{1}{2\sqrt{y}} [f_X(\sqrt{y}) + f_X(-\sqrt{y})] u(y).$$



► **Esempio 4.3.** Consideriamo la trasformazione $Y = \cos(X)$, che è rappresentata graficamente in Fig. 4.4, e supponiamo in particolare che sia $X \sim U(0, 2\pi)$. Se $y < -1$, si ha evidentemente $F_Y(y) = P(Y \leq y) = P(\cos(X) \leq y < -1) = P(\emptyset) = 0$. Viceversa, se $y \geq 1$, risulta $F_Y(y) = P(\cos(X) \leq y) = P(\Omega) = 1$. Infine, per $-1 \leq y < 1$, si ha (vedi Fig. 4.4) che la disequazione $\cos(X) \leq y$ è soddisfatta, all'interno dell'intervallo $(0, 2\pi)$, dai valori di $X \in [x_1, x_2]$, con $x_1 = \arccos(y)$ e $x_2 = 2\pi - \arccos(y)$.³ Pertanto, per tali valori di y si ha, poiché X è uniforme in $(0, 2\pi)$,

$$F_Y(y) = P(X \in [x_1, x_2]) = \frac{x_2 - x_1}{2\pi} = 1 - \frac{1}{\pi} \arccos(y).$$

In definitiva, allora, la CDF di Y si può esprimere come:

$$F_Y(y) = \begin{cases} 0, & y < -1; \\ 1 - \frac{1}{\pi} \arccos(y), & -1 \leq y < 1; \\ 1, & y \geq 1; \end{cases}$$

ed è raffigurata in Fig. 4.5 (si noti che è una funzione continua). La pdf si ottiene derivando la $F_Y(y)$:

$$f_Y(y) = \begin{cases} 0, & y < -1; \\ \frac{1}{\pi} \frac{1}{\sqrt{1-y^2}}, & -1 < y < 1; \\ 0, & y > 1; \end{cases}$$

ed è raffigurata in Fig. 4.6. Si noti che tale pdf non contiene impulsi, perché la CDF è continua; inoltre, essa non è definita (diverge) nei punti ± 1 , tuttavia ha comunque area unitaria. ◀

³Evidentemente la disuguaglianza è soddisfatta, per la periodicità della funzione coseno, anche in qualunque intervallo del tipo $[x_1 + 2k\pi, x_2 + 2k\pi]$; tuttavia, poiché per ipotesi X assume valori in $(0, 2\pi)$, non è necessario considerare tali ulteriori intervalli, ma è sufficiente limitarsi all'intervallo $(0, 2\pi)$.

► **Esempio 4.4 (amplificatore con saturazione).** Consideriamo la trasformazione in Fig. 4.7, che può essere espressa matematicamente come segue:

$$g(x) = \begin{cases} -d_y, & x < -d_x; \\ a x, & -d_x \leq x < d_x; \\ d_y, & x \geq d_x. \end{cases}$$

con $a \triangleq \frac{d_y}{d_x} > 0$. Tale legge è quella caratteristica di un dispositivo che amplifica (se $a > 1$) a patto che

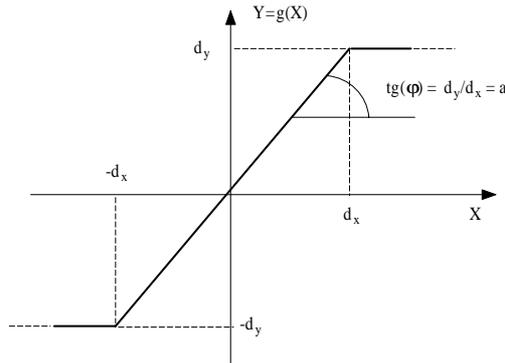


Fig. 4.7. La trasformazione $Y = g(X)$ caratteristica di un amplificatore con saturazione.

$X \in [-d_x, d_x]$, altrimenti l'uscita è limitata ("satura") al valore $-d_y$ oppure d_y .

Veniamo ora al calcolo della CDF di Y . Se $y < -d_y$, si ha evidentemente $F_Y(y) = P(Y \leq y) = P(g(X) \leq y < -d_y) = P(\emptyset) = 0$. Viceversa, se $y \geq d_y$, risulta $F_Y(y) = P(g(X) \leq y) = P(\Omega) = 1$. Per $-d_y \leq y < d_y$, infine, si ha:

$$F_Y(y) = P(Y \leq y) = P(aX \leq y) = P\left(X \leq \frac{y}{a}\right) = F_X\left(\frac{y}{a}\right)$$

In definitiva, la CDF è data da:

$$F_Y(y) = \begin{cases} 0, & y < -d_y; \\ F_X(y/a), & -d_y \leq y < d_y; \\ 1, & y \geq d_y; \end{cases}$$

ed è riportata in Fig. 4.8. Notiamo che per $y = -d_y$ tale CDF è discontinua, perché il suo limite da sinistra vale 0, mentre il limite da destra vale $F_X(-d_x)$. Allo stesso modo, la CDF è discontinua nel punto $y = d_y$, in quanto il limite da sinistra vale 1, mentre il limite da destra vale $F_X(d_x)$. Pertanto, quando calcoleremo la pdf, compariranno, oltre alla derivata convenzionale, due impulsi di Dirac, centrati in $y = -d_y$ e $y = d_y$, e

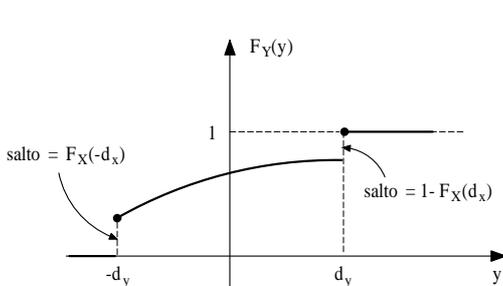


Fig. 4.8. La CDF $F_Y(y)$ della variabile aleatoria Y all'uscita di un amplificatore con saturazione.

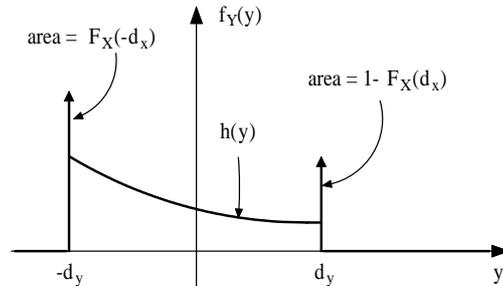


Fig. 4.9. La pdf $f_Y(y)$ della variabile aleatoria Y all'uscita di un amplificatore con saturazione.

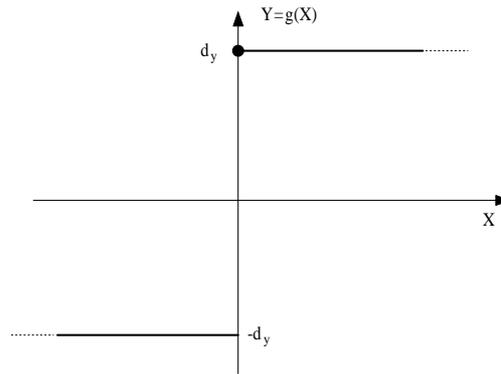


Fig. 4.10. La trasformazione $Y = g(X)$ caratteristica di un hard limiter.

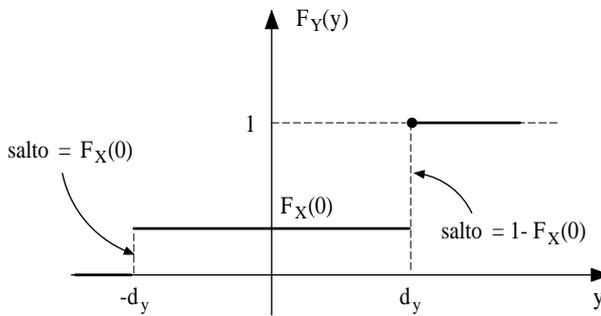


Fig. 4.11. La CDF $F_Y(y)$ della variabile aleatoria Y all'uscita di un hard limiter.

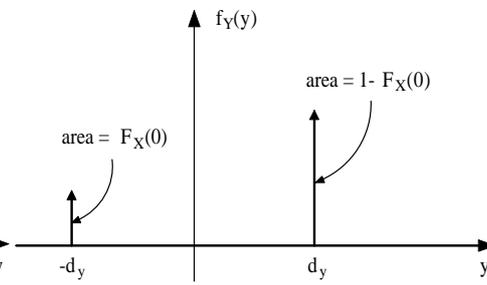


Fig. 4.12. La pdf $f_Y(y)$ della variabile aleatoria Y all'uscita di un hard limiter.

di area rispettivamente pari a $F_X(-d_x)$ ed $1 - F_X(d_x)$. La derivata convenzionale $h(y)$ vale

$$h(y) = \begin{cases} 0, & y < -d_y; \\ \frac{1}{a} f_X\left(\frac{y}{a}\right), & -d_y < y < d_y; \\ 0, & y > d_y; \end{cases}$$

mentre la pdf si ottiene aggiungendo alla derivata convenzionale gli impulsi di Dirac:

$$f_Y(y) = h(y) + F_X(-d_x) \delta(y + d_y) + [1 - F_X(d_x)] \delta(y - d_y)$$

ed è raffigurata in Fig. 4.9. In conclusione, la variabile aleatoria Y ha una CDF discontinua ma non costante a tratti, per cui costituisce un primo esempio di una variabile aleatoria *mista*. ◀

► **Esempio 4.5 (hard limiter).** Consideriamo la trasformazione in Fig. 4.10, che può essere espressa matematicamente come

$$g(x) = d_y \operatorname{sgn}(x),$$

dove $d_y > 0$, e $\operatorname{sgn}(x)$ è la funzione signum:

$$\operatorname{sgn}(x) \triangleq \begin{cases} 1, & x \geq 0; \\ -1, & x < 0. \end{cases}$$

Tale legge è quella caratteristica di un dispositivo limitatore ideale o *hard limiter*, e si può vedere come caso limite di un amplificatore con saturazione per $d_x \rightarrow 0$.

Passiamo al calcolo della CDF di Y . Se $y < -d_y$, si ha $F_Y(y) = P(Y \leq y) = P(g(X) \leq y < -d_y) = P(\emptyset) = 0$. Viceversa, se $y \geq d_y$, $F_Y(y) = P(\Omega) = 1$. Per $-d_y \leq y < d_y$, si ha $F_Y(y) = P(g(X) \leq y) = P(X \leq$

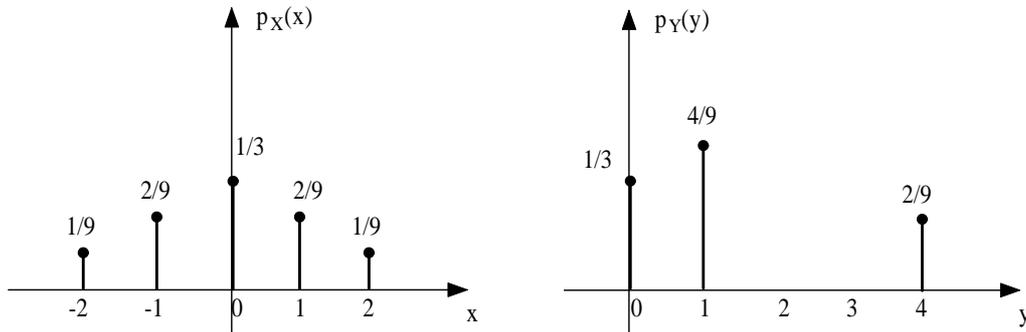


Fig. 4.13. La DF di X (a sinistra) e di $Y = X^2$ (a destra).

$0) = F_X(0)$. In definitiva, allora, la CDF di Y si può esprimere come:

$$F_Y(y) = \begin{cases} 0, & y < -d_y; \\ F_X(0), & -d_y \leq y < d_y; \\ 1, & y \geq d_y. \end{cases}$$

ed è raffigurata in Fig. 4.11; si noti che è una funzione costante a tratti, per cui Y è una variabile aleatoria discreta, e la sua pdf, puramente impulsiva, è data da

$$f_Y(y) = F_X(0) \delta(y + d_y) + [1 - F_X(0)] \delta(y - d_y)$$

ed è raffigurata in Fig. 4.12. ◀

4.2.2 Calcolo della DF di $Y = g(X)$

Il calcolo della DF è appropriato quando X è una variabile aleatoria discreta. In tal caso, è immediato osservare che, qualunque sia la trasformazione $g(x)$, anche la variabile aleatoria $Y = g(X)$ è discreta, ed assume i valori $y = g(x) \in \mathcal{Y}$, dove $x \in \mathcal{X}$ sono i valori assunti da X . In effetti, il calcolo della DF di Y è immediato: basta osservare che

$$P(Y = y) = \sum_{x \in \mathcal{X}, g(x)=y} P(X = x),$$

ovvero la probabilità che la variabile aleatoria Y assuma un qualunque valore $y \in \mathcal{Y}$ si ottiene come somma delle probabilità che la variabile aleatoria X assuma i valori x , con $y = g(x)$. Pertanto, introducendo le DF di Y ed X , possiamo scrivere sinteticamente

$$p_Y(y) = \sum_{x \in \mathcal{X}, g(x)=y} p_X(x). \quad (4.2)$$

► **Esempio 4.6.** Si consideri la seguente variabile aleatoria discreta X , che assume i valori $\mathcal{X} = \{-2, -1, 0, 1, 2\}$ con la seguente DF (Fig. 4.13):

$$p_X(x) = \begin{cases} 1/3, & x = 0; \\ 2/9, & x = \pm 1; \\ 1/9, & x = \pm 2. \end{cases}$$

Calcoliamo la DF della variabile aleatoria $Y = X^2$. La variabile aleatoria Y è ancora discreta, e assume i valori $y \in \mathcal{Y} = \{0, 1, 4\}$. Applicando la (4.2), si ha:

$$\begin{aligned} p_Y(0) &= P(Y = 0) = P(X = 0) = 1/3; \\ p_Y(1) &= P(Y = 1) = P(X = -1) + P(X = 1) = 2/9 + 2/9 = 4/9; \\ p_Y(4) &= P(Y = 4) = P(X = -2) + P(X = 2) = 1/9 + 1/9 = 2/9; \end{aligned}$$

per cui la DF si scrive in forma compatta come:

$$p_Y(y) = \begin{cases} 1/3, & y = 0; \\ 4/9, & y = 1; \\ 2/9, & y = 4. \end{cases}$$

ed è raffigurata in Fig. 4.13. Si noti che Y è una variabile aleatoria *positiva*. ◀

4.2.3 Calcolo della pdf di $Y = g(X)$

Affrontiamo adesso il problema di determinare la pdf di $Y = g(X)$ in funzione della pdf di X . Di importanza fondamentale è il seguente teorema, nel quale $g'(x)$ indica la derivata prima di $g(x)$:

Teorema 4.1 (teorema fondamentale sulle trasformazioni di variabili aleatorie). Sia X una variabile aleatoria avente pdf $f_X(x)$, e si consideri la trasformazione $Y = g(X)$; la pdf di Y è data da:

$$f_Y(y) = \begin{cases} 0, & \text{se l'equazione } y = g(x) \text{ non ammette soluzioni;} \\ \sum_i \frac{f_X(x_i)}{|g'(x_i)|}, & \text{dove } x_i \text{ è una soluzione dell'equazione } y = g(x). \end{cases}$$

Prova. La pdf $f_Y(y)$ si può ottenere sulla base della seguente relazione (per $dy > 0$):

$$f_Y(y) dy = P(y < Y \leq y + dy) = P(y < g(X) \leq y + dy).$$

Se y è un valore tale che l'equazione $g(x) = y$ non ammette soluzioni, allora $f_Y(y) = 0$. Infatti, se y non appartiene alla frontiera del codominio di $g(x)$, è possibile scegliere dy sufficientemente piccolo tale che

$$\{y < g(X) \leq y + dy\} = \emptyset \Rightarrow f_Y(y) = 0.$$

Se invece y appartiene alla frontiera del codominio di $g(x)$, posso comunque porre $f_Y(y) = 0$, perchè la frontiera è un insieme di misura nulla, e quindi il valore della pdf su un insieme di misura nulla è inessenziale. Viceversa, si consideri il caso in cui y appartenga al codominio di $g(x)$, cioè sia un valore tale che l'equazione $g(x) = y$ ammette una o più soluzioni. Per semplicità, supponiamo che le soluzioni siano tre, x_1, x_2, x_3 , come in Fig. 4.14. Allora:

$$f_Y(y) dy = P(y < Y \leq y + dy) = P(x_1 < X \leq x_1 + dx_1) + P(x_2 + dx_2 < X \leq x_2) + P(x_3 < X \leq x_3 + dx_3),$$

dove $dx_1 > 0$, $dx_2 < 0$, $dx_3 > 0$. (Fig. 4.14) e, poiché dy è infinitesimo, i tre insiemi cui appartiene X sono mutuamente esclusivi. Poichè:

$$\begin{aligned} P\{x_1 < X \leq x_1 + dx_1\} &= f_X(x_1) dx_1; \\ P\{x_2 + dx_2 < X \leq x_2\} &= f_X(x_2) |dx_2|; \\ P\{x_3 < X \leq x_3 + dx_3\} &= f_X(x_3) dx_3; \end{aligned}$$

ed inoltre

$$\begin{aligned} dx_1 &= dy/g'(x_1); \\ dx_2 &= dy/g'(x_2); \\ dx_3 &= dy/g'(x_3); \end{aligned}$$

dove (Fig. 4.14) $g'(x_1) > 0$, $g'(x_2) < 0$, e $g'(x_3) > 0$, risulta

$$f_Y(y) dy = \frac{f_X(x_1)}{g'(x_1)} dy + \frac{f_X(x_2)}{|g'(x_2)|} dy + \frac{f_X(x_3)}{g'(x_3)} dy,$$

ed eliminando dy , si ha l'asserto. ◻

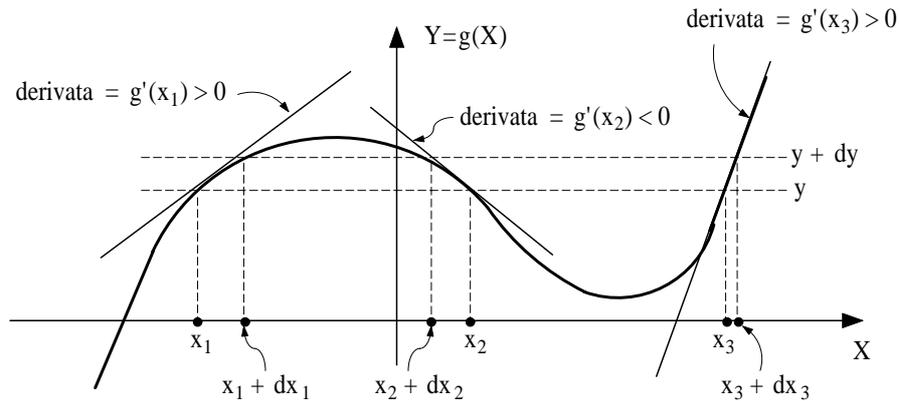


Fig. 4.14. Dimostrazione del teorema fondamentale sulle trasformazioni di variabili aleatorie. Le soluzioni dell'equazione $y = g(x)$ sono x_1 , x_2 , ed x_3 .

Nel seguito, sono riportati numerosi esempi di applicazione del teorema precedente per alcune trasformazioni di particolare interesse. Notiamo che la caratterizzazione di Y in termini di pdf è appropriata se Y è una variabile aleatoria continua oppure mista, il che richiede necessariamente che X sia una variabile aleatoria continua oppure mista anch'essa (se X fosse discreta, anche Y sarebbe tale). Per semplicità, molte delle considerazioni fatte negli esempi, assumono (esplicitamente o implicitamente) che X sia una variabile aleatoria *continua*.

► **Esempio 4.7.** Consideriamo nuovamente la trasformazione lineare dell'esempio 4.1:

$$Y = aX + b,$$

raffigurata in Fig. 4.2. Qualunque sia $y \in \mathbb{R}$, e per ogni $a \neq 0$, l'equazione $y = g(x) = ax + b$ ammette l'unica soluzione

$$x = \frac{y - b}{a},$$

ed inoltre risulta

$$|g'(x)| = |a|,$$

per cui:

$$f_Y(y) = \frac{1}{|a|} f_X\left(\frac{y - b}{a}\right)$$

che coincide con il risultato ottenuto, derivando la CDF, nell'esempio 4.1. ◀

► **Esempio 4.8.** Consideriamo nuovamente la trasformazione quadratica dell'esempio 4.2:

$$Y = X^2$$

raffigurata in Fig. 4.3. Se $y < 0$, l'equazione $y = g(x) = x^2$ non ha soluzioni, e quindi $f_Y(y) = 0$. Se $y > 0$, si hanno due soluzioni:

$$x_1 = \sqrt{y}, \quad x_2 = -\sqrt{y}$$

ed inoltre

$$|g'(x)| = 2|x|,$$

per cui:

$$f_Y(y) = \frac{1}{2\sqrt{y}} [f_X(\sqrt{y}) + f_X(-\sqrt{y})] u(y),$$

che è lo stesso risultato ottenuto nell'esempio 4.2.⁴ Come applicazione del precedente risultato, si consideri il caso in cui $X \sim N(0, 1)$:

$$f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

e sia $Y = X^2$. Dalla relazione precedente, tenendo conto che X ha una pdf pari, si ha:

$$f_Y(y) = \frac{1}{\sqrt{y}} f_X(\sqrt{y}) u(y) = \frac{1}{\sqrt{2\pi y}} e^{-y/2} u(y)$$

che è la pdf di una variabile aleatoria di tipo *chi-square* con un grado di libertà, che si denota $Y \sim \chi^2(1)$. Notiamo che per $y \rightarrow 0$ tale pdf diverge, ma l'area sottesa si mantiene comunque finita. ◀

► **Esempio 4.9.** Consideriamo la trasformazione *iperbolica*:

$$Y = 1/X.$$

Per $y \neq 0$, l'equazione $y = g(x) = 1/x$ ha l'unica soluzione

$$x = \frac{1}{y},$$

ed inoltre si ha

$$|g'(x)| = \frac{1}{x^2},$$

per cui:

$$f_Y(y) = \frac{1}{y^2} f_X\left(\frac{1}{y}\right) \quad (4.3)$$

Per quanto riguarda il caso $y = 0$, l'equazione $y = g(x)$ non ha soluzione per $y = 0$, per cui la pdf di Y è nulla in tal punto. Come applicazione del precedente risultato, si consideri il caso in cui $X \sim \text{Cauchy}(\alpha)$, ovvero X ha una pdf di tipo Cauchy di parametro α :

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

Si verifica facilmente, applicando la (4.3), che la variabile aleatoria $Y = 1/X$ risulta anch'essa Cauchy, e precisamente $Y \sim \text{Cauchy}(1/\alpha)$:

$$f_Y(y) = \frac{1/(\alpha\pi)}{y^2 + 1/\alpha^2}.$$

In altri termini, la classe delle variabili aleatorie di Cauchy è *chiusa* rispetto all'operazione di reciprocazione. Notiamo che sebbene l'espressione generale (4.3) sia stata ricavata per $y \neq 0$, la $f_Y(y)$ di Cauchy ottenuta può essere prolungata per continuità in $y = 0$. ◀

Negli esempi precedenti, abbiamo incontrato casi in cui il teorema non è applicabile, e precisamente per quei valori di $y = g(x)$ in corrispondenza dei quali la derivata $g'(x)$ si annulla. Se tali punti y sono isolati, il valore di $f_Y(y)$ è inessenziale, in quanto la pdf compare solo in relazioni *integrali*, e quindi il suo valore in un punto isolato non è rilevante (l'integrale della pdf non cambia). Può accadere che, nei punti y in cui il teorema non è applicabile, la pdf sia divergente (cfr. la variabile aleatoria chi-square dell'esempio 4.8 per $y = 0$), oppure che essa si possa prolungare per continuità (cfr. la variabile aleatoria Cauchy dell'esempio 4.9 per $y = 0$).

Diversa è la situazione se, per un determinato y , l'equazione $y = g(x)$ ammette una infinità *continua* di soluzioni, come accade ad esempio se $g(x)$ presenta uno o più tratti costanti con

⁴ Se $y = 0$ il teorema non è applicabile, poiché anche se l'equazione $y = g(x)$ ha una sola soluzione $x = 0$, in corrispondenza di tale punto la derivata $g'(x)$ è nulla. Tuttavia, essendo $y = 0$ un punto isolato, il valore della pdf nel punto $y = 0$ è inessenziale, a patto, ovviamente, di essere sicuri che in $y = 0$ non ci sia un impulso (si veda la discussione che segue l'esempio 4.9).

ordinata pari ad y (si noti che in tal caso si ha anche $g'(x) = 0$ per tutti i valori x corrispondenti al tratto costante). In tal caso, la pdf di Y presenta nel punto y un *impulso di Dirac*, la cui area va determinata direttamente calcolando $P(Y = y)$. L'esempio che segue chiarirà meglio questo concetto.

► **Esempio 4.10.** Consideriamo nuovamente la trasformazione (amplificatore con saturazione) dell'esempio 4.4, raffigurata in Fig. 4.7. Tale trasformazione ha due tratti costanti, di ordinata $y = -d_y$ e $y = d_y$; anticipiamo pertanto la presenza di due impulsi di Dirac, centrati in $y = \pm d_y$, le cui aree dobbiamo determinare. Applichiamo comunque il teorema nei punti dove è consentito. Per $|y| > d_y$, l'equazione $y = g(x)$ non ha soluzioni, per cui $f_Y(y) = 0$. Per $|y| < d_y$, l'equazione $y = g(x)$ ha una sola soluzione $x = y/a$. Il calcolo della derivata prima per $|y| < d_y$ fornisce

$$|g'(x)| = a;$$

pertanto per tutti i valori $y \neq \pm d_y$ l'applicazione del teorema fondamentale fornisce la parte "convenzionale" $h(y)$ della pdf (corrispondente alla derivata convenzionale della CDF):

$$h(y) = \begin{cases} 0, & y < -d_y; \\ \frac{1}{a} f_X\left(\frac{y}{a}\right), & -d_y < y < d_y; \\ 0, & y > d_y; \end{cases}$$

Passiamo ora a determinare le aree degli impulsi. Si ha:

$$\begin{aligned} P(Y = -d_y) &= P(X \leq -d_x) = F_X(-d_x), \\ P(Y = d_y) &= P(X \geq d_x) = 1 - F_X(d_x), \end{aligned}$$

e quindi la pdf si scrive come:

$$f_Y(y) = h(y) + F_X(-d_x) \delta(y + d_y) + [1 - F_X(d_x)] \delta(y - d_y),$$

che coincide con il risultato ottenuto nell'esempio 4.4 derivando la CDF. ◀

4.3 Problema inverso: determinazione di $g(x)$

Finora ci siamo occupati del problema di caratterizzare la variabile aleatoria Y ottenuta dalla variabile aleatoria X mediante una nota trasformazione $g(x)$. Questo problema è denominato *problema diretto*, per contrasto con il seguente, che denomineremo *problema inverso*: date due variabili aleatorie X ed Y , con CDF (o pdf, o DF) assegnate, trovare la particolare trasformazione $Y = g(X)$ che consente di ottenere Y da X . Nel seguito, supporremo in particolare che le due variabili aleatorie X ed Y siano caratterizzate per mezzo delle loro CDF $F_X(x)$ e $F_Y(y)$, e supporremo inoltre che X ed Y siano due variabili aleatorie continue con CDF strettamente monotone, e quindi invertibili.⁵

Per semplificare lo studio del problema, converrà riguardare la trasformazione da X ad Y come realizzata in due passi (Fig. 4.15), ovvero come composta da due successive trasformazioni g_1 e g_2 : (i) mediante g_1 , si passa da X ad una variabile aleatoria uniforme $U \sim U(0, 1)$; (ii) mediante g_2 , si passa da una variabile aleatoria uniforme $U \sim U(0, 1)$ ad Y .

- (i) *Da X ad una variabile aleatoria uniforme*: vogliamo passare dalla variabile aleatoria X ad una variabile aleatoria $U \sim U(0, 1)$. È immediato verificare che la trasformazione appropriata è $g_1(x) = F_X(x)$.

⁵L'ipotesi di invertibilità delle CDF non è strettamente necessaria, nel paragrafo 4.3.1 vedremo una importante generalizzazione ottenuta rimuovendo tale ipotesi.

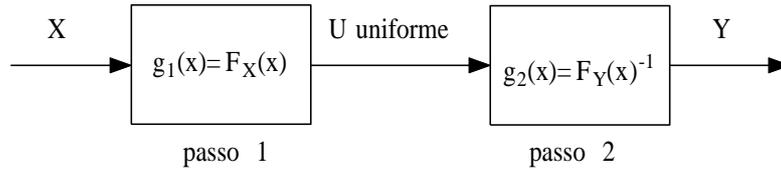


Fig. 4.15. La trasformazione di una variabile aleatoria X in una variabile aleatoria Y si articola in due passi: a partire da X , si genera una variabile aleatoria $U \sim U(0, 1)$; successivamente, da U si genera la variabile aleatoria Y .

Prova. Infatti, consideriamo la trasformazione $U = g_1(X) = F_X(X)$ e calcoliamo la CDF di U . Se $u < 0$ si ha:

$$F_U(u) = P(U \leq u) = P[F_X(X) \leq u] = 0,$$

in quanto i valori assunti da una CDF sono sempre non negativi, mentre se $u \geq 1$ si ha:

$$F_U(u) = P(U \leq u) = P[F_X(X) \leq u] = 1,$$

in quanto i valori assunti da una CDF sono sempre non superiori ad 1. Infine, se $u \in [0, 1[$, si ha:

$$F_U(u) = P(U \leq u) = P[F_X(X) \leq u] = P[X \leq F_X^{-1}(u)] = F_X[F_X^{-1}(u)] = u.$$

Si noti che abbiamo applicato la $F_X^{-1}(\cdot)$ ad entrambi i membri della disuguaglianza perchè abbiamo supposto che la CDF di X sia strettamente monotona (crescente) e quindi invertibile. In definitiva, mettendo insieme i tre casi, la CDF di U è data da:

$$F_U(u) = \begin{cases} 0, & u < 0; \\ u, & u \in [0, 1[; \\ 1, & u \geq 1; \end{cases}$$

e questa è proprio la CDF di una variabile aleatoria $U \sim U(0, 1)$, per cui resta dimostrato l'asserto. \square

- (ii) *Da una variabile aleatoria uniforme a Y :* abbiamo a disposizione una variabile aleatoria $U \sim U(0, 1)$ e vogliamo trasformarla in una variabile aleatoria $Y = g_2(U)$ con preassegnata CDF $F_Y(y)$. Si può verificare in tal caso che la trasformazione cercata è $g_2(x) = F_Y^{-1}(x)$, coincide cioè con l'inversa (che abbiamo supposto esistente) della CDF desiderata.

Prova. Per verificarlo, denotiamo con $\widehat{F}_Y(y)$ la CDF di $Y = F_Y^{-1}(U)$ e dimostriamo che essa coincide con $F_Y(y)$. Si ha:

$$\widehat{F}_Y(y) = P(Y \leq y) = P[F_Y^{-1}(U) \leq y] = P[U \leq F_Y(y)] = F_U[F_Y(y)] = F_Y(y)$$

perchè $F_Y(y) \in [0, 1]$ ed U è una variabile aleatoria uniforme in $(0, 1)$, quindi con CDF $F_U(u) = u$ per $u \in [0, 1]$. Resta pertanto dimostrato che $\widehat{F}_Y(y) = F_Y(y)$, e quindi la trasformazione $g_2(x)$ coincide proprio con l'inversa della CDF di Y .

Una dimostrazione più semplice, e meno formale, è la seguente: abbiamo mostrato in precedenza che per passare da una variabile aleatoria Y qualsiasi ad una variabile aleatoria uniforme U occorre effettuare la trasformazione $U = F_Y(Y)$. Se, allora, $F_Y(y)$ è invertibile, la variabile aleatoria Y si otterrà dalla variabile aleatoria uniforme U come $Y = F_Y^{-1}(U)$. \square

A questo punto, come precedentemente osservato, il caso della trasformazione di una variabile aleatoria X con CDF arbitraria in una variabile aleatoria Y con CDF arbitraria può essere affrontato sfruttando i risultati dei due casi precedenti, e cioè articolando la trasformazione in due passi (Fig. 4.15):

- (i) nel primo passo, si trasforma X in una variabile aleatoria uniforme $U \sim U(0, 1)$, mediante la trasformazione $g_1(x) = F_X(x)$;

(ii) nel secondo passo, dalla variabile aleatoria uniforme $U \sim U(0, 1)$, si ottiene Y mediante la trasformazione $g_2(x) = F_Y^{-1}(x)$.

La trasformazione g complessiva è chiaramente la *funzione composta* di g_1 (funzione interna) e g_2 (funzione esterna), e cioè:

$$g(x) = g_2[g_1(x)] = F_Y^{-1}[F_X(x)] \quad (4.4)$$

per cui la variabile Y si ottiene da X con la trasformazione

$$Y = F_Y^{-1}[F_X(X)].$$

► *Esempio 4.11.* Determiniamo la trasformazione $g(x)$ che consente di passare da una variabile aleatoria esponenziale $X \sim \text{Exp}(\lambda)$ ad una variabile aleatoria Rayleigh $Y \sim \text{Rayleigh}(b)$. La CDF di X è:

$$F_X(x) = (1 - e^{-\lambda x}) u(x)$$

mentre quella di Y è:

$$F_Y(y) = (1 - e^{-\frac{y^2}{b}}) u(y).$$

Per individuare la $g(x)$, conviene riscrivere la (4.4) nella forma:

$$F_Y[g(x)] = F_X(x),$$

che va riguardata come un'equazione nell'incognita $g(x)$ e risolta rispetto all'incognita. Sostituendo le espressioni delle CDF, si ha:

$$\left(1 - e^{-\frac{g^2(x)}{b}}\right) u[g(x)] = (1 - e^{-\lambda x}) u(x).$$

Per $x < 0$, il secondo membro si annulla, per cui posso assumere $g(x) \equiv 0$ per $x < 0$; invece, per $x \geq 0$, il secondo membro si annulla solo per $x = 0$, per cui deve risultare necessariamente $g(x) \geq 0$ cosicché $u[g(x)] = 1$; in tal caso, si ha:

$$1 - e^{-\frac{g^2(x)}{b}} = 1 - e^{-\lambda x},$$

da cui con semplici passaggi algebrici si ricava:

$$g^2(x) = \lambda x b \Rightarrow g(x) = \sqrt{\lambda x b}.$$

Si noti che nella risoluzione abbiamo scelto la soluzione non negativa per $g(x)$ per tenere conto della condizione $g(x) \geq 0$ ricavata in precedenza; questo corrisponde al fatto che, poiché la variabile aleatoria di Rayleigh è positiva, allora la trasformazione $g(x)$ cercata deve essere non negativa. ◀

4.3.1 Generazione di una variabile aleatoria con CDF assegnata

Un'importantissima applicazione dei risultati del precedente paragrafo è quella della *generazione* di una variabile aleatoria con CDF assegnata. Infatti, se vogliamo generare una variabile aleatoria X con CDF $F_X(x)$ (supposta invertibile), basta generare (Fig. 4.16) una variabile aleatoria $U \sim U(0, 1)$ e trasformarla secondo la legge $g(x) = F_X^{-1}(x)$. È sufficiente allora disporre di un generatore di variabili aleatorie uniformi in $(0, 1)$ che, come vedremo nel § 4.3.2, può essere facilmente realizzato mediante un calcolatore.

Notiamo che, poiché l'inversa della CDF è la legge che definisce il *percentile* (cfr. § 3.2.3), tale tecnica di generazione è anche denominata *metodo della trasformazione percentile*.

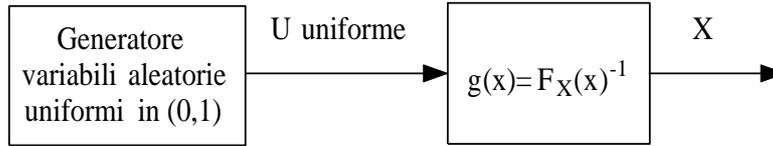


Fig. 4.16. La generazione di una variabile aleatoria X con CDF $F_X(x)$ invertibile si può effettuare a partire da un generatore di variabili aleatorie uniformi $U \sim U(0, 1)$, applicando all'uscita di quest'ultimo la trasformazione $g(x) = F_X^{-1}(x)$.

► **Esempio 4.12.** Determiniamo la trasformazione che consente di generare una variabile aleatoria esponenziale $X \sim \text{Exp}(\lambda)$ a partire da una variabile aleatoria uniforme $U \sim U(0, 1)$. Poiché:

$$F_X(x) = (1 - e^{-\lambda x}) u(x),$$

allora si ha:

$$g(x) = F_X^{-1}(x) = -\frac{1}{\lambda} \ln(1 - x).$$

Osserviamo però che, se $U \sim U(0, 1)$, allora anche $1 - U \sim U(0, 1)$. Allora, più semplicemente, possiamo scrivere:

$$g(x) = -\frac{1}{\lambda} \ln(x).$$

Poiché X è una variabile aleatoria positiva, la trasformazione $g(x)$ è non negativa. ◀

► **Esempio 4.13.** Determiniamo la trasformazione che consente di generare una variabile aleatoria Rayleigh $X \sim \text{Rayleigh}(b)$ a partire da una variabile aleatoria uniforme $U \sim U(0, 1)$. Poiché:

$$F_X(x) = (1 - e^{-\frac{x^2}{b}}) u(x),$$

allora si ha:

$$g(x) = F_X^{-1}(x) = \sqrt{-b \ln(1 - x)},$$

dove nella determinazione dell'inversa abbiamo scelto la soluzione positiva perchè la variabile aleatoria di Rayleigh è positiva. Anche qui, poiché se $U \sim U(0, 1)$, anche $1 - U \sim U(0, 1)$, possiamo scrivere più semplicemente:

$$g(x) = \sqrt{-b \ln(x)}.$$

Poiché X è una variabile aleatoria positiva, la trasformazione $g(x)$ è non negativa. ◀

In molti casi, la $F_X(x)$ non ha una espressione analitica semplice e pertanto, sebbene sia strettamente monotona, non è semplice calcolarne l'inversa $F_X^{-1}(x)$; ciò accade, ad esempio, se $X \sim N(\mu, \sigma)$, e quindi la CDF è espressa in termini della funzione non elementare $\mathbb{G}(x)$. Se allora si riesce a mettere in relazione la variabile aleatoria X con altre variabili aleatorie Z_1, Z_2, \dots, Z_n di più semplice generazione, mediante una legge del tipo $X = f(Z_1, Z_2, \dots, Z_n)$, è possibile risolvere il problema della generazione di X in due passi:

- (i) si genera ciascuna delle variabili aleatorie Z_1, Z_2, \dots, Z_n con il metodo della trasformazione percentile;
- (ii) si applica alle Z_1, Z_2, \dots, Z_n la trasformazione f per ottenere X .

Un esempio di applicazione di tale tecnica, di fondamentale importanza nelle applicazioni pratiche, riguarda proprio la generazione delle *variabili aleatorie gaussiane* e sarà discusso nell'esempio 6.10 (in quanto richiede concetti che saranno introdotti nel seguito).

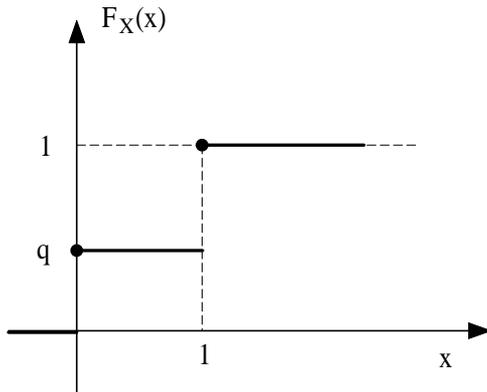


Fig. 4.17. La CDF $F_X(x)$ della variabile aleatoria $X \sim \text{Bern}(p)$.

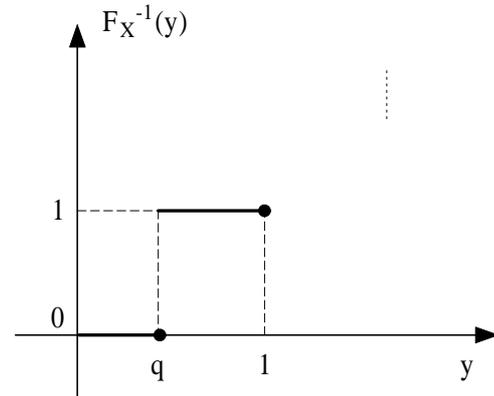


Fig. 4.18. L'inversa sinistra $F_X^{-1}(y)$ della CDF della variabile aleatoria $X \sim \text{Bern}(p)$ raffigurata in Fig. 4.17.

Un altro caso in cui il metodo della trasformazione percentile non sembrerebbe applicabile è quello in cui la $F_X(x)$ non è strettamente monotona e, quindi, non è invertibile. Tale limitazione escluderebbe l'impiego della tecnica di generazione nei casi in cui la CDF della variabile aleatoria da generare presenti uno o più tratti costanti; in particolare, ne precluderebbe l'uso per la generazione di variabili aleatorie discrete, che hanno una CDF costante a tratti. In realtà, per applicare il metodo della trasformazione percentile anche a questi casi, è sufficiente generalizzare la definizione di inversa della CDF, ponendo, per ogni $y \in [0, 1]$,

$$F_X^{-1}(y) \triangleq \inf\{x \in \overline{\mathbb{R}} \text{ tali che } F_X(x) \geq y\}. \quad (4.5)$$

Osserviamo che, se $F_X(x)$ è strettamente monotona, la $F_X^{-1}(y)$ definita dalla (4.5) si riduce all'inversa convenzionale; altrimenti, se ad esempio la CDF $F_X(x)$ presenta un tratto costante nell'intervallo $[x_1, x_2]$ di ordinata pari a y , è facile verificare che $F_X^{-1}(y) = x_1$. La funzione definita dalla (4.5) viene a volte denominata inversa "sinistra", in quanto si può facilmente verificare che $F_X[F_X^{-1}(y)] = y$, mentre in generale risulta $F_X^{-1}[F_X(x)] \neq x$; inoltre poichè $F_X(x)$ è monotona crescente, anche la funzione $F_X^{-1}(y)$ definita dalla (4.5) è monotona crescente. Si può allora facilmente verificare che la dimostrazione sviluppata nel § 4.3 al punto (ii) rimane valida, a patto di sostituire all'inversa convenzionale l'inversa sinistra. In particolare, il metodo della trasformazione percentile risulta ancora applicabile, come mostrato dal seguente esempio.

► *Esempio 4.14.* Si vuole generare una variabile aleatoria $X \sim \text{Bern}(p)$, la cui CDF è raffigurata in Fig. 4.17. Calcoliamo prima l'inversa sinistra $F_X^{-1}(y)$, in accordo alla (4.5). Si ha:

$$\begin{aligned} y = 0 &\Rightarrow \inf\{x \in \overline{\mathbb{R}} \text{ tali che } F_X(x) \geq y\} = \inf\{-\infty, +\infty\} = -\infty \\ y \in]0, q] &\Rightarrow \inf\{x \in \overline{\mathbb{R}} \text{ tali che } F_X(x) \geq y\} = \inf\{[0, +\infty]\} = 0 \\ y \in]q, 1] &\Rightarrow \inf\{x \in \overline{\mathbb{R}} \text{ tali che } F_X(x) \geq y\} = \inf\{[1, +\infty]\} = 1 \end{aligned}$$

per cui:

$$F_X^{-1}(y) = \begin{cases} -\infty, & y = 0; \\ 0, & y \in]0, q]; \\ 1, & y \in]q, 1]; \end{cases}$$

raffigurata in Fig. 4.18. Si può osservare che l'inversa sinistra $F_X^{-1}(x)$ è continua da sinistra (mentre la CDF è continua da destra), e che si può ottenere con una procedura grafica molto semplice: a partire dalla CDF,

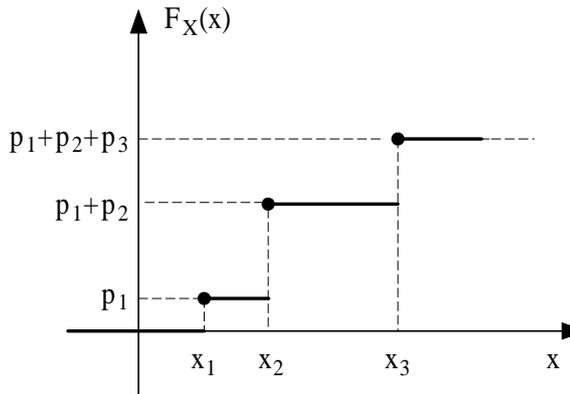


Fig. 4.19. La CDF $F_X(x)$ di una variabile aleatoria discreta X .

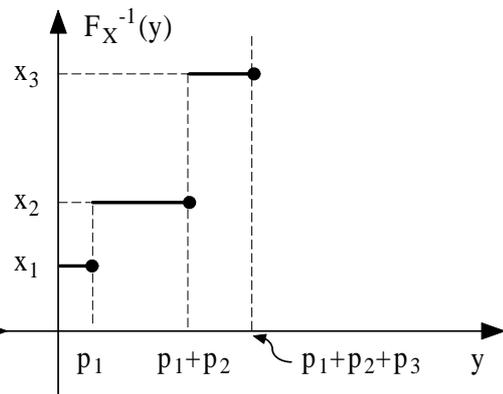


Fig. 4.20. L'inversa sinistra $F_X^{-1}(y)$ della CDF della variabile aleatoria X discreta raffigurata in Fig. 4.19.

si scambiano gli assi x ed y . Pertanto, a partire da $U \sim U(0, 1)$ e tenendo conto della forma dell'inversa sinistra, la tecnica di generazione è molto semplice:

1. si genera un valore $U \in [0, 1]$;
2. se $U \in [0, p_1]$, si pone $X = x_1$;
3. se $U \in [p_1, 1]$, si pone $X = x_2$.

Si noti che abbiamo arbitrariamente assegnato anche a $U = 0$ il valore $X = x_1$, e non il valore $X = -\infty$; questo non altera la probabilità che $X = x_1$, e quindi la distribuzione di X , poiché, essendo U una variabile aleatoria continua, si ha $P(U = 0) = 0$. ◀

Con un modesto sforzo di generalizzazione, possiamo estendere la procedura delineata nell'esempio 4.14 alla generazione di una *qualunque* variabile aleatoria discreta. Infatti, siano $\mathcal{X} = \{x_1, x_2, x_3, \dots\}$ i valori assunti dalla variabile aleatoria X , che supporremo, senza ledere la generalità, *ordinati in senso crescente*, vale a dire $x_1 < x_2 < x_3 < \dots$, e siano p_1, p_2, p_3, \dots le rispettive probabilità, dove $p_i \triangleq P(X = x_i)$. La CDF di X sarà una funzione costante a tratti, con gradini (Fig. 4.19) posti alle quote $p_1, p_1 + p_2, p_1 + p_2 + p_3, \dots$ e posizionati orizzontalmente in x_1, x_2, x_3, \dots ; l'inversa sinistra, ottenuta scambiando gli assi della CDF, sarà ancora una funzione costante a tratti, con gradini (Fig. 4.20) situati alle quote x_1, x_2, x_3, \dots e posizionati orizzontalmente in $p_1, p_1 + p_2, p_1 + p_2 + p_3, \dots$. Pertanto, a partire da $U \sim U(0, 1)$ e tenendo conto della forma dell'inversa sinistra, la generazione di X avviene nel seguente modo:

1. si genera un valore $U \in [0, 1]$;
2. se $U \in [0, p_1]$, si pone $X = x_1$;
3. altrimenti, si determina il valore di $k \in \{2, 3, \dots\}$ per cui si ha

$$p_1 + p_2 + \dots + p_{k-1} < U \leq p_1 + p_2 + \dots + p_k,$$

e si pone $X = x_k$.

In pratica, la procedura è equivalente a suddividere l'intervallo $(0, 1)$ in tanti sottointervalli di ampiezza pari a p_1, p_2, p_3, \dots : se il valore di U cade nel k -esimo sottointervallo, allora $X = x_k$. La complessità dell'algoritmo di generazione [11] coincide sostanzialmente con quella dell'algoritmo di ricerca del sottointervallo nel quale cade la variabile aleatoria U .

4.3.2 Generazione automatica di numeri casuali

Nel precedente paragrafo abbiamo visto che, a partire da una variabile aleatoria U uniforme in $(0, 1)$, è possibile generare una variabile aleatoria X con CDF arbitraria $F_X(x)$ mediante la trasformazione percentile $X = F_X^{-1}(U)$; tuttavia non abbiamo fornito una soluzione per generare, in maniera computazionalmente efficiente, i valori di una variabile aleatoria uniforme $(0, 1)$. Questo problema può essere risolto, in via approssimata, se disponiamo di una tecnica per la generazione di numeri casuali *interi*.

Evidentemente, la tecnica più semplice e teoricamente corretta per generare un numero casuale intero è quella di *eseguire* un esperimento aleatorio, ed associare ad ogni risultato un numero intero, in accordo con la definizione stessa di variabile aleatoria. Ad esempio, lanciando un dado potremo generare numeri casuali da 1 a 6; estraendo una pallina da un paniere della tombola, numeri casuali da 1 a 90, e così via. Tali procedure, tuttavia, hanno lo svantaggio di non poter essere facilmente automatizzate, ed essendo inoltre di tipo manuale o meccanico, risultano estremamente lente.

Con l'avvento dei calcolatori elettronici, si è pensato di ricorrere a procedure aritmetiche di tipo ricorsivo, che fossero di semplice realizzazione e consentissero quindi la generazione rapida di sequenze x_0, x_1, \dots di numeri casuali molto lunghe (da migliaia a milioni di campioni). La maggior parte di tali procedure generano il valore x_{n+1} della sequenza a partire dal valore precedente x_n , mediante una legge *ricorsiva* del tipo $x_{n+1} = g(x_n)$, dove g è una funzione opportuna, ed il primo valore x_0 della sequenza è denominato *seme* (in inglese, "seed"). È chiaro che, osservato un valore della sequenza e conoscendo la legge g , è possibile prevedere esattamente tutti i valori successivi; per questo motivo la sequenza non è realmente aleatoria, ma viene detta *pseudo-aleatoria*. Quello che realmente interessa, allora, è che la sequenza pseudo-aleatoria generata non sia distinguibile, da parte di un osservatore esterno che non conosca la legge g , da una sequenza realmente aleatoria, generata cioè con metodi puramente casuali.

4.3.3 Algoritmo "middle-square" (Von Neumann)

Una delle prime procedure di tipo ricorsivo ad essere proposte fu l'algoritmo cosiddetto "middle-square", ideato da John Von Neumann nel 1946. L'algoritmo è il seguente: si parte da un seme x_0 di 4 cifre e lo si eleva al quadrato, ottenendo un numero del quale si conservano le 4 cifre intermedie (si eliminano le ultime due cifre); tali cifre costituiscono il numero x_1 , che viene nuovamente elevato al quadrato, e così via. Ad esempio, la sequenza generata a partire dal seme 5232 è la seguente:

$$\begin{aligned} x_0 &= 5232 \\ 5232^2 &= 27|3738|24 \rightarrow x_1 = 3738 \\ 3738^2 &= 13|9726|44 \rightarrow x_2 = 9726 \\ 9726^2 &= 94|5950|76 \rightarrow x_3 = 5950 \\ &\dots \end{aligned}$$

Dalla sequenza intera ottenuta è possibile ottenere numeri interi in $(0, 1)$ semplicemente spostando la virgola in prima posizione, ovvero dividendo ciascun numero per 10 000: ad esempio, la sequenza del precedente esempio genera la seguente successione di valori in $(0, 1)$:

$$0.5232 \quad 0.3738 \quad 0.9726 \quad 0.5950 \quad \dots$$

È chiaro che, essendo solo 10 000 i numeri di quattro cifre, e poichè ogni numero dipende solo da quello precedentemente generato, la sequenza ottenuta sarà necessariamente periodica, con periodo *al più* pari a 10 000. In realtà, il principale svantaggio di tale procedura è che le proprietà della sequenza generata dipendono in maniera critica dalla scelta del seme iniziale; ad esempio, la scelta $x_0 = 0000$ produce la sequenza banale

0000 0000 0000 ...

Ma anche scelte meno banali del seme possono portare a risultati altrettanto sgradevoli: ad esempio, scegliendo $x_0 = 2100$, si ottiene la sequenza composta dai soli quattro valori interi

2100 4100 8100 6100

che si ripetono indefinitamente. Proprio a causa della sensibilità rispetto alla scelta del seme iniziale, il metodo “middle-square” è stato presto abbandonato, e l’attenzione degli studiosi si è spostata verso tecniche ricorsive che fossero al tempo stesso più efficienti computazionalmente (l’algoritmo “middle-square” ricorre ad una elevazione al quadrato, che ha una complessità algoritmica non trascurabile) e tali da garantire proprietà ottimali o quasi ottimali delle sequenze generate.

4.3.4 Algoritmo lineare congruente

Al giorno d’oggi, l’algoritmo più diffusamente utilizzato per la generazione di numeri casuali è quello cosiddetto *lineare congruente*, nel quale la ricorsione $x_{n+1} = g(x_n)$ può essere espressa come:

$$x_{n+1} = (a x_n + c) \pmod{m} \quad (4.6)$$

dove il *moltiplicatore* a , l’*incremento* c ed il *modulo* m sono tutti numeri interi non negativi. L’equazione (4.6) descrive sostanzialmente una ricorsione *lineare*, in cui tuttavia il risultato è preso in aritmetica “modulo m ”, considerando cioè il resto della divisione per m e ottenendo quindi sempre numeri interi compresi tra 0 ed $m - 1$, estremi inclusi. Ad esempio, scegliendo $a = c = 7$, $m = 10$ ed un seme $x_0 = 7$, si ottiene la sequenza:

7 6 9 0 7 6 9 0 ...

che risulta chiaramente periodica di periodo 4. Tale periodicità è una proprietà generale del generatore lineare congruente: tutte le sequenze generate in base alla (4.6) saranno periodiche di periodo minore o uguale ad m , in quanto composte al più da m valori. Per avere un buon generatore, allora, dovremo scegliere m molto grande: in pratica converrebbe scegliere m pari al massimo numero intero rappresentabile nella parola macchina del calcolatore, quindi ad esempio $m = 2^{16}$ per un calcolatore a 16 bit, oppure $m = 2^{32}$ per un calcolatore a 32 bit. Inoltre dobbiamo assicurarci che la sequenza generata sia a *massimo periodo*: affinché ciò accada, devono valere le seguenti condizioni [8]:

1. c ed a devono essere primi tra loro;
2. $a - 1$ dev’essere multiplo di ogni fattore primo di m ;
3. $a - 1$ dev’essere multiplo di 4 se m è multiplo di 4.

È chiaro che, nel caso di sequenze a massimo periodo, il periodo m dovrà eccedere significativamente la lunghezza tipica delle sequenze che utilizzeremo in una singola simulazione; se così non fosse, la periodicità della sequenza generata sarebbe chiaramente individuabile, e ciò ne comprometterebbe la natura pseudo-aleatoria.⁶

Una volta progettato un buon generatore di numeri casuali interi x_n tra 0 ed $m - 1$, possiamo ottenere un generatore⁷ di numeri casuali y_n tra 0 ed 1 semplicemente dividendo x_n per m :

$$y_n = \frac{x_n}{m}.$$

I numeri y_n così generati non riempiono tutto l'intervallo $(0, 1)$, ma si dispongono su un reticolo monodimensionale con spaziatura $1/m$; in pratica, non otterremo tutti i numeri reali tra 0 ed 1, ma soltanto i numeri *razionali* del tipo p/m , con $p \in \{0, 1, \dots, m - 1\}$. Se però m è molto grande, il reticolo è sufficientemente fitto da potersi ritenere una buona approssimazione⁸ dei numeri nell'intervallo $(0, 1)$.

Osserviamo inoltre che, se la sequenza è a massimo periodo, ogni valore p tra 0 ed $m - 1$ sarà assunto una ed una sola volta nel periodo, e quindi ogni valore razionale p/m sarà anch'esso assunto una ed una sola volta nel periodo; in altri termini, osservando una sequenza di lunghezza pari al periodo m , otterrei una distribuzione perfettamente *uniforme* (sebbene discreta) di valori in $(0, 1)$. In pratica, se m è sufficientemente elevato, è possibile osservare solo sequenze di lunghezza molto minore del periodo, per cui la legge di distribuzione dei valori è solo approssimativamente uniforme, se il segmento di sequenza osservato è sufficientemente lungo.

Una classe di generatori lineari congruenti particolarmente utilizzata è quella dei generatori *puramente moltiplicativi*, per i quali cioè $c = 0$. In questo caso, la condizione $c = 0$ impedisce di raggiungere il massimo periodo m , perché dobbiamo escludere dai valori ammissibili per la sequenza il valore 0, che porterebbe il generatore in un ciclo composto da tutti 0; tuttavia esistono condizioni che garantiscono comunque la possibilità di raggiungere un periodo massimo pari ad $m - 1$, e precisamente ciò accade se:

1. m è primo;
2. a è una *radice primitiva*⁹ di m ;
3. il seme x_0 è diverso da zero.

Ad esempio, il generatore `rand` utilizzato in Matlab¹⁰ è di tipo puramente moltiplicativo, con $c = 0$, $m = 2^{31} - 1 = 2147483647$ ed $a = 7^5 = 16807$, e periodo pari a $m - 1 = 2^{31} - 2 = 2147483646$. Tale generatore è stato proposto per la prima volta da S. K. Park e K. W. Miller in [9] ed è quello più comunemente implementato nella maggior parte dei linguaggi di programmazione moderni (generatore di Park e Miller).

⁶Una regola pratica [10] è che il periodo del generatore deve eccedere il *quadrato* della massima lunghezza delle sequenze generate in una simulazione.

⁷Tali generatori fanno parte delle funzioni di libreria dei moderni linguaggi di programmazione, nei quali assumono la denominazione di funzione "rand", o similari.

⁸Consideriamo anche che se m rappresenta il massimo numero rappresentabile in macchina, la differenza $1/m$ tra due numeri razionali consecutivi è la minima che posso rappresentare su una macchina con registri di dimensione finita.

⁹Un numero intero a si dice [3] [8] radice primitiva di m se il più piccolo valore di n tale che $a^n - 1 = 0 \pmod{m}$ è $n = m - 1$.

¹⁰Fino alla versione 4: nella versione 5 e successive si utilizza un generatore basato su un algoritmo più sofisticato di quello lineare congruente, che assicura un periodo pari a 2^{1492} (si veda <http://www.mathworks.com/company/newsletter/pdf/Cleve.pdf> per maggiori dettagli sui generatori impiegati in Matlab).

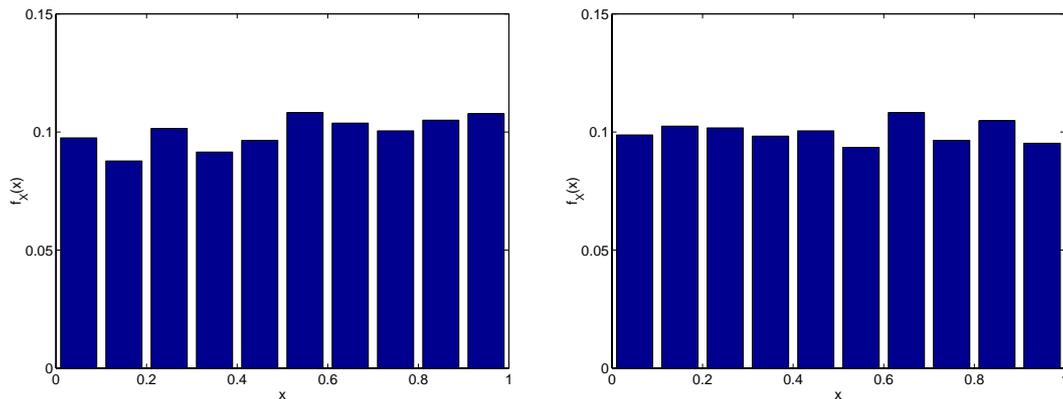


Fig. 4.21. Istogrammi di $N = 4000$ valori generati dal generatore “good” (a sinistra) e dal generatore “bad” (a destra).

4.3.5 Test statistici sui generatori★

Se osserviamo un intero periodo di una sequenza all’uscita di un generatore lineare congruente, la condizione di massimo periodo implica che ogni numero intero si presenti una ed una sola volta, e quindi garantisce l’uniformità dei numeri generati; tuttavia, tale condizione da sola non è sufficiente per assicurare alla sequenza generata una natura realmente *aleatoria* o *pseudo-aleatoria*. Si pensi, ad esempio, al seguente generatore lineare congruente:

$$x_{n+1} = (x_n + 1) \pmod{m}; \quad (4.7)$$

esso è evidentemente a massimo periodo, in quanto inizializzato ad esempio con $x_0 = 0$ genererà la sequenza:

$$0 \ 1 \ 2 \ \dots \ m-1 \ 0 \ 1 \ \dots$$

che ha periodo m , tuttavia la sequenza generata (una “rampa” di valori discreti) non ha evidentemente proprietà aleatorie o pseudo-aleatorie. Per casi meno evidenti, gli studiosi hanno messo a punto svariati *test statistici* [8], che consentono di verificare la capacità di un generatore di simulare il comportamento aleatorio.

Un test semplice per verificare la distribuzione uniforme o quasi uniforme dei numeri generati consiste nel generare una sequenza sufficientemente lunga (ma di lunghezza molto minore del massimo periodo) e calcolarne l’*istogramma* dei valori (il numero dei valori che cadono tra 0.0 e 0.1, tra 0.1 e 0.2 e così via), il che corrisponde in pratica, se si divide il numero dei valori in ciascun intervallo per l’ampiezza Δx dell’intervallo, ad effettuare una stima empirica della pdf dei valori generati. Ad esempio, consideriamo i due seguenti generatori:

- “good”: generatore di Park e Miller (lineare congruente con $c = 0$, $m = 2^{31} - 1 = 2147483647$ ed $a = 7^5 = 16807$);
- “bad”: generatore lineare congruente con $c = 0$, $m = 2^{17} = 131072$ ed $a = 1277$.

In Fig. 4.21, sono riportati gli istogrammi di $N = 4000$ valori generati in $(0, 1)$ da ciascuno dei due generatori: si può notare che per entrambi l’ipotesi di distribuzione uniforme è ben verificata.

Tuttavia, abbiamo osservato che non basta che la distribuzione sia uniforme, ma occorre verificare che non ci sia una “regolarità” facilmente identificabile nella sequenza generata. Un test semplice per individuare tali regolarità consiste nel diagrammare su un piano cartesiano le

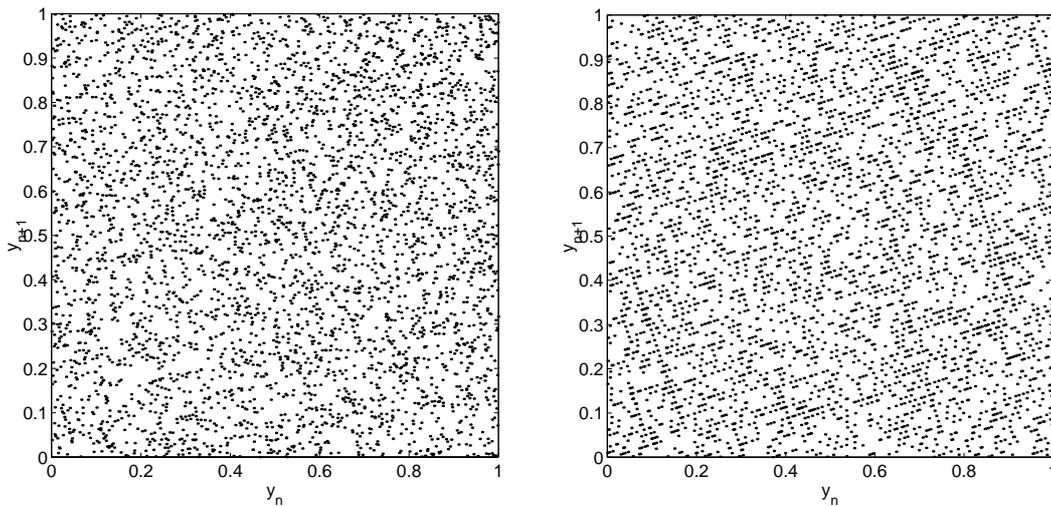


Fig. 4.22. Diagrammi delle coppie (y_n, y_{n+1}) di valori consecutivi generati in $(0, 1)$ per il generatore “good” (a sinistra) ed il generatore “bad” (a destra).

coppie di valori (x_n, x_{n+1}) generate: poichè $x_{n+1} = f(x_n)$, un “cattivo” generatore tenderà a presentare delle configurazioni regolari abbastanza evidenti. Ad esempio, il diagramma per il generatore (4.7) “a rampa” sarebbe composto dalle coppie

$$(0, 1) \quad (1, 2) \quad (2, 3) \quad (3, 4) \quad \dots$$

che si dispongono chiaramente a formare una retta. I corrispondenti diagrammi per i due generatori considerati sono riportati in Fig. 4.22, con riferimento ai valori y_n generati nell’intervallo $(0, 1)$: mentre il generatore “good” (a sinistra) non presenta schemi o regolarità facilmente individuabili, è facile invece notare che nel diagramma del generatore “bad” (a destra) i valori tendono a disporsi su rette oblique, il che induce a ritenere che “bad” non sia un buon generatore. Ovviamente esistono test più sofisticati di questi menzionati, per i quali si rimanda a [8] e [11]; per verificare la bontà di un generatore, è buona norma sottoporlo a più di un test. A tale proposito, gli studiosi di generatori di numeri casuali sono soliti citare l’affermazione: “Un generatore può ingannare un test qualche volta, e qualche test tutte le volte, ma non tutti i test tutte le volte”.

4.4 Esercizi proposti

Esercizio 4.1. Sia $X \sim N(0, 1)$, mostrare che $Y = |X|$ ha CDF $F_Y(y) = (2\mathbb{G}(y) - 1)u(y)$. Determinare inoltre la pdf di Y e rappresentarla graficamente.

Esercizio 4.2. Sia $X \sim N(0, 1)$, mostrare che $Y = 1/X^2$ ha CDF $F_Y(y) = 2[1 - \mathbb{G}(1/\sqrt{y})]u(y)$. Determinare inoltre la pdf di Y e rappresentarla graficamente.

Esercizio 4.3. Sia X una variabile aleatoria $X \sim \text{Cauchy}(1)$.

a) Dimostrare che la sua CDF è:

$$F_X(x) = \frac{1}{2} + \frac{1}{\pi} \arctan x.$$

b) Determinare CDF e pdf della variabile aleatoria Y ottenuta attraverso la seguente trasformazione:

$$Y = \begin{cases} 0, & X \leq 0; \\ X, & X > 0. \end{cases}$$

e rappresentarle graficamente.

c) Determinare CDF e pdf della variabile aleatoria Y ottenuta attraverso la seguente trasformazione:

$$Y = \begin{cases} -1, & X \leq 0; \\ X, & X > 0. \end{cases}$$

e rappresentarle graficamente.

Esercizio 4.4. Sia X la variabile aleatoria che descrive il numero di teste che si ottengono nel lancio di 3 monete bilanciate. Determinare la DF della variabile aleatoria $Y = 3 - X$.

Esercizio 4.5. Sia X una variabile aleatoria discreta che assume tutti i valori interi tra -2 e 2 (estremi inclusi) in maniera equiprobabile.

- Determinare la DF di $Y = |X|$ e rappresentarla graficamente;
- ripetere il punto 1 per la variabile aleatoria $Y = \text{sgn}(X) + X$;
- ripetere il punto 1 per la variabile aleatoria $Y = X^2 - 1$.

★ **Esercizio 4.6.** Mostrare che se $X \sim U(0, 2\pi)$, allora $Y = \tan(X)$ è $Y \sim \text{Cauchy}(1)$.

Esercizio 4.7. Si determini la pdf di Y definita attraverso la seguente trasformazione:

$$Y = \begin{cases} X, & |X| \leq X_{\max}; \\ X_{\max} \text{sgn}(X), & |X| > X_{\max}. \end{cases}$$

in termini della pdf di X . Particolarizzare il risultato al caso in cui $X \sim N(0, 3X_{\max})$.

★ **Esercizio 4.8.** Si determini la pdf di $Y = \sin(X + \phi)$, con $X \sim U(0, 2\pi)$ e ϕ costante.

Esercizio 4.9. Sia $X \sim U(-1, 3)$ una variabile aleatoria uniforme.

- Determinare la pdf di $Y = \sqrt{|X+1|}u(X+1)$ e rappresentarla graficamente;
- ripetere il punto 1 per $Y = |X|$;
- ripetere il punto 1 per $Y = \sqrt{|X|}$.

Esercizio 4.10. Sia $X \sim N(0, 1)$, e si consideri la trasformazione $Y = g(X)$, con

$$g(x) = \begin{cases} 0, & |x| < 1; \\ x - 1, & x \geq 1; \\ x + 1, & x \leq -1. \end{cases}$$

Determinare la pdf di Y e rappresentarla graficamente.

Esercizio 4.11. Sia $X \sim \text{Lap}(\lambda)$, e si consideri la trasformazione $Y = g(X)$, con

$$g(x) = \begin{cases} 1, & |x| < \frac{1}{\lambda}; \\ x - \frac{1}{\lambda}, & x \geq \frac{1}{\lambda}; \\ x + \frac{1}{\lambda}, & x \leq -\frac{1}{\lambda}. \end{cases}$$

Determinare la pdf di Y e rappresentarla graficamente.

Esercizio 4.12. Sia $X \sim \text{Exp}(\lambda)$, determinare la pdf di $Y = e^X$.

Esercizio 4.13. Sia $X \sim N(\mu, \sigma)$, determinare la pdf di $Y = e^X$ (pdf lognormale).

Esercizio 4.14. Sia X una variabile aleatoria con pdf $f_X(x) = 2e^{-2x}u(x)$.

- Determinare la pdf della variabile aleatoria $Y = 2X - 5$, e rappresentare le pdf di X ed Y sullo stesso diagramma;
- ripetere il punto 1 per $Y = -2X + 1$.

Esercizio 4.15. Sia X una variabile aleatoria con pdf $f_X(x) = e^{-x}u(x)$, e sia $Y = g(X)$ la variabile aleatoria ottenuta mediante la seguente trasformazione:

$$g(x) = \begin{cases} x, & x \leq 1; \\ 1/x, & x > 1. \end{cases}$$

Determinare la pdf della variabile aleatoria Y e rappresentarla graficamente.

Esercizio 4.16. Determinare la trasformazione che consente di generare una variabile aleatoria $X \sim U(0, 2\pi)$ a partire da una variabile aleatoria $U \sim U(0, 1)$.

[Risposta: $g(x) = 2\pi x$]

Esercizio 4.17. Determinare la trasformazione che consente di generare una variabile aleatoria $X \sim \text{Cauchy}(\alpha)$ a partire da una variabile aleatoria $U \sim U(0, 1)$.

[Risposta: $g(x) = \alpha \tan[\pi(x - 0.5)]$]

Esercizio 4.18. Determinare la trasformazione che consente di generare una variabile aleatoria $X \sim \text{Lap}(\lambda)$ a partire da una variabile aleatoria $U \sim U(0, 1)$.

[Risposta: $g(x) = (1/\lambda) \ln(2x)$, per $x \leq 1/2$; $g(x) = -(1/\lambda) \ln[2(1-x)]$, per $x \geq 1/2$]

Esercizio 4.19. Determinare la trasformazione che consente, a partire da una variabile aleatoria $U \sim U(0, 1)$, di generare una variabile aleatoria X di tipo *Weibull*, avente cioè pdf:

$$f_X(x) = \alpha x^{\alpha-1} e^{-x^\alpha} u(x),$$

con $\alpha \geq 0$. [Risposta: $g(x) = [-\ln(x)]^{1/\alpha}$]

Esercizio 4.20. Determinare la trasformazione che consente, a partire da una variabile aleatoria $U \sim U(0, 1)$, di generare una variabile aleatoria X di tipo *Pareto*, avente cioè pdf:

$$f_X(x) = \frac{\alpha - 1}{x^\alpha} u(x - 1)$$

con $\alpha > 1$. [Risposta: $g(x) = \left(\frac{1}{x}\right)^{\frac{1}{\alpha-1}}$]

Esercizio 4.21. Determinare la trasformazione che consente, a partire da una variabile aleatoria $U \sim U(0, 1)$, di generare una variabile aleatoria X avente pdf

$$f_X(x) = \begin{cases} 12(x - 0.5)^2, & 0 < x < 1; \\ 0, & \text{altrimenti.} \end{cases}$$

Esercizio 4.22. Si consideri la variabile aleatoria X definita come

$$X = \begin{cases} (2U)^{1/2}, & U < 0.5; \\ 2 - (2 - 2U)^{1/2}, & U \geq 0.5. \end{cases}$$

con $U \sim U(0, 1)$. Mostrare che X ha una pdf *triangolare* in $(0, 2)$.

Caratterizzazione sintetica di una variabile aleatoria

In questo capitolo si introducono le principali grandezze (media, varianza, valor quadratico medio) che consentono di fornire la cosiddetta caratterizzazione sintetica di una variabile aleatoria, mostrando altresì che tali grandezze appartengono alla più ampia classe dei momenti di una variabile aleatoria. Si introduce poi il teorema fondamentale della media, che semplifica notevolmente il calcolo dei momenti di una variabile aleatoria Y ottenuta mediante trasformazione $Y = g(X)$ da un'altra variabile aleatoria X . Infine, si introducono le disuguaglianze fondamentali (Markov, Bienaymé, Chebishev) che consentono di legare tra loro alcuni momenti con i valori di probabilità; in particolare, la disuguaglianza di Chebishev fornisce un'interpretazione della varianza come indice di dispersione e mette in relazione i valori assunti dalla varianza con quelli della probabilità che una variabile aleatoria assuma valori in prossimità della sua media.

5.1 Introduzione

Abbiamo visto che una variabile aleatoria X è completamente descritta (“completamente caratterizzata”, in gergo probabilistico) dalla conoscenza della sua CDF, pdf, o DF. In molti casi pratici, tuttavia, tale informazione è eccessivamente dettagliata oppure è difficile da ottenere, mentre invece è interessante conoscere solo alcuni parametri *numerici* della variabile aleatoria, che sono genericamente denominati *momenti*. Tali parametri forniscono informazioni *sintetiche* (rispetto alla conoscenza della CDF, pdf, o DF) sulla variabile aleatoria: si parla infatti in tal caso di *caratterizzazione sintetica* della variabile aleatoria in oggetto. Il primo passo per introdurre la caratterizzazione sintetica è quello di fornire la definizione di *media* (statistica) di una variabile aleatoria.

5.2 Media di una variabile aleatoria

La definizione dei momenti di una variabile aleatoria discende in maniera diretta del concetto fondamentale di *media* (statistica):

Definizione (media di una variabile aleatoria). La media (statistica) $E(X)$ di una variabile aleatoria X con pdf $f(x)$ è:

$$E(X) \triangleq \int_{-\infty}^{\infty} x f(x) dx \quad (5.1)$$

se tale integrale esiste finito.

Osserviamo che la media di una variabile aleatoria è un numero reale. Nella notazione $E[X]$, la E deriva dalla denominazione anglosassone di media come “expectation” (in italiano, diremmo “valore atteso”). Talvolta si usa indicare la media di una variabile aleatoria con la lettera greca μ ; per specificare, poi, che si tratta della media della variabile aleatoria X , useremo anche la notazione μ_X .

► **Esempio 5.1 (media di una variabile aleatoria uniforme).** Sia $X \sim U(a, b)$, allora si ha:

$$E(X) = \int_a^b x \frac{1}{b-a} dx = \frac{1}{b-a} \left[\frac{x^2}{2} \right]_{x=a}^{x=b} = \frac{a+b}{2},$$

per cui la media di X coincide con il punto medio dell'intervallo $[a, b]$. ◀

► **Esempio 5.2 (media di una variabile aleatoria esponenziale).** Sia $X \sim \text{Exp}(\lambda)$, allora si ha:

$$\begin{aligned} E(X) &= \int_0^{\infty} x \lambda e^{-\lambda x} dx = \int_0^{\infty} x \frac{d}{dx} [-e^{-\lambda x}] dx = (\text{per parti}) = \\ &= \left[-xe^{-\lambda x} \right]_{x=0}^{x=\infty} + \int_0^{\infty} e^{-\lambda x} dx = \frac{1}{\lambda}, \end{aligned}$$

per cui la media di X coincide con il reciproco del parametro λ . ◀

Che cosa rappresenta in pratica la media, o valore atteso? Dal punto di vista matematico, l'integrale nella (5.1) effettua una “media pesata” dei valori x , dove la “pesatura” è rappresentata dal valore $f(x)$ della pdf nel punto x , e quindi i valori x in corrispondenza dei quali la pdf è più grande vengono pesati maggiormente, e contribuiscono in maggior misura al valore della media. Con una similitudine tratta dalla fisica, possiamo pensare alla media $E(X)$ come al valore “baricentrico” della distribuzione (pdf) di probabilità (e difatti la definizione di media è formalmente identica alla definizione del baricentro di una distribuzione lineare di masse). In pratica, la media è una grandezza deterministica che si può interpretare come “rappresentativa” dei valori assunti dalla variabile aleatoria, ed in questo senso si parla di “valore atteso”; è questo l'uso che si fa correntemente della media quando si fanno affermazioni del tipo “i maschi italiani sono alti in media 172 cm” oppure “uno studente di Ingegneria impiega in media 2.3 mesi a preparare un esame”. Si noti tuttavia che, a dispetto dell'interpretazione precedente, per particolari forme della pdf la media potrebbe non coincidere con alcuno dei valori assunti dalla variabile aleatoria (ciò accade spesso per variabili aleatorie discrete). Altre grandezze deterministiche che possono essere assunte come “rappresentative” della variabile aleatoria sono la *mediana* (ovvero il valore

che non è superato con probabilità pari a 0.5, vedi § 3.2.3) e la *moda* (ovvero il valore in cui la pdf ha un massimo locale, vedi § 3.3.1).

► *Esempio 5.3 (media di una variabile aleatoria di Cauchy).* Per particolari pdf la media potrebbe non essere definita, nel senso che la funzione integranda nella (5.1) potrebbe non essere *sommabile*. È questo il caso di una variabile aleatoria $X \sim \text{Cauchy}(a)$, che ha pdf $f(x) = \frac{\alpha/\pi}{x^2 + a^2}$, per la quale l'integrale nella (5.1) si scrive esplicitamente come:

$$E(X) \triangleq \int_{-\infty}^{\infty} x f(x) dx = \int_{-\infty}^{\infty} x \frac{\alpha/\pi}{x^2 + a^2} dx.$$

La funzione integranda non è sommabile, in quanto decade a zero all'infinito come $1/x$. Pertanto, la media $E(X)$ di una variabile aleatoria di Cauchy non è definita.¹ ◀

Osserviamo che, se la media $E(X)$ esiste, e se la retta verticale di equazione $x = a$ è un asse di simmetria per $f(x)$, cioè se

$$f(a+x) = f(a-x), \quad \forall x \in \mathbb{R},$$

allora è facile dimostrare che $E(X) = a$.² In particolare, se $f(x)$ è una funzione pari, $x = 0$ è un asse di simmetria, per cui $E(X) = 0$ (variabile aleatoria a media nulla).

► *Esempio 5.4 (media di una variabile aleatoria gaussiana).* Sia $X \sim N(\mu, \sigma)$, ricordiamo che la sua pdf è (cfr. § 3.5.7)

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

Poiché tale funzione ha chiaramente $x = \mu$ come asse di simmetria, allora risulta necessariamente $E(X) = \mu$ (notiamo che la media esiste, in quanto la funzione $x f(x)$ è sicuramente sommabile, in quanto di tipo esponenziale). Pertanto il parametro μ , caratteristico di una variabile aleatoria gaussiana, ne rappresenta la media $E(X)$. ◀

Vediamo come si particularizza la definizione di media al caso in cui X è una variabile aleatoria discreta. In tal caso, la pdf $f(x)$ si riduce (cfr. § 3.3) ad una somma discreta di impulsi di Dirac, del tipo

$$f(x) = \sum_{x_j \in \mathcal{X}} p_j \delta(x - x_j),$$

dove $p_j = P(X = x_j)$, per cui, sostituendo la pdf nella definizione di media, si ottiene con facili passaggi:

$$\begin{aligned} E(X) &= \int_{-\infty}^{\infty} x f(x) dx = \int_{-\infty}^{\infty} x \sum_{x_j \in \mathcal{X}} p_j \delta(x - x_j) dx \\ &= \sum_{x_j \in \mathcal{X}} p_j \int_{-\infty}^{\infty} x \delta(x - x_j) dx = \sum_{x_j \in \mathcal{X}} p_j x_j \\ &= \sum_{x_j \in \mathcal{X}} x_j P(X = x_j) = \sum_{x_j \in \mathcal{X}} x_j p(x_j), \end{aligned}$$

ovvero la media si può esprimere, anziché attraverso un integrale, mediante una *sommatoria* dei valori $x_j \in \mathcal{X}$ della variabile aleatoria discreta X , ciascuno pesato per la DF $p(x)$ calcolata nel punto x_j ("somma pesata"). Se i valori x_j sono in numero finito ed equiprobabili, la media statistica si riduce alla semplice *media aritmetica* dei valori x_j .

¹Notiamo che se, viceversa, si adoperasse nella (5.1) la definizione di *integrale a valor principale secondo Cauchy* o *integrale improprio*, la media risulterebbe nulla per la simmetria della pdf.

²In questo caso, si può anche verificare facilmente che, se $x = a$ è anche un punto di massimo locale della pdf, allora media, moda e mediana coincidono.

► *Esempio 5.5 (media di una variabile aleatoria di Bernoulli).* Sia $X \sim \text{Bern}(p)$, allora

$$E(X) = 0 \cdot P(X=0) + 1 \cdot P(X=1) = 0 \cdot q + 1 \cdot p = p.$$

Si noti come la media (salvo nei casi, peraltro poco interessanti, in cui $p=0$ oppure $p=1$) non coincide con alcun valore assunto dalla variabile aleatoria X . ◀

► *Esempio 5.6 (media di una variabile aleatoria binomiale).* Sia $X \sim B(n, p)$, allora

$$E(X) = \sum_{k=0}^n k P(X=k) = \sum_{k=0}^n k \binom{n}{k} p^k (1-p)^{n-k} = np.$$

Per ottenere tale risultato, occorre sfruttare opportunamente le proprietà dei coefficienti binomiali. Si ha:

$$\begin{aligned} E(X) &= \sum_{k=0}^n k \binom{n}{k} p^k (1-p)^{n-k} = \sum_{k=1}^n k \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} \\ &= \sum_{k=1}^n \frac{n!}{(k-1)!(n-k)!} p^k (1-p)^{n-k} = \sum_{h=0}^{n-1} \frac{n!}{h!(n-h-1)!} p^{h+1} (1-p)^{n-h-1} \\ &= np \underbrace{\sum_{h=0}^{n-1} \frac{(n-1)!}{h!(n-1-h)!} p^h (1-p)^{n-1-h}}_{=(p+q)^{n-1}=1} = np. \end{aligned}$$

Vedremo nel seguito (cfr. esempio 8.2) che il risultato precedente si può giustificare assai più semplicemente riguardando la variabile aleatoria binomiale come la *somma* di n variabili aleatorie bernoulliane. Osserviamo infine che solo nel caso in cui np sia intero, la media coincide con uno dei valori assunti dalla variabile aleatoria X . ◀

► *Esempio 5.7 (media di una variabile aleatoria indicatrice di un evento).* Sia X_A la variabile aleatoria indicatrice di un evento A (vedi esempio 3.10), e cioè:

$$X_A(\omega) = \begin{cases} 1, & \text{se } \omega \in A; \\ 0, & \text{se } \omega \notin A. \end{cases}$$

Tale variabile aleatoria è ovviamente discreta, e assume i valori 1 e 0 con probabilità $P(A)$ e $P(\bar{A})$. Si ha, allora:

$$E(X_A) = 1 \cdot P(A) + 0 \cdot P(\bar{A}) = P(A).$$

Questo esempio evidenzia che la probabilità di un evento A si può interpretare come media della variabile aleatoria indicatrice dell'evento stesso. ◀

► *Esempio 5.8 (media di una costante).* Sia $X = a$ una variabile aleatoria costante, che assume l'unico valore reale a con probabilità 1. Poiché la sua pdf è $f(x) = \delta(x-a)$, si ha:

$$E(X) = \int_{-\infty}^{\infty} x \delta(x-a) dx = a$$

e pertanto $E(a) = a$. ◀

5.2.1 Teorema fondamentale della media

Occupiamoci adesso del problema, ricorrente nelle applicazioni, di calcolare la media di una variabile aleatoria $Y = g(X)$ ottenuta come trasformazione di un'altra variabile aleatoria X . Applicando la definizione di media per Y , si ha:

$$E(Y) \triangleq \int_{-\infty}^{\infty} y f_Y(y) dy.$$

Pertanto, per determinare $E(Y)$, sembra necessario calcolare la pdf $f_Y(y)$, il che può farsi adoperando il teorema fondamentale 4.1 sulle trasformazioni di variabili aleatorie. Tale conclusione non è però del tutto corretta, in virtù del seguente *teorema fondamentale della media*, che enunciamo senza dimostrazione:

Teorema 5.1 (teorema fondamentale della media). Sia $Y = g(X)$ una trasformazione della variabile aleatoria X avente pdf $f_X(x)$, si ha:

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

se tale integrale esiste finito.

Nel caso in cui X sia una variabile aleatoria discreta con DF $p(x)$, abbiamo avuto già modo di osservare che anche $Y = g(X)$ sarà una variabile aleatoria discreta, ed il teorema fondamentale della media si può esprimere come:

$$E(Y) = E[g(X)] = \sum_{x_i \in \mathcal{X}} g(x_i) P(X = x_i) = \sum_{x_i \in \mathcal{X}} g(x_i) p_X(x_i)$$

ovvero la media di Y si esprime in termini della DF $p_X(x)$ di X . In questo caso si può fornire una dimostrazione semplice del teorema (si veda [4] oppure [5]).

► *Esempio 5.9.* Sia $X \sim U(0, 2\pi)$, e si voglia calcolare la media di $Y = \cos(X)$. Applicando il teorema fondamentale, scriviamo:

$$E(Y) = E[\cos(X)] = \frac{1}{2\pi} \int_0^{2\pi} \cos(x) dx = \frac{1}{2\pi} [\sin(x)]_{x=0}^{x=2\pi} = 0$$

per cui $E(Y) = 0$ e non è stato necessario calcolare esplicitamente la pdf di Y , la cui espressione abbiamo peraltro derivato nell'esempio 4.3 (si tratta di una pdf pari, per cui effettivamente $E(Y) = 0$). ◀

5.2.2 Proprietà della media

L'operazione di media gode delle seguenti proprietà, che vengono enunciate senza dimostrazione:

1. Siano $g(\cdot)$ e $h(\cdot)$ funzioni reali, e siano a e b costanti reali. Si ha:

$$E[ag(X) + bh(X)] = aE[g(X)] + bE[h(X)].$$

In particolare, si ha:

$$E(aX + b) = aE(X) + b,$$

in quanto $E(b) = b$. Tale fondamentale proprietà va sotto il nome di *linearità della media*.

2. Se $g(x) \geq 0$ per ogni x , allora $E[g(X)] \geq 0$.
3. Se $g_1(x) \geq g_2(x)$ per ogni x , allora $E[g_1(X)] \geq E[g_2(X)]$.
4. Se $a \leq g(x) \leq b$ per ogni x , allora $a \leq E[g(X)] \leq b$.

5.3 Varianza e valor quadratico medio di una variabile aleatoria

Passiamo ora a definire un altro importante parametro sintetico di una variabile aleatoria X , ovvero la sua *varianza*:

Definizione (varianza di una variabile aleatoria). La varianza $\sigma^2 = \text{Var}(X)$ di una variabile aleatoria X con media $\mu = E(X)$ è:

$$\sigma^2 = \text{Var}(X) \triangleq E[(X - \mu)^2] = \int_{-\infty}^{\infty} (x - \mu)^2 f(x) dx, \quad (5.2)$$

se tale integrale esiste finito.

Notiamo che la definizione precedente si può interpretare anche come l'applicazione del teorema fondamentale della media al calcolo della media di $Y = g(X) = (X - \mu)^2$.

La varianza è una quantità non negativa: la sua radice quadrata $\sigma \triangleq \sqrt{\text{Var}(X)}$ prende il nome di *deviazione standard* della variabile aleatoria X ; si noti che la varianza è dimensionalmente omeogenea al quadrato della variabile aleatoria, mentre la deviazione standard ha le stesse dimensioni della variabile aleatoria. Useremo anche la notazione σ_X per denotare esplicitamente che si tratta della deviazione standard della variabile aleatoria X .

Sviluppando algebricamente il quadrato che compare nella definizione di varianza, ed adoperando la proprietà di linearità della media, si ha, con semplici passaggi,

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

ovvero la relazione fondamentale

$$\text{Var}(X) = E(X^2) - E^2(X). \quad (5.3)$$

La quantità $E(X^2)$ (anch'essa non negativa) si calcola applicando il teorema fondamentale della media e prende il nome di *valore quadratico medio* (vqm):

Definizione (valor quadratico medio di una variabile aleatoria). Il valore quadratico medio $E(X^2)$ di una variabile aleatoria X è:

$$E(X^2) \triangleq \int_{-\infty}^{\infty} x^2 f(x) dx,$$

se tale integrale esiste finito.

La radice quadrata $x_{\text{rms}} \triangleq \sqrt{E(X^2)}$ del valore quadratico medio prende il nome di *valore efficace* della variabile aleatoria X , ed è dimensionalmente omogeneo ad X .³

La relazione (5.3) tra varianza, media e valor quadratico medio è fondamentale, e mostra che solo due tra questi tre parametri possono essere assegnati arbitrariamente, in quanto il terzo dipende univocamente dagli altri due. Inoltre, la relazione (5.3) mostra che, per variabili aleatorie a media nulla, la varianza coincide con il valor quadratico medio, e quindi la deviazione standard coincide con il valore efficace.

► **Esempio 5.10 (varianza di una variabile aleatoria uniforme a media nulla).** Consideriamo il caso di $X \sim U(-\Delta/2, \Delta/2)$, allora $\mu = E(X) = 0$, e si ha:

$$\sigma^2 = E[X^2] = \int_{-\Delta/2}^{\Delta/2} x^2 \frac{1}{\Delta} dx = \frac{1}{\Delta} \left[\frac{x^3}{3} \right]_{x=-\Delta/2}^{x=\Delta/2} = \frac{\Delta^2}{12}.$$

Osserviamo come la varianza cresca al crescere dell'ampiezza Δ dell'intervallo in cui la variabile aleatoria X assume i suoi valori. Il caso di una variabile aleatoria uniforme a media non nulla è trattato nell'esempio 5.14. ◀

► **Esempio 5.11 (varianza di una costante).** Sia $X = a$ una variabile aleatoria costante, che assume l'unico valore reale a con probabilità 1. È immediato verificare che la sua varianza è nulla, in quanto risulta $X - \mu = a - a = 0$. ◀

Qual è l'interpretazione della varianza? Notiamo che l'integrale (5.2) effettua una media pesata, con funzione di peso $f(x)$, degli scarti quadratici $(x - \mu)^2$ tra i valori assunti dalla variabile aleatoria e la sua media. La varianza σ^2 , pertanto, misura la *concentrazione* (o, equivalentemente, la *dispersione*) di X intorno alla sua media μ . In altri termini, se una variabile aleatoria ha varianza *piccola*, allora essa è poco dispersa intorno alla sua media (assumerà con maggior probabilità valori intorno alla media); viceversa, se una variabile aleatoria ha varianza *grande*, allora essa è molto dispersa intorno alla sua media (assumerà con probabilità non trascurabile valori assai lontani dalla media).⁴ Possiamo equivalentemente dire che la varianza è una misura dell'*incertezza* associata ai valori della variabile aleatoria X ; infatti una variabile aleatoria costante ($X = a$) ha varianza nulla, perchè non c'è nessuna incertezza sui valori che può assumere.

Adoperando una similitudine fisica, come la media è equivalente al *baricentro* di una distribuzione di masse, così la varianza rappresenta (e la sua espressione matematica è formalmente equivalente) il *momento di inerzia* della distribuzione di masse rispetto al baricentro.

► **Esempio 5.12 (varianza di una variabile aleatoria gaussiana).** Sia $X \sim N(\mu, \sigma)$: vogliamo verificare che σ^2 rappresenta proprio la varianza di X , e quindi σ la sua deviazione standard. Per provarlo, ricorriamo ad un artificio: avendo già dimostrato che $\mu = E(X)$, consideriamo l'integrale (condizione di normalizzazione per una pdf):

$$\int_{-\infty}^{\infty} f(x) dx = 1,$$

che per la pdf gaussiana si scrive esplicitamente:

$$\int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = 1,$$

³Il pedice "rms" sta per "root mean square", che è la denominazione inglese per "radice del valor quadratico medio".

⁴Un legame quantitativo più preciso tra il valore della varianza e la probabilità con cui la variabile aleatoria assume valori nell'intorno della media è fornito dalla fondamentale *disuguaglianza di Chebishev* (vedi § 5.5).

ovvero:

$$\int_{-\infty}^{\infty} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = \sigma\sqrt{2\pi}.$$

Poiché quest'identità vale per ogni $\sigma > 0$, deriviamola rispetto a σ :

$$\int_{-\infty}^{\infty} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \frac{(x-\mu)^2}{\sigma^3} dx = \sqrt{2\pi}$$

da cui, con banali manipolazioni algebriche,

$$\text{Var}(X) = \int_{-\infty}^{\infty} (x-\mu)^2 \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = \sigma^2$$

che è quanto volevamo provare. ◀

Se X è una variabile aleatoria discreta, allora la sua pdf è del tipo

$$f(x) = \sum_{x_i \in \mathcal{X}} p_i \delta(x - x_i)$$

e quindi, sostituendo la pdf nella definizione di varianza, con facili passaggi si ottiene:

$$\sigma^2 = \sum_{x_i \in \mathcal{X}} p_i (x_i - \mu)^2 = \sum_{x_i \in \mathcal{X}} p(x_i) (x_i - \mu)^2$$

dove $p(x)$ è la DF di X . Ovviamente, anche per variabili aleatorie discrete vale la fondamentale relazione (5.3) tra varianza, media e valor quadratico medio.

► *Esempio 5.13 (varianza di una variabile aleatoria di Bernoulli).* Sia $X \sim \text{Bern}(p)$, allora, poiché:

$$\begin{aligned} E(X) &= 1 \cdot p + 0 \cdot q = p, \\ E(X^2) &= 1^2 \cdot p + 0^2 \cdot q = p, \end{aligned}$$

applicando la (5.3) si ha:

$$\sigma^2 = E(X^2) - E^2(X) = p - p^2 = p(1-p) = pq.$$

Si noti che tale varianza, al variare di p e q , assume il valore massimo per $p = q = 0.5$ (condizione di massima incertezza). ◀

5.3.1 Proprietà della varianza

Come è evidente dalla sua definizione, la varianza *non* è un operatore lineare, ma *quadratico*: si pone allora il problema di come calcolare la varianza di $Y = aX + b$. Se X è una variabile aleatoria con varianza finita, qualunque siano le costanti reali a e b , si ha la fondamentale relazione:

$$\boxed{\text{Var}(aX + b) = a^2 \text{Var}(X)}. \quad (5.4)$$

Prova. Sulla base della definizione, si ha:

$$\text{Var}(aX + b) = E\{[aX + b - E(aX + b)]^2\}$$

Adoperando la linearità della media, con semplici passaggi possiamo scrivere:

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

che è il risultato cercato. ◻

Notiamo che, nella trasformazione $Y = aX + b$, la varianza di Y non dipende da b , e quindi in particolare la varianza di $Y = X + b$ coincide con quella di X per qualunque valore della traslazione b . Tale risultato è intuitivamente chiaro se si considera che la varianza misura la dispersione intorno alla media: una traslazione di b modifica evidentemente la media, ma non la dispersione intorno ad essa. Questa proprietà di *invarianza per traslazione* della varianza consente di scegliere opportunamente b nei casi pratici in maniera tale da semplificare il calcolo della varianza. In particolare, scegliendo $b = -\mu_X$ si costruisce la variabile aleatoria *centrata* $Y = X - \mu_X$ che ha media nulla e la stessa varianza di X . Notiamo che, per il teorema sulle trasformazioni di variabili aleatorie (cfr. § 4.2.3), la pdf di Y si otterrà semplicemente per traslazione della pdf di X , ovvero $f_Y(y) = f_X(y + \mu_X)$.

► *Esempio 5.14 (varianza di una variabile aleatoria uniforme).* Sia $X \sim U(a, b)$, allora $\mu_X = E(X) = \frac{a+b}{2}$. La variabile aleatoria centrata $Y = X - \mu_X$ avrà media nulla e sarà ancora uniforme, ma nell'intervallo $(-\Delta/2, \Delta/2)$, con $\Delta = b - a$. Pertanto, ricordando il risultato dell'esempio 5.10, si ha:

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

In questo caso, l'applicazione della proprietà (5.4) ha consentito di semplificare il calcolo, riconducendosi ad una variabile aleatoria uniforme con la stessa varianza ma avente media nulla. ◀

In molti casi, a partire da una variabile aleatoria X , si desidera ottenere una variabile aleatoria *standard*, ovvero una variabile aleatoria Z a media nulla e varianza unitaria. È allora sufficiente dividere la variabile aleatoria centrata $Y = X - \mu_X$ per la deviazione standard di X , costruendo Z come:

$$Z = \frac{X - \mu_X}{\sigma_X}.$$

Infatti, è banale verificare che $E(Z) = 0$, mentre applicando la proprietà (5.4) si ha:

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

Notiamo che, poiché Z si ottiene da X attraverso una trasformazione lineare del tipo $Z = aX + b$, la pdf di Z sarà data (per il teorema fondamentale sulle trasformazioni di variabili aleatorie, cfr. § 4.2.3) da

$$f_Z(z) = \sigma_X f_X(\sigma_X z + \mu_X),$$

e pertanto tale trasformazione di *normalizzazione* non modifica sostanzialmente la “famiglia” a cui la pdf appartiene (nel senso che una variabile aleatoria uniforme resta uniforme, una variabile aleatoria gaussiana resta gaussiana, e così via) ma cambia solo la posizione della pdf sull'asse dell'ascisse (per effetto della traslazione di μ_X) e la scala della pdf (per effetto della moltiplicazione per σ_X sia dell'argomento che dei valori assunti). In particolare, se $X \sim N(\mu_X, \sigma_X)$, la variabile aleatoria $Z = \frac{X - \mu_X}{\sigma_X} \sim N(0, 1)$ è ancora gaussiana con media nulla e varianza unitaria, e prende il nome di *normale standard* (la sua CDF è proprio la funzione $\mathbb{G}(x)$ riportata in Appendice C).

► *Esempio 5.15.* La procedura di normalizzazione precedentemente descritta è particolarmente utile per il calcolo di valori di probabilità riguardanti le variabile aleatoria gaussiane, in quanto consente di ricondurre tale calcolo al caso di una gaussiana standard $Z \sim N(0, 1)$, che può essere effettuato utilizzando la funzione $\mathbb{G}(x)$. Si consideri ad esempio la variabile aleatoria $X \sim N(3, 0.5)$, della quale si desidera calcolare la

probabilità che assuma valori nell'intervallo $[2, 4]$. Si ha:

$$\begin{aligned} P(X \in [2, 4]) &= P(2 \leq X \leq 4) = P\left(\frac{2-3}{0.5} \leq \frac{X-3}{0.5} \leq \frac{4-3}{0.5}\right) \\ &= P(-2 \leq Z \leq 2) = \mathbb{G}(2) - \mathbb{G}(-2) = 2\mathbb{G}(2) - 1 = 0.9546 \end{aligned}$$

dove abbiamo sfruttato le proprietà della funzione $\mathbb{G}(x)$ e la tabella dei valori riportata in Appendice C. ◀

5.4 Momenti di una variabile aleatoria

La media, la varianza ed il valor quadratico medio appartengono ad una classe di grandezze sintetiche più generali, i *momenti* di una variabile aleatoria:⁵

Definizione (momento). Il momento di ordine $n \in \mathbb{N}$ di una variabile aleatoria X è:

$$\mu_n \triangleq E(X^n) = \int_{-\infty}^{\infty} x^n f(x) dx,$$

se l'integrale esiste finito.

Definizione (momento centrale). Il momento centrale di ordine $n \in \mathbb{N}$ di una variabile aleatoria X con media $\mu = E(X)$ è:

$$\sigma_n \triangleq E[(X - \mu)^n] = \int_{-\infty}^{\infty} (x - \mu)^n f(x) dx,$$

se l'integrale esiste finito.

Definizione (momento assoluto). Il momento assoluto di ordine $n \in \mathbb{N}$ di una variabile aleatoria X è:

$$E[|X|^n] = \int_{-\infty}^{\infty} |x|^n f(x) dx,$$

se l'integrale esiste finito.

Definizione (momento generalizzato/assoluto). Il momento generalizzato rispetto ad a di ordine $n \in \mathbb{N}$ di una variabile aleatoria X è:

$$E[(X - a)^n] = \int_{-\infty}^{\infty} (x - a)^n f(x) dx,$$

oppure nella versione assoluta è:

$$E[|X - a|^n] = \int_{-\infty}^{\infty} |x - a|^n f(x) dx,$$

se i corrispondenti integrali esistono finiti.

Le definizioni precedenti si particolarizzano facilmente al caso di variabili aleatorie discrete. Notiamo poi che, di tali definizioni, le più utilizzate sono quelle relative ai momenti propriamente

⁵La definizione delle grandezze che seguono si può sempre ricondurre all'applicazione del teorema fondamentale della media ad opportune trasformazioni della variabile aleatoria X .

detti (μ_n) ed ai momenti centrali (σ_n) . In particolare, osserviamo che la media $E(X) = \mu$ coincide con il momento μ_1 di ordine $n = 1$, che la varianza $\sigma^2 = \text{Var}(X)$ coincide con il momento centrale σ_2 di ordine $n = 2$, e infine che il valor quadratico medio $E(X^2)$ coincide con il momento μ_2 di ordine $n = 2$. I momenti con $n > 2$ sono meno utilizzati, e prendono il nome di *momenti di ordine superiore*.

Notiamo infine che la caratterizzazione di una variabile aleatoria in termini di momenti viene detta *caratterizzazione sintetica*, in quanto fornisce un'informazione ridotta (per l'appunto, "sintetica") rispetto alla conoscenza della CDF, pdf o DF. Infatti, mentre assegnare la CDF, pdf o DF di una variabile aleatoria X (caratterizzazione statistica o caratterizzazione completa) consente di calcolare un qualunque momento, la conoscenza di un sottoinsieme di momenti di X (caratterizzazione sintetica) non consente in generale di risalire alla CDF, pdf o DF.⁶

5.4.1 Relazione tra momenti e momenti centrali

È immediato ricavare i momenti centrali in funzione di quelli non centrali, sfruttando la formula per lo sviluppo della potenza n -esima di un binomio e la linearità della media. Si ha:

$$\begin{aligned}\sigma_n &= E[(X - \mu)^n] = E\left[\sum_{k=0}^n \binom{n}{k} X^k (-\mu)^{n-k}\right] = \\ &= \sum_{k=0}^n \binom{n}{k} E(X^k) (-\mu)^{n-k} = \sum_{k=0}^n \binom{n}{k} \mu_k (-\mu)^{n-k}.\end{aligned}$$

Osserviamo che il momento centrale σ_n di ordine n dipende dalla media μ e da tutti i momenti μ_k di ordine $k \leq n$.

Altrettanto immediato è ricavare i momenti non centrali in funzione di quelli centrali. Si ha:

$$\begin{aligned}\mu_n &= E[X^n] = E[(X - \mu + \mu)^n] = E\left[\sum_{k=0}^n \binom{n}{k} (X - \mu)^k \mu^{n-k}\right] \\ &= \sum_{k=0}^n \binom{n}{k} E[(X - \mu)^k] \mu^{n-k} = \sum_{k=0}^n \binom{n}{k} \sigma_k \mu^{n-k}.\end{aligned}\tag{5.5}$$

Anche qui il momento μ_n di ordine n dipende dalla media μ e da tutti i momenti centrali σ_k di ordine $k \leq n$.

► **Esempio 5.16 (momenti di una variabile aleatoria gaussiana).** Sia $X \sim N(\mu, \sigma)$: vogliamo calcolarne i momenti e i momenti centrali di ordine n qualsiasi.

Iniziamo con il calcolo dei momenti della normale standard $Z \sim N(0, 1)$. Infatti, poichè possiamo esprimere una generica gaussiana $X \sim N(\mu, \sigma)$ in termini della normale standard Z , come $X = \sigma Z + \mu$, potremo poi esprimere i momenti di X in funzione dei momenti di Z .

Poiché Z è a media nulla, momenti e momenti centrali coincidono: dobbiamo allora calcolare il generico momento di ordine n , dato da:

$$\mu_n = \sigma_n = E[Z^n] = \int_{-\infty}^{\infty} x^n f_Z(x) dx,$$

dove

$$f_Z(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}.$$

Notiamo che tali momenti esistono tutti finiti, perchè la funzione $x^n f_Z(x)$, per la natura esponenziale di $f_Z(x)$, è sommabile per ogni $n \in \mathbb{N}$. Poiché poi $f_Z(x)$ è una funzione pari, i momenti per n dispari risultano

⁶Il discorso è diverso se si suppone di conoscere *tutti* i momenti; in tal caso, sotto opportune ipotesi, è possibile risalire alla CDF, pdf o DF attraverso l'uso della *funzione caratteristica* (vedi ad esempio [3, § 5-5])

nulli, essendo definiti attraverso l'integrale di una funzione dispari; il calcolo va allora affrontato solo per n pari. Poiché il calcolo diretto dell'integrale per n pari è tuttavia complicato, utilizziamo un artificio simile a quello dell'esempio 5.12, ovvero partiamo dall'identità

$$\int_{-\infty}^{\infty} e^{-\alpha x^2} dx = \sqrt{\pi} \alpha^{-1/2},$$

valida per ogni $\alpha > 0$, che si può ottenere a partire dalla condizione di normalizzazione della pdf per una variabile aleatoria $X \sim N(0, \sigma)$ con $\sigma^2 = 1/(2\alpha)$. Derivando k volte rispetto ad α tale identità, si ottiene:

$$\begin{aligned} \int_{-\infty}^{\infty} e^{-\alpha x^2} (-x^2) dx &= \sqrt{\pi} \left(-\frac{1}{2}\right) \alpha^{-3/2} \\ \int_{-\infty}^{\infty} e^{-\alpha x^2} (-x^2)(-x^2) dx &= \sqrt{\pi} \left(-\frac{1}{2}\right) \left(-\frac{3}{2}\right) \alpha^{-5/2} \\ &\dots \\ \int_{-\infty}^{\infty} e^{-\alpha x^2} \underbrace{(-x^2)(-x^2)\dots(-x^2)}_{k \text{ termini}} dx &= \sqrt{\pi} \underbrace{\left(-\frac{1}{2}\right) \left(-\frac{3}{2}\right) \dots \left(-\frac{2k-1}{2}\right)}_{k \text{ termini}} \alpha^{-(2k+1)/2} \end{aligned}$$

L'ultima relazione può essere riscritta, con semplici manipolazioni algebriche, nella forma:

$$\int_{-\infty}^{\infty} e^{-\alpha x^2} x^{2k} dx = \sqrt{\frac{\pi}{\alpha}} 1 \cdot 3 \cdot \dots \cdot (2k-1) (2\alpha)^{-k}$$

da cui, portando $\sqrt{\frac{\pi}{\alpha}}$ al primo membro e ponendo $\alpha = 1/2$ si ottiene:

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^{2k} e^{-x^2/2} dx = 1 \cdot 3 \cdot \dots \cdot (2k-1) \triangleq (2k-1)!!$$

ovvero proprio il momento $E(Z^n)$ con $n = 2k$ pari.⁷ In definitiva, se $Z \sim N(0, 1)$, si ha:

$$E(Z^n) = \begin{cases} 0, & n \text{ dispari}; \\ (n-1)!!, & n \text{ pari}. \end{cases}$$

Possiamo adesso affrontare il caso più generale di $X \sim N(\mu, \sigma)$. Procediamo dapprima considerando il caso di una variabile aleatoria X con $\mu = 0$ (a media nulla), per la quale ovviamente i momenti ed i momenti centrali coincidono, ed inoltre si ha $X = \sigma Z$, per cui $E(X^n) = \sigma^n E(Z^n)$, e quindi:

$$E(X^n) = \begin{cases} 0, & n \text{ dispari}; \\ \sigma^n (n-1)!!, & n \text{ pari}. \end{cases} \quad (5.6)$$

In particolare, per una variabile aleatoria $X \sim N(0, \sigma)$, molto utilizzato è il momento del quarto ordine $E(X^4) = 3\sigma^4$.

Infine, se $\mu \neq 0$, e quindi $X \sim N(\mu, \sigma)$, i momenti centrali σ_n coincidono con quelli di $Y = X - \mu$ che è a media nulla, e quindi sono dati ancora dalla relazione (5.6):

$$\sigma_n \triangleq E[(X - \mu)^n] = \begin{cases} 0, & n \text{ dispari}; \\ \sigma^n (n-1)!!, & n \text{ pari}; \end{cases}$$

mentre i momenti non centrali si ottengono a partire da quelli centrali sfruttando la relazione (5.5). ◀

⁷Si noti che abbiamo utilizzato il simbolo !! (doppio fattoriale) per denotare il prodotto dei soli numeri *dispari* fino ad un numero specificato (vedi Appendice A).

5.5 Disuguaglianze notevoli

In questo paragrafo introdurremo tre disuguaglianze notevoli, che mettono in relazione tra loro momenti e probabilità. Procederemo introducendo la disuguaglianza di *Markov*, dalla quale discende quella di *Bienaymé*, ed infine particularizzando quest'ultima otterremo la fondamentale disuguaglianza di *Chebyshev*, la più importante sia teoricamente che nelle applicazioni. Quest'ultima ci consentirà, in particolare, di approfondire l'interpretazione della varianza di una variabile aleatoria X come indice di dispersione dei valori assunti dalla variabile aleatoria.

Teorema 5.2 (Disuguaglianza di Markov). Sia Y una variabile aleatoria positiva, cioè tale che $f_Y(y) \equiv 0$ per ogni $y < 0$, e con media $E(Y)$ finita. Si ha:

$$P(Y \geq \alpha) \leq \frac{E(Y)}{\alpha}$$

per ogni $\alpha > 0$.

Prova. Si ha, con facili passaggi, la seguente catena di disuguaglianze:

$$E(Y) = \int_0^{\infty} y f_Y(y) dy \geq \int_{\alpha}^{\infty} y f_Y(y) dy \geq \int_{\alpha}^{\infty} \alpha f_Y(y) dy = \alpha P(Y \geq \alpha),$$

da cui l'asserto. □

L'utilità della disuguaglianza di Markov consiste nella possibilità di valutare approssimativamente la probabilità che una variabile aleatoria positiva ecceda un dato valore α . In effetti, poiché $P(Y \geq \alpha)$ per una variabile aleatoria continua rappresenta la CDF complementare $\bar{F}(\alpha) = 1 - F_X(\alpha)$, allora la disuguaglianza di Markov fornisce un limite superiore per l'andamento della CDF complementare di una variabile aleatoria positiva, che non può decrescere più lentamente di $1/\alpha$. Tuttavia, in molti casi pratici la rapidità di decadimento a zero della CDF complementare è molto più rapido (ad esempio, è di tipo esponenziale) di quello previsto dalla disuguaglianza di Markov, come mostrato dal seguente esempio.

► *Esempio 5.17.* Sia $Y \sim \text{Exp}(\lambda)$, con $E(Y) = \frac{1}{\lambda}$. Essendo $F_Y(y) = [1 - e^{-\lambda y}] u(y)$, possiamo calcolare direttamente $P(Y \geq \alpha) = 1 - F_Y(\alpha) = e^{-\lambda \alpha}$. La disuguaglianza di Markov si scrive allora esplicitamente nella forma:

$$e^{-\lambda \alpha} \leq \frac{1}{\lambda \alpha}.$$

Tale disuguaglianza è senz'altro verificata, ma l'errore relativo tra primo membro (che decade con legge esponenziale) e secondo membro (che decade con legge iperbolica) cresce senza limiti al crescere di $\lambda \alpha$, come dimostrato dai valori riportati in Tab. 5.5. ◀

Teorema 5.3 (disuguaglianza di Bienaymé). Sia X una variabile aleatoria e sia b un numero reale. Si ha:

$$P(|X - b| \geq \varepsilon) \leq \frac{E(|X - b|^n)}{\varepsilon^n},$$

per ogni $n \in \mathbb{N}$ ed $\varepsilon > 0$.

Prova. Si ottiene banalmente dalla disuguaglianza di Markov ponendo $Y = |X - b|^n$ ed $\alpha = \varepsilon^n$, ed osservando che, poiché la funzione $y = x^n$ è monotona crescente per $n \in \mathbb{N}$, si ha

$$P(Y \geq \alpha) = P(|X - b|^n \geq \varepsilon^n) = P(|X - b| \geq \varepsilon).$$

Si osservi che $E(|X - b|^n)$ deve esistere finito. □

$\lambda\alpha$	$P(Y \geq \alpha)$ (Markov)	$P(Y \geq \alpha)$ (esatto)
2	$5 \cdot 10^{-1}$	$1.35 \cdot 10^{-1}$
5	$2 \cdot 10^{-1}$	$6.74 \cdot 10^{-3}$
10	$1 \cdot 10^{-1}$	$4.54 \cdot 10^{-5}$
20	$5 \cdot 10^{-2}$	$2.06 \cdot 10^{-9}$
50	$2 \cdot 10^{-2}$	$1.93 \cdot 10^{-22}$
100	$5 \cdot 10^{-2}$	$3.72 \cdot 10^{-44}$

Tab. 5.1. Confronto tra i valori di probabilità previsti dalla disuguaglianza di Markov e quelli esatti per una variabile aleatoria esponenziale Y di parametro λ .

La probabilità che compare nella disuguaglianza di Bienaymé è quella che la variabile aleatoria X non appartenga all'intervallo $(b - \varepsilon, b + \varepsilon)$. Tale probabilità a parità di ε , è tanto più piccola quanto più è piccolo il momento assoluto $E[|X - b|^n]$ rispetto a b , che quindi va interpretato come un *indice di dispersione* della variabile aleatoria intorno a b . Notiamo, in particolare, che se $b = \mu = E(X)$ e se n è pari, $E[|X - b|^n]$ coincide con il momento centrale σ_n di ordine n , che pertanto va interpretato, per n pari, come un indice di dispersione intorno alla media. Tale risultato vale in particolare per $n = 2$, e quindi per la varianza $\sigma_2 = \sigma^2$, ed è tanto importante da prendere il nome di *disuguaglianza di Chebishev*:

Teorema 5.4 (disuguaglianza di Chebishev). Sia X una variabile aleatoria con media μ e varianza σ^2 finite. Si ha:

$$P(|X - \mu| \geq \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2}, \quad (5.7)$$

per ogni $\varepsilon > 0$.

Prova. Si ottiene dalla disuguaglianza di Bienaymé per $b = \mu$ ed $n = 2$. □

Sulla base della disuguaglianza di Chebishev, la varianza può essere interpretata come il più semplice indice di dispersione dei valori assunti da una variabile aleatoria intorno alla sua media. Infatti, ponendo $\varepsilon = k\sigma$, possiamo anche riscrivere la (5.7) come

$$P(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}, \quad (5.8)$$

o equivalentemente come:

$$P(|X - \mu| < k\sigma) \geq 1 - \frac{1}{k^2}. \quad (5.9)$$

In particolare, la (5.9) consente di ottenere un limite inferiore per la probabilità che la variabile aleatoria X assuma valori nell'intervallo $(\mu - k\sigma, \mu + k\sigma)$, come evidenziato in Tab. 5.5, nella quale tali valori sono riportati per i valori di k da 1 a 5.

Ad esempio, per $k = 4$ la variabile aleatoria assume valori in $\mu \pm 4\sigma$ con probabilità superiore al 93%; pertanto, quanto più σ è piccola, tanto più tale intervallo sarà piccolo, e conterrà comunque una frazione superiore al 93% dei valori assunti dalla variabile aleatoria. In questo senso, σ misura la dispersione o variabilità della variabile aleatoria intorno alla media μ , ed è questo il motivo per cui, in ultima analisi, σ^2 è denominata *varianza*.

Osserviamo infine che poiché la disuguaglianza di Chebishev discende da quella di Markov, valgono per essa considerazioni analoghe a quelle già effettuate per la disuguaglianza di Markov relativamente allo scostamento – che può essere notevole – tra i valori effettivi di probabilità

k	intervallo	$P(X \in \text{intervallo})$
1	$\mu \pm \sigma$	≥ 0
2	$\mu \pm 2\sigma$	≥ 0.75
3	$\mu \pm 3\sigma$	≥ 0.89
4	$\mu \pm 4\sigma$	≥ 0.9375
5	$\mu \pm 5\sigma$	≥ 0.96

Tab. 5.2. Probabilità che la variabile aleatoria X appartenga ad un intervallo centrato intorno alla media previsti dalla disuguaglianza di Chebishev.

ed il limite previsto dalla disuguaglianza. L'utilità della disuguaglianza di Chebishev non sta tanto nell'accuratezza con la quale è in grado di fornire i valori della probabilità che la variabile aleatoria X appartenga ad un intervallo centrato intorno alla media, ma nella sua *generalità* e *semplicità*, in quanto consente di ottenere stime di tale probabilità senza richiedere la conoscenza esplicita della pdf o CDF della variabile aleatoria, ma solo della sua varianza.

5.6 Esercizi proposti

Esercizio 5.1. Calcolare la media e la varianza di una variabile aleatoria $X \sim \text{Bern}(p)$. [Risposta: $\mu = p$, $\sigma^2 = pq$.]

Esercizio 5.2. Calcolare la media e la varianza di una variabile aleatoria $X \sim \text{B}(n, p)$. [Risposta: $\mu = np$, $\sigma^2 = npq$.]

Esercizio 5.3. Calcolare la media e la varianza di una variabile aleatoria $X \sim \text{Geom}(p)$. [Risposta: $\mu = 1/p$, $\sigma^2 = q/p^2$.]

Esercizio 5.4. Calcolare la media e la varianza di una variabile aleatoria $X \sim \text{Poiss}(\lambda)$. [Risposta: $\mu = \lambda$, $\sigma^2 = \lambda$.]

Esercizio 5.5. Calcolare la media e la varianza di una variabile aleatoria $X \sim \text{U}(0, 2\pi)$. [Risposta: $\mu = \pi$, $\sigma^2 = \frac{\pi^2}{3}$.]

Esercizio 5.6. Calcolare la media e la varianza di una variabile aleatoria $X \sim \text{Exp}(\lambda)$. [Risposta: $\mu = 1/\lambda$, $\sigma^2 = 1/\lambda^2$.]

Esercizio 5.7. Calcolare la media e la varianza di una variabile aleatoria $X \sim \text{Lap}(\lambda)$. [Risposta: $\mu = 0$, $\sigma^2 = 2/\lambda^2$.]

Esercizio 5.8. Calcolare la media e la varianza di una variabile aleatoria $X \sim \text{Rayleigh}(b)$. [Risposta: $\mu = \sqrt{\pi b/4}$, $\sigma^2 = b(1 - \pi/4)$.]

Esercizio 5.9. Calcolare la media e la varianza di una variabile aleatoria X di tipo *Pareto*, avente cioè pdf:

$$f_X(x) = \frac{\alpha - 1}{x^\alpha} u(x - 1)$$

con $\alpha > 1$. [Risposta: $\mu = \frac{\alpha-1}{\alpha-2}$, per $\alpha > 2$; $\sigma^2 = \frac{\alpha-1}{(\alpha-3)(\alpha-2)^2}$, per $\alpha > 3$.]

Esercizio 5.10. Per ciascuna delle seguenti variabili aleatorie X , calcolare media e varianza.

- X variabile aleatoria continua con pdf $f_X(x) = \alpha x^{\alpha-1}$, $0 \leq x \leq 1$, $\alpha > 0$;
- X variabile aleatoria discreta con DF $p_X(k) = 1/n$, $k \in \{1, 2, \dots, n\}$, $n \in \mathbb{N}$;
- X variabile aleatoria continua con pdf $f_X(x) = \frac{3}{2}(x-1)^2$, $0 \leq x \leq 2$.

[Risposta: a) $\mu = \frac{\alpha}{\alpha+1}$, $\sigma^2 = \frac{\alpha}{(\alpha+2)(\alpha+1)^2}$; b) $\mu = \frac{n+1}{2}$, $\sigma^2 = \frac{n^2-1}{12}$; c) $\mu = 1$, $\sigma^2 = 3/5$.]

★ **Esercizio 5.11.** Sia X una variabile aleatoria continua non negativa. Mostrare che:

$$E(X) = \int_0^\infty [1 - F_X(x)] dx$$

Suggerimento: integrare per parti l'integrale tra $(0, y)$ e far tendere y ad infinito.

Esercizio 5.12. Dovete aprire la porta del vostro nuovo ufficio, ed il portiere vi ha dato un mazzo con n chiavi simili tra loro. Decidete di provarle tutte, a caso. In particolare, siete indecisi tra due strategie:

- non eliminare dal mazzo le chiavi che si dimostrano inutili;
- eliminare dal mazzo le chiavi che si dimostrano inutili.

Detta X la variabile aleatoria che conta il numero di tentativi che dovete effettuare per aprire la porta, determinare la DF di X ed il numero medio di tentativi utilizzando le due strategie. [Risposta: $E(X) = n$ (strategia 1), $E(X) = \frac{n+1}{2}$ (strategia 2).]

Esercizio 5.13. Se X è una variabile aleatoria con media e valor quadratico medio unitari, calcolare media e varianza della variabile aleatoria $Y = X + 1$.

Esercizio 5.14. Calcolare la media della variabile aleatoria $Y = -\ln(X)$, con $X \sim U(0, 1)$. [Risposta: $\mu = 1$]

Esercizio 5.15. Se $X \sim N(0, 1)$, calcolare media e varianza di $Y = |X|$. [Risposta: $\mu = \sqrt{2/\pi}$, $\sigma^2 = 1 - 2/\pi$]

Esercizio 5.16. Calcolare media e valore efficace della variabile aleatoria $Y = \cos(X)$, con $X \sim U(0, 2\pi)$. [Risposta: $\mu = 0$, $y_{\text{rms}} = \frac{1}{\sqrt{2}}$]

Esercizio 5.17. Sia X una variabile aleatoria avente la seguente pdf

$$f_X(x) = \begin{cases} \frac{1}{2}(1+x), & |x| \leq 1; \\ 0, & \text{altrimenti.} \end{cases}$$

Calcolare la media e la varianza di $Y = X^2$. [Risposta: $\mu = 1/3$; $\sigma^2 = 4/45$]

Esercizio 5.18. Un proiettile viene lanciato dal suolo con velocità iniziale v_0 e con angolo θ rispetto al suolo uniformemente distribuito tra 0 e $\pi/2$. Detta X la variabile aleatoria che rappresenta la distanza tra il punto in cui il proiettile è stato lanciato e quello di atterraggio, determinare la distanza mediamente percorsa dal proiettile (considerare il proiettile soggetto alla sola accelerazione di gravità g). [Risposta: $E(X) = \frac{2v_0^2}{\pi g}$]

Esercizio 5.19. Si supponga che la durata X , espressa in secondi, di una telefonata da un cellulare sia una variabile aleatoria esponenziale $X \sim \text{Exp}(\lambda)$, con media $E(X) = 180$. Il gestore A offre un piano tariffario a 3 lire al secondo con scatto di 200 lire alla risposta, per cui il costo della telefonata (in lire) si esprime come:

$$Y = \begin{cases} 200, & 0 < X \leq 3 \\ 200 + 3(X - 3), & X > 3 \end{cases}$$

Il gestore B offre un piano tariffario a 4 lire al secondo senza scatto alla risposta, per cui il costo della telefonata (in lire) si esprime semplicemente come $Y = 4X$.

Stabilire qual è il piano tariffario più conveniente con riferimento al *costo medio* di una telefonata.

Esercizio 5.20. Dimostrare che la media μ di una variabile aleatoria X è il valore b che rende minimo il momento generalizzato $E[(X - b)^2]$.

★ *Esercizio 5.21.* Dimostrare che la mediana m di una variabile aleatoria X è il valore b che rende minimo il momento generalizzato assoluto $E(|X - b|)$.

Suggerimento: utilizzare la formula di Leibnitz (Appendice F) per la derivazione.

Esercizio 5.22. Data una variabile aleatoria $X \sim N(\mu, \sigma)$, calcolare la probabilità che essa appartenga ad un intervallo $(\mu - k\sigma, \mu + k\sigma)$, con $k \in \{1, 2, 3, 4, 5\}$, e confrontare il risultato con i valori previsti dalla disuguaglianza di Chebishev.

